Nationwide Lake Impairment and Property Value

Saleh Mamun[1,2,3,*], Stephen Polasky[1,2,4], and Christoph Nolte[5,6]

July 15, 2022

(Here is the latest version)

[1] Department of Applied Economics, University of Minnesota, St. Paul, MN
[2] The Natural Capital Project, University of Minnesota, St. Paul, MN
[3] Natural Resources Research Institute, University of Minnesota – Duluth, Duluth, MN
[4] Department of Ecology, Evolution & Behavior, University of Minnesota, St. Paul, MN
[5] Department of Earth & Environment, Boston University, Boston, MA
[6] Faculty of Computing & Data Sciences, Boston University, Boston, MA

**corresponding author, email: salmamun@d.umn.edu*

Abstract

Under the U.S. Clean Water Act, states are required to assess, monitor, and list a waterbody as impaired when it fails to support its designated use(s). We examine the effect of lake impairment status on property values throughout the contiguous United States using the hedonic pricing model and nationwide geospatial data of sales transactions and property characteristics linked to a lake impairment database. Previous literature concentrates on regional level analysis based on measures of water quality that is subject to home buyers' and sellers' perception, that adds to information asymmetry. We contribute to the body of literature by using explicit expert-developed waterbody impairment status instead of non-explicit water quality measures, thus reducing bias generated from information asymmetry between house buyers and sellers. We also contribute by extending our estimates to the national level where previous studies are at the regional level. We estimate that information disclosed via impairment status induces a 4.47% reduction in sales prices within 150m of a lake. On average, the value of a property within 150m of an impaired lake decreases by $20,455. Our capitalization estimates imply that total property value depreciation of $5.4-16.9 billion can be attributed to properties within 150m of all lakes larger than 4ha. Although hedonic models provide lower bound estimates, our nationally representative robust capitalization effect demonstrates the importance of protecting waterbodies from being impaired.

JEL Classification: Q51, Q53, Q57,

Keywords: Ecosystem Services Valuation, Hedonic, Property Values, Water Quality, Water Impairment, ATTAINS, TMDL, WQS, LAGOS-NE.

1 INTRODUCTION

Surface freshwater bodies such as lakes, rivers, wetlands, and streams are crucial for people and the planet for providing critical ecosystem services including drinking water supply, water-based recreation, aesthetic and cultural values, and biodiversity-friendly habitats. Despite their importance, many waterbodies are highly degraded due to freshwater extraction and pollution. Freshwater extraction and pollution of waterbodies are ever-growing with the increased pressure from human economic activity. When a particular waterbody fails to meet designated use of the waterbody such as fishing or swimming, the waterbody is referred to as impaired. 38% of the world's waterbodies in the world are reported to be lacking good ambient water quality and hence fail to support designated uses (UNEP, 2021). According to US national statistics, 55.48% of rivers and streams length and 70.26% of lakes, ponds, and reservoirs area (excluding the Great Lakes) are declared impaired by state agencies (USEPA, 2017).

The leading causes of waterbody impairment are anthropogenic. Industrialization, urbanization, water engineering, land-use change, and agricultural practices (e.g., fertilizers and pesticide use) all are major sources of waterbody impairment in the US. When a waterbody fails to provide designated uses, it can have negative impacts on biodiversity and human health. Examples of negative effects include degraded aquatic habitat quality, eutrophication, drinking water contamination, and heavy metal in fish tissues. The benefit of an unimpaired waterbody can be similarly manifold: an unimpaired waterbody supports its designated uses, such as the provision of clean drinking water, water supply for agricultural and industrial uses, healthy aquatic habitat, recreational opportunities, or aesthetic amenities, among others.

In the US, states and federal agencies work together to improve water quality so that waterbodies can support more of their designated uses. Under the Clean Water Act 1972 (CWA), states, territories, and tribal authorities (hereinafter *reporting agency* in short) are required to monitor water quality in their jurisdiction every two years and report to the United States Environmental Protection Agency (USEPA). The reporting agencies 'list' a waterbody[1] as impaired if the water quality of the waterbody is too degraded to support its designated uses. This list of a reporting

[1] Instead of monitoring and listing individual waterbodies, the reporting agencies monitor areas of interest that they call *assessment units*. An assessment unit can be part of a waterbody, a whole waterbody, or multiple waterbodies.

agency's impaired waters is often referred to as "CWA 305(d)", after the corresponding section of the Clean Water Act. Once a waterbody is listed as impaired, the reporting agencies develop a total maximum daily load (TMDL) for pollutants in the waterbody as a starting point to develop clean-up plans. When the water quality is improved, as per the standards adopted by the reporting agencies in consultation with USEPA, the reporting agencies 'delist' the waterbody. A detailed description of the listing and delisting process is provided in Appendix A.

Under the CWA, billions of dollars have been directed to improve water quality (Liao, Wilhelm, & Solomon, 2016). However, the measurement of the true cost and benefit of water quality improvement is challenging due to the complexity of measuring water quality and its impact on human well-being. As water quality is a non-market good (a good that is not traded in the market), the benefit of water quality improvements (or cost of water quality deteriorations) cannot be captured through direct cost-benefit analysis due to lack of objective value signaling. Therefore, analysts have dedicated substantial effort to estimating the value of water quality benefits through non-market valuation techniques, such as stated preferences, hedonic analysis, travel cost method, and others (Bishop et al., 2020). In this study, we use a nationwide hedonic analysis to estimate the value of unimpaired lakes.

Hedonic analyses of water quality primarily rely on biophysical measurements of water quality to estimate the extent to which water quality is capitalized in property values, as observed in property sales transactions. This approach captures households' willingness to pay to live in proximity to a waterbody with better water quality. Water quality indicators used in these analyses include measures of water clarity (e.g., Secchi depth) (Gibbs, Halstead, Boyle, & Huang, 2002; Moore, Doubek, Xu, & Cardinale, 2020), biological productivity (e.g., chlorophyll A) (Weng et al., 2020; J. Zhang, Phaneuf, & Schaeffer, 2022), the presence of chemicals, toxins, or pathogens (Netusil, Kincaid, & Chang, 2014), invasive species (Horsch & Lewis, 2009; C. Zhang & Boyle, 2010), as well as composite water quality indicators. Appendix B provides an overview of prior hedonic analyses of water quality, including their geographic location, inferential methods, data sources, water quality measures, and key findings.

Two issues are not addressed in these hedonic analyses. First, the majority of these analyses use water clarity (Secchi depth) and chlorophyll A as measures of water quality, as these indicators are more frequently collected and shared than other water quality parameters. While many users prefer clearer waterbodies, water clarity does not necessarily correlate well with an expert-based definition of a healthy waterbody. A waterbody can be very clear even when it has a higher level of mercury and long-lasting Perfluoroalkyl and Polyfluoroalkyl Substances (PFAS). It might be aesthetically pleasing but contact or consumption can have serious health effects. Water clarity is also not necessarily associated with desirable ecological conditions, as in the example of invasive mussels in the Great Lake region. In addition, hedonic studies that rely on ambient water quality measures may fail to fully capture buyers' and sellers' true preferences for water quality due to information asymmetry between buyers and sellers. Second, because the cost of acquiring and aggregating data on water quality measures and sales transactions can be high (Sprague, Oelsner, & Argue, 2017), most analyses are conducted at relatively small spatial scales, with only few exceptions (Moore *et al.* 2020; J. Zhang, Phaneuf, & Schaeffer, 2022).

Our study contributes to the broader hedonic literature by addressing those two issues. Instead of using non-explicit water quality measures, we rely on explicit experts-developed objective measures of water quality criteria that indicate whether or not a waterbody is impaired and cannot support its designated uses. Poor *et al.* (2001) argues that objective measures of water quality are a better predictor of the housing market compared to subjective measures. Moreover, using impairment status is advantageous in the hedonic framework as it reduces information asymmetry between buyer and seller. Impairment status is public information and before listing an impaired waterbody, the reporting agencies are required to arrange public hearings and comments. Studies indicate that failure to provide full information in housing market transactions leads to misrepresentative conclusions (Gao, Song, & Timmins, 2021; Pope, 2008).

There are very few studies that estimate the effect of waterbody impairment on property values. Cho, Roberts and Kim (2011) uses impairment status of a portion of Pigeon River (10 out of its 18 subwatersheds) in Tennessee and North Carolina and found that there is a proximity benefit for the unimpaired parts of that waterbody. For waterbody impairment due to mercury pollution, Tang, Heintzelman and Holsen (2018) found a 6-7% decrease in property values using data from 34 counties in northern New York State. Papenfus (2019) found an 11% decrease in property value due to waterbody impairments in two counties of Washington state – Kitsap and King County located near Puget Sound. These three studies are conducted at the regional level, and none tracks impairment status over time. We extend their analytical framework both geographically to the contiguous US level and temporally by tracking water impairment status over time. We use the largest currently available data sources for both impairment status and property transactions. We collected, cleaned, processed, and solved identification issues for water impairment data from a national repository (the Assessment, Total Maximum Daily Load (TMDL) Tracking, and Implementation System, namely *ATTAINS*) (USEPA, 2021) and property data from Zillow's ZTRAX database (Zillow, 2019).

After careful consideration for sources of biases in hedonic model specifications, and following best practices laid out by Bishop *et al.* (2020), we estimate a significant negative effect of lake impairment on property prices. Our main estimate shows that there is a 4.5% decrease in lakeshore (within 0-150m buffer of a lake) property values if the lake is listed as impaired. Implicit price effects suggest that, on average, property values within 150m buffer are depressed by $20,455 (in 2019 U.S. dollars) if the lake is listed as impaired. Capitalization estimates of all lakes greater than 4ha in the contiguous US range from $5.4 to $16.9 billion. This is a lower bound estimate of lake impairment as our research design only captures effects on property prices. We also found significant heterogeneity of our estimates depending on the causes of lake impairment and the designated uses that were impaired. Our estimate suggests that the impairment of visual and recreational uses has the most significant negative effect on property values. Similar to previous studies of water quality, our results also confirm the distance decay effect of lake impairment.

The rest of the paper is organized as follows: Section 2 describes the econometric model, data, and database construction, section 3 presents the results, and section 4 discusses and concludes.

2 METHOD AND DATA

2.1 Method

We use a hedonic valuation model to estimate the impact of lake impairment on property value. Lancaster's modern consumer theory states that a good itself does not provide utility, rather the characteristics of a good rise in utility (Lancaster, 1966). This allows us to decompose a good into several attributes and get the value of each attribute. In hedonic analysis, we usually use market transactions of properties as a signal for environmental amenities or hazards. Suppose in a competitive market, a house buyer is willing to pay for each of the attributes the house entails. It is not necessarily limited to characteristics of the house structure only, instead all the neighborhood and environmental amenities or hazards associated with the house. Our main hedonic estimation model is log-linear form:

$$
\begin{aligned}
ln(P_{hct}) = \beta_i \times PostImpair_{hct} + \beta_l LakeShore_{hc} \\
+ \theta\ PostImpair_{hct} \times LakeShore_{hc} + \gamma X_{hct} + \sigma L_w + \eta_{ct} + \epsilon_{hct}
\end{aligned} \tag{1}
$$

The dependent variable, $ln(P_{hct})$ is the natural log of the price of property h, in the neighborhood (census tract) c, in the year of sale t, adjusted using the seasonally adjusted housing price index (Federal Housing Finance Agency, 2021). The normally distributed error term is given by ϵ_{hct}. We use census by sale year fixed effects (η_{ct}) to control for fine-scale spatial and temporal variation in property prices (unless mentioned otherwise). The standard errors are clustered at the census tract level to account for correlation within a census tract. We control for building and parcel characteristics (X_{hct}) and lake size (L_w). The dummy variable $PostImpair$ indicate if the nearest lake of the property is impaired or not at the time of sale. Following the literature, we hypothesize that the lake impairment status will have a decaying effect as the distance from the lake to the property increases. We capture this distance decay effect by including a $LakeShore$ dummy variable for properties located within 150m buffer of the lake.[2] Our variable of interest is the interaction term $PostImpair \times LakeShore$; in other words, our '*treatment*' properties are located within 150m of an impaired waterbody at the time of sale, and we compare these treatment properties with '*control*' properties that are either located within a 150-1500m distance buffer or within 1500m of a lake that is not impaired at the time of sale.[3]

Identification of '*treatment*' and '*control*' properties can be challenging using ATTAINS database. Distinction between impaired lakes and non-impaired lakes is complex due to several reasons. First, some lakes are only partially assessed, e.g., when the lake has multiple bays, or the lake falls into multiple state jurisdictions[4]. We therefore exclude lakes from the ATTAINS database if we find, in direct comparison to the corresponding lakes in the National Hydrography Dataset, that only a fraction of the lake was assessed. On the other hand, an assessment unit can be a watershed

[2] We choose 150m buffer as LakeShore because the mean distance from waterfront parcel centroid to lake boundary is about 150m in our sample.
[3] For one of the models, we used transaction till 2000m of the waterbody.
[4] A lake with multiple bays or jurisdictions can be assessed and found to have conflicting impairment statuses. Impairment status of part of a lake cannot be representation of whole lake when major part of a lake is not assessed.

and encompass multiple lakes[5]. We therefore also exclude five state reporting agencies (Florida, Louisiana, New Jersey, Ohio, Utah) that report water impairment status by watershed. Second, we use the ATTAINS parameter database to determine if a lake is impaired. The database groups assessment units (here: lakes) into five distinct categories. Category 1 is the list of lakes that fully support all designated uses. Category 2 and 3 lakes do not have enough information to determine if they are impaired or not. Categories 4 and 5 are impaired lakes where sub-categories 4A, 4B, 4C, and 5A have either TMDL or alternative clean-up plans, whichever are appropriate and approved by the USEPA. We classify categories 4 and 5 as impaired lakes, category 1 as unimpaired lakes, and exclude categories 2 and 3. Third, we determine if the lake is impaired at the time of property sale. A lake can be listed as impaired before it is included in the ATTAINS database. We therefore exclude property sales whose time of sale falls between the first time a lake was designated as impaired by reporting agency and the first time it was included in the ATTAINS database. This way we ensure that our treatment properties are within 150m of a fully impaired lake at the time of sale. Fourth, we exclude fishery-other, aquatic life, agricultural, and other categories of designated uses as those may be salient but housing market participants are less likely to pay attention to these use impairments[6]. We provide a robustness check that includes these designated uses to verify this assumption. A detailed identification strategy is provided in Appendix D (Identification strategy).

We consider a set of building and parcel characteristics as potential control variables, X_{hct}. Hedonic analyses commonly include any subset of control variables: building area, number of bedrooms, number of rooms, number of bathrooms, building age, lot size (area of the parcel), average slope of the parcel, average elevation of the parcel, distance from highway, and/or travel time to the nearest city. We also consider a set of neighborhood variables as potential controls, namely, building footprint within 5km of the property, population density in the census block group, and median income in the census block group. We consider missingness, multicollinearity, and variation within census tract by year as criteria to guide our final selection. Building area,

[5] The impairment status of a subset of assessment units (or lakes) within a watershed can be used to determine the impairment status of a whole watershed that has multiple lakes. The houseowner typically does not care about impairment status of watershed, instead they care about impairment status of nearby lake.
[6] The ATTAINS database provide a long list of use impairments. We grouped them together with the help of limnologist. Fishery-other represents different type of commercial fishing, and aquatic life represent habitat quality of aquatic species including algae, different species of fishes etc.

number of bedrooms, number of rooms, and number of bathrooms are highly correlated with each other and building area has the least missing observations. Travel time to the nearest city and distance to highway are also highly correlated to each other and travel time to the nearest city has the most missing values compared to the distance to highway. We also drop out neighborhood variables as they do not have much variation within our census-tract-by-year fixed effect. Our final model includes building area, building age, lot size, average slope of the parcel, average elevation of the parcel, and distance from highway as control variables. We also include the surface area of the waterbody variable (L_w) to control for the impact of waterbody sizes.

We estimate equation (1) for 1,838 lakes located in 39 states of the US. The number of property transactions in our base model is 660,707. We also estimate several variations of equation (1) depending on alternative definitions of variables and robustness checks. Two coefficients are of central interest in our estimation: β_i captures the impact of lake impairment on sales prices within 1500m of an impaired lake ($PostImpair$), while θ captures the additional impact of impairment on sales prices of lakeshore properties (within 150m of an impaired lake, $LakeShore$). As the functional form is log-linear, we can estimate the effect of switching lake impairment status from fully supporting to impaired by taking the derivative of equation (1):

$$\frac{\delta P}{\delta PostImpair} = e^{(\theta \times LakeShore + \beta_i)} - 1 \qquad (2)$$

From equation (2) we can calculate the effect of impairment on *LakeShore* properties based on the significance of θ and β_i. In case one of the parameters (θ or β_i) is not significant, we need to estimate the joint significance of the parameters. We estimate the joint significance of the equation parameters θ and β_i using the Chi-square test.

This is the first national-level study estimating the effect of water impairment on property value, made possible by recent advances in data collection, synthesis, and access that enable analyses with broad geographic coverage. However, there are still data limitations that challenge interpretations of the results. We could not estimate the effect of delisting waterbodies at a national scale due to the limited number of waterbodies that can be completely identified as delisted from the impairment list. Only a handful of waterbodies are delisted very recently in the 2020 reporting cycle, but our property data covers transactions between 2000-2019. Our estimates cannot be interpreted as having the similar opposite positive benefit of waterbody delisting from impairment.

Scaling up to a nationwide hedonic study poses several methodological challenges as the housing market is not homogenous and there is always a possibility of crossing over from one housing market to another – thus failing to fulfill the ‘law of one price’ (Bishop et al., 2020). One of the ways to address the issue is to let the coefficient of housing characteristics vary by spatial unit and temporal unit. This requires a significantly large number of observations to achieve statistical power of model specification, and hence demands a vibrant housing market with a significant number of transactions. Guignet and Nolte (2021) used this specification to estimate the impact of hazardous waste sites on property values. Note that those waste sites are within urban settings with over 9 million property transactions. On the other hand, lakeshore properties are less frequently up for sale at fair market value. We adapted for the second-best alternative and used fine-scale spatial fixed effect as suggested by Bishop *et al.* (2020). In our paper, we laid out a detailed identification strategy to reduce biases, but the selection of waterbody assessment units is not necessarily random. The reporting agencies select which waterbody to monitor using both target monitoring and probabilistic sampling. The targeted selection of assessment units can be a potential source of bias that we could not account for.

2.2 Property data and control variables

We used the Private-Land Conservation Evidence System (PLACES) dataset that is based on digital geospatial parcel maps. PLACES connects parcel boundary data with Zillow's Transaction and Assessment Database (ZTRAX, version: Oct 09, 2019) (Zillow, 2019). Using PLACES dataset is advantageous over directly using ZTRAX dataset as it ensures the correct location of a property and increases the likelihood of its fair market value. We improve the geographic information of properties by linking parcel boundaries with ZTRAX using the assessor's parcel number and a customized pattern matching algorithm (Nolte, 2020). We ignored the transactions if parcel subdivision and consolidation resulted in unsuccessful or partial linking of parcels with the ZTRAX dataset. To increase the likelihood that sales prices reflect fair market values, we remove transactions that are inter-family transfers, foreclosure, or undervalued public transactions. We also removed transactions if the sale price is less than $10,000 (Gindelsky et al., 2019) and the top 1 percentile of the sale price to address the remaining outliers.

We calculated the lot area from digital parcel boundary maps. The National Elevation dataset is spatially joined with parcel boundary maps to extract the average slope and elevation of the parcel

(U.S. Geological Survey, 2017a). We used the National Hydrography Dataset (NHD) to calculate the distance from a lake or river to the centroid of a parcel (U.S. Geological Survey, 2017b). Our final property dataset contains 660,707 single-family residential houses within 1500m from lakes greater than 4 ha for the years 2000-2019.

2.3 Water impairment data

Water impairment status data comes from USEPA's ATTAINS database (USEPA, 2021). USEPA stores integrated reports[7] submitted by reporting agencies in two separate data repositories – geospatial and parameter. The ATTAINS geospatial database provides the latest condition of assessment units along with its geographic information, while the ATTAINS parameter database provides the condition of assessment units over time. The former contains three separate layers of polygons (n=68,088), lines (n=352,479) and points (n=3,999). Lakes, impoundments, and reservoirs are represented in all three shape forms.

Although the geospatial database lists several causes of impairment and several designated impaired uses, it does not track the impairment status of the assessment unit over time. We linked ATTAINS geospatial database with parameter data to construct our water impairment data so that we can track impairment over time and space. The nomenclatures used for ATTAINS parameter database for designated uses, parameters, and causes are not standard and vary among reporting agencies. Detailed data cleaning and structuring techniques are provided in Appendix C (Data cleaning and database construction).

2.4 Spatially joining water impairment data with property data

The water impairment status in ATTAINS database is provided by assessment units which consist of multiple layers –polygon, lines, and points. As we use distance-based variables in our model, distance calculated from property location to various forms of assessment units may misrepresent the actual distance if we use all three layers the same way. To circumvent this problem, we use the NHD waterbody polygon layer as an intermediate step to spatially join water impairment data with property data. However, assessment units do not always correspond to the whole lake. We exclude lakes that are partially assessed in ATTAINS database. In addition, a lake can be divided into multiple assessment units and each of these assessment units can have different impairment

[7] It includes CWA section 303 (d) and section 305(b) listing of waterbody in reporting agencies' jurisdiction.

statuses. We also exclude those waterbodies from our analysis. A detailed description of the identification of waterbody impairment status is provided in Appendix D (Identification strategy). After identifying the corresponding lake and its impairment status over time we use NHD identifier to link ATTAINS data with property data.

2.5 Descriptive statistics

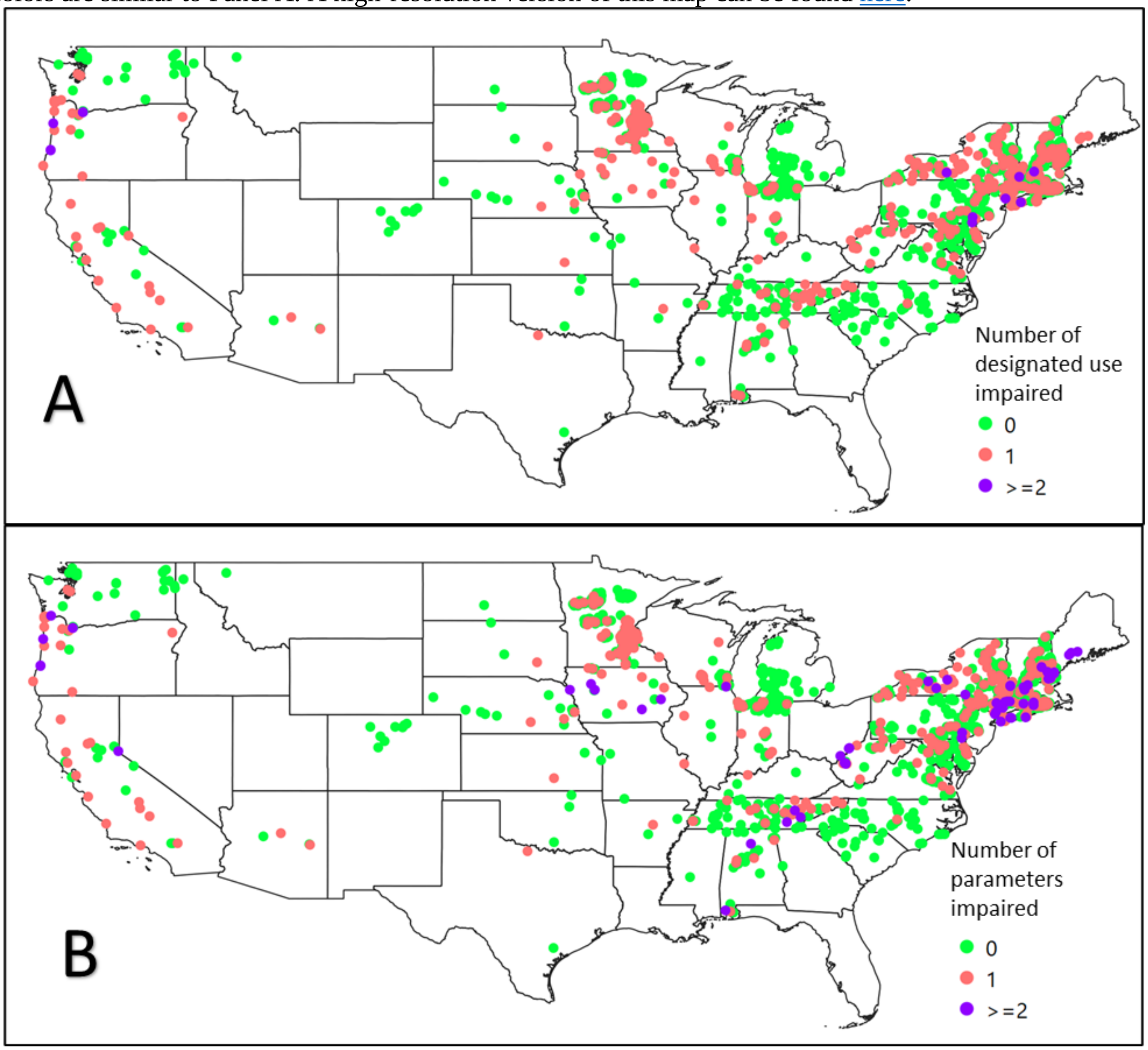
We started with nationwide data from the property database and the water impairment database. After carefully identifying treatment and control observations, we end up with 660,707 property transactions surrounding 1,838 lakes in 39 states of the US for our baseline model. Table 1 shows the descriptive statistics of our dataset. On average, 12% of properties in our data are located within 150m from the lakeshore. The dummy variable *PostImpair* is 1 when the nearest lake from the property is listed as impaired at the time of sale.

Table 1: Descriptive statistics of housing and lake characteristics in our sample. Our baseline model sample includes 1,838 lakes and 660,707 transactions of surrounding properties. Note that models of some robustness checks might have more or fewer transactions depending on the specifications of each model.

	min	max	mean	std dev
LakeShore (0-150m)	0	1	0.12	0.32
PostImpair	0	1	0.21	0.41
Lot area (m^2)	100	3,076,309	3,084	16,125
Average slope of the parcel (deg)	0	38.52	3.75	3.69
Average elevation of the parcel (m)	2	3041	291	367
Distance to highway (m)	15	3000	967	763
Building age (years)	1	323	37	31
Building area (m^2)	1	129,402	2,149	1,933
Lake size (km^2)	0.04	1124.14	4.98	38.64
Price ($2019)	10,000	3,008,330	360,088	265,685
Lakeshore property price ($2019)	10,000	3,008,330	457,903	356,829
Lake level summary				
Number of lakes			1,838	
Number of impaired lakes			550	
Mean number of sales per lake			440	

The average lot area of our sample is about 3000 m^2. The average slope and elevation of our sample is 3.75 degree and 290m respectively. For the distance to highway variable, we use a maximum cut-off of 3km to reduce computational complexity, which might exclude some very remote properties, but we do not think it will create a selection bias in our case. The building age and building area of our sample are 37 years and 2,149 m^2 on average respectively. The lakes included in our sample have a lower bound size of 4ha (0.04 km^2) with an average of 4.98 km^2. We used a lower bound cut-off of $10,000 for property prices before adjusting for inflation. The mean property price after adjusting for inflation is $360,088 ($2019) and $457,903 for *LakeShore* properties. Our final base model included 1,838 lakes (550 impaired lakes) in 39 US states where the average number of transactions per lake is 440.

Figure 1: The maps show selected lakes (n=1,838) in our analysis with impairment status and the number of parameters and designated uses impaired. Panel A shows the number of parameters impaired where green means waterbody is not impaired (n=1,288), red means only one parameter impairment and purple corresponds to multiple parameters impairment. Panel B shows number of impaired designated uses where colors are similar to Panel A. A high-resolution version of this map can be found here.



The study map of lake location along with impairment status and number of designated uses and parameters impaired and their intensity is shown in the Figure 1. The location of lakes that fulfill all the criteria described in the identification strategy is shown in panel A. The green color corresponds to category 1 (n=1,288) while red and purple correspond to category 4 or 5 (n=550). The red and purple color is assigned based on the number of designated uses and parameters impaired. Although our analysis is unprecedented in its spatial coverage, the number of lakes in the Mountain and West South-Central regions is low in our sample. One reason for less representation of our dataset in those regions is that they have a smaller number of lakes with surrounding properties. Another reason is the absence of transaction data from non-disclosure states, many of which are located in these regions. Only Arizona and Colorado are full disclosure states among the 10 states in these two regions.

3 RESULTS

3.1 Lake impairment negatively affects property values

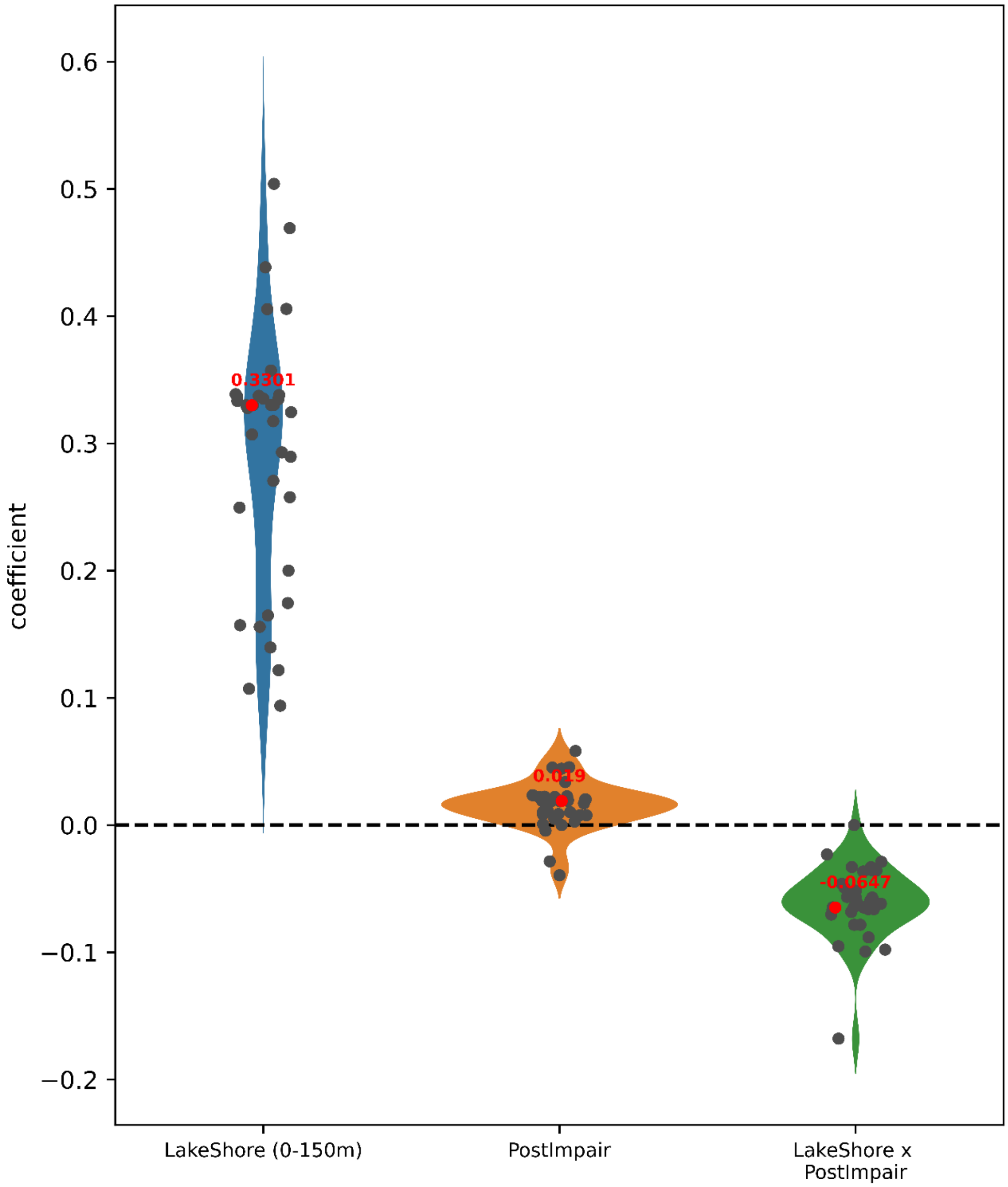
Table 2 summarizes the key parameter results of the baseline model. The parameter *PostImpair* measures the insignificant general impairment effect. Some literature also found an insignificant overall effect of water quality (Wolf & Klaiber, 2017), and lake impairment status (Papenfus, 2019), indicating that the effect is localized. The coefficient β_l captures the premium housing market participants pay to live close to a lake. On average, we find properties within 150m of the waterbody to sell for 39.11% more than properties within 150-1500m. The coefficient of interaction terms, θ indicates that lake impairment status depresses property values within this 150m buffer by 4.47%. Using average price properties within 150m buffer as $457,903, lakeshore properties decrease value by $20,455, on average.

Table 2: Key parameter estimates of the baseline model. A full estimation result is shown in Appendix E: Results Table S3.

	Description	**Baseline Model**
PostImpair [β_i]	=1, if the nearest lake is impaired at the time of sale	0.019 (0.0132)
LakeShore (0-150m) [β_l]	=1, if the property is located within 150m of the lake	0.3301*** (0.0126)
PostImpair x LakeShore [θ]	Interaction term of impairment indicator and lake proximity indicator	-0.0647*** (0.0194)

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1, n=660,707. Clustered standard errors are in parenthesis.

Figure 2: Violin plot showing water impairment effect on property values across all model specifications and robustness checks whenever all three of these variables are used in the same model (n=41). The red dots and values in red indicate our baseline model. The fatness of each violin indicates less variability, but it does not show significance level.



We plotted the estimation results for all the model specifications and robustness checks (n=41) in Figure 2[8]. The coefficient of the Lakeshore variable has relatively large variations across alternative definitions of treatment and control properties based on distance buffers. Regardless, the fat violin of the coefficient of $PostImpair$ and $PostImpair \times LakeShore$ variables indicates that they are almost similar in all the models – overall β_i is insignificant, close to zero and θ is small but significant and negative. In addition to key variables, we find significant and expected signs of control variables in our models.

3.2 Lake impairment impacts vary by uses and parameters impaired

We hypothesize that house buyers have heterogenous preferences for lake impairment based on designated use and pollutant that causes the impairments. Table 3 shows results for three frequently impaired designated uses that can potentially impact the housing market. We find a significant effect for aesthetic amenities and recreation use impairment. We find a 10.39% and 5.03% decrease in housing stock value due to aesthetic use and recreational use impairment. It is no wonder that the housing market has such a strong negative preference for aesthetic amenities impairment as those are highly salient. ‘Water supply’ use does not have a significant effect. For an average American housing market participant, water supply is not a concern as they get treated water in their house. Although water supply agencies pass the water treatment cost to end users, the end users usually do not decide where the water is supplied from and complexities of water pricing along with wastewater disposal make the end users perplexed about water supply.

Table 4 shows results for four categories of water quality parameters that are the most common reasons for impairment. We find a significant effect for pathogens, clarity, and nutrient load as these parameters lead to aesthetic and recreational use impairment which has significant value to the housing market participants. We did not find a significant effect for chlorine. As pollution or parameters can cause different use impairments, housing market participants respond directly to the use impairment instead of what pollutant is causing that impairment.

[8] Results of all models are provided in Appendix F, Dataset S1 (Excel file). There are altogether 45 models we run. We ploted results of 41 models where we used PostImpair x LakeShore as a regressor. We also calculate percentage of time our variables shows expected signs in tab ‘All_model_mag_sig’. Link: https://osf.io/gyd6f/files/osfstorage/63370e6e31d653045f2ddd7c

Table 3: Results of lake impairment on property values by designated use impairment. Note that, we could not estimate the 'Fishery' group separately as it has very few $PostImpair$ observations. We regrouped Fishery with Recreation as Fishery includes only three $useName$ from ATTAINS – 'Fishing', 'fishing', and 'Shellfishing' that reflects recreational fishing other than commercial fishing that is grouped in 'Fishing-other'.

	Aesthetic	**Recreation**	**Water Supply**	**Baseline Model**
PostImpair	0.0582**	0.0187	0.0078	0.019
	(-0.0231)	(-0.0137)	(-0.0468)	(-0.0132)
LakeShore (0-150m)	0.3380***	0.3278***	0.3387***	0.3301***
	(-0.014)	(-0.0127)	(-0.0139)	(-0.0126)
PostImpair x LakeShore	-0.1679***	-0.0703***	-0.0231	-0.0647***
	(-0.045)	(-0.0202)	(-0.042)	(-0.0194)
N	438,693	653,290	455,938	660,707
adj. R-squared	0.6466	0.6598	0.6481	0.6594

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

Table 4: Waterbody impairment effect by water quality parameters.

	Pathogens	**Clarity**	**Nutrient**	**Chlorine**	**Baseline Model**
PostImpair	-0.014	-0.0236	0.0086	-0.0036	0.019
	(-0.0379)	(-0.0412)	(-0.0156)	(-0.0338)	(-0.0132)
LakeShore (0-150m)	0.4746***	0.4759***	0.4760***	0.4782***	0.3301***
	(-0.016)	(-0.0163)	(-0.0156)	(-0.0165)	(-0.0126)
PostImpair x LakeShore	-0.0952*	-0.0866*	-0.1079***	-0.0163	-0.0647***
	(-0.0492)	(-0.0492)	(-0.03)	(-0.0417)	(-0.0194)
N	501,721	454,847	560,935	458,718	660,707
adj. R-squared	0.6645	0.6518	0.6537	0.6519	0.6594

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

3.3 Distance decaying effect of lake impairment

The distance decaying effect is captured in water quality hedonic literature by one of the three approaches: (a) the use of a distance-based dummy (Papenfus, 2019; Poor, Pessagno, & Paul, 2007; Walsh, Griffiths, Guignet, & Klemick, 2017)[9]; (b) the inclusion of a continuous distance (Liu, Opaluch, & Uchida, 2017) or an inverse distance (Wolf & Kemp, 2021) variable in the regression; and (c) the restriction of the sample to properties in proximity to the waterbody (Kashian, Eiswerth, & Skidmore, 2006; Poor et al., 2001).

[9] Poor et al (2006) used waterfront properties and adjacent waterfront properties, Papenfus used 150m buffer, Walsh et al. (2017) used Bayfront properties.

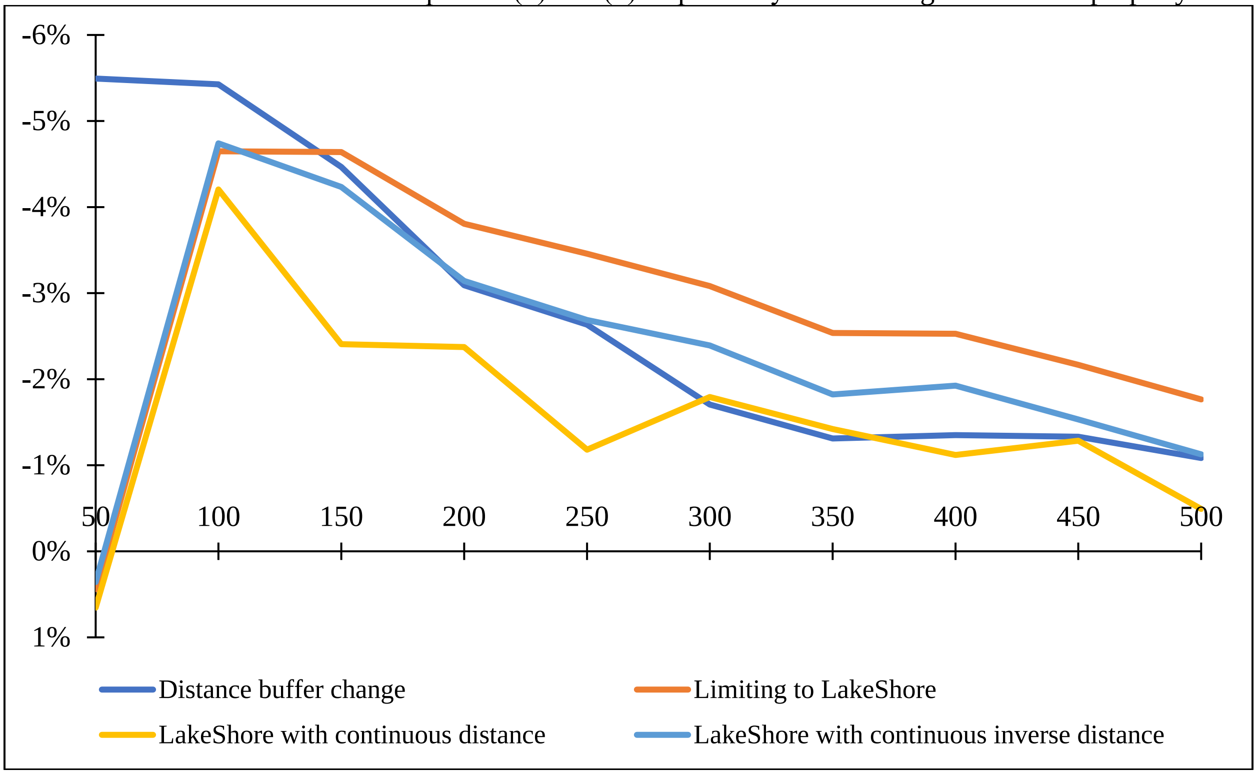
In our baseline model, we identify properties located close to a lake with our $LakeShore$ dummy variable, which identifies properties within 0-150m of a lake. We relax this definition of $LakeShore$ and estimate the effects of impairment on sales prices varying this threshold in steps of 50m up to a total distance of 500m. In addition, we use continuous distance and continuous inverse distance variables by limiting our dataset to different bins of 50m increments up to 500m. For continuous distance and inverse distance, our model specifications are equations (3) and (4) respectively. We also run a variant of equation (1) where we limited our sample to properties located within a certain distance from the waterbody and dropped any type of distance variable.

$$\begin{aligned} ln(P_{hct}) = \beta_i \times PostImpair_{hct} + \beta_l LakeDist_{hc} \\ + \theta PostImpair_{hct} \times LakeDist_{hc} + \gamma X_{hct} + \sigma L_w + \eta_{ct} + \epsilon_{hct} \end{aligned} \tag{3}$$

$$\begin{aligned} ln(P_{hct}) = \beta_i \times PostImpair_{hct} + \beta_l \left(\frac{1}{LakeDist_{hc}}\right) \\ + \theta PostImpair_{hct} \times \frac{1}{LakeDist_{hc}} + \gamma X_{hct} + \sigma L_w + \eta_{ct} + \epsilon_{hct} \end{aligned} \tag{4}$$

The results of using different methods of estimating distance decay effect are shown in Figure 3. Results show a general trend of diminishing effects of lake impairment as the distance from property parcel centroid to lake boundary increases. For most of the estimates, limiting the sample to properties within 50m of a lake provides small and insignificant results as there is a limited number of properties within that distance band. Although the impairment effects keep reducing till 500m, after 400m the effect size is small and generally insignificant. Literature on distance decay effect of lake water quality also supports the same conclusion. Wolf and Klaiber (2017) find insignificant water quality estimates beyond 250m. Walsh et al. (2017) found heterogenous distance decaying effect for counties around the Chesapeake Bay. While in some counties, the water quality effects are not significant beyond the waterfront, the water quality effect can go out to 1,000m from the shoreline for some counties (Walsh et al. 2017).

Figure 3: Figure showing distance decaying effect of lake impairment. As the housing units are farther away from the lake, the effect of waterbody impairment generally diminishes. For "distance buffer change", we used our baseline model with changing the definition of *LakeShore* dummy variable. For a second set of results, "limiting to lakeshore", we only used lakeshore property data and the dummy variable is defined if the property is waterfront or not. For "lakeshore with continuous distance" and "lakeshore with continuous inverse distance" models we used equation (3) and (4) respectively and limiting to lakeshore property data.



3.4 Welfare implications

The most critical limitation of hedonic studies is that they usually end in estimating an implicit price for a marginal change in environmental amenities or hazards (Banzhaf, 2021). Failure to provide welfare estimates leads to weak contributions in policy formulations. We provide a capitalization effect that requires additional assumptions to be made before interpreting it as a welfare estimate. The hedonic surface gradients need to be the same across the spatial and temporal dimensions. We did not allow coefficients to vary by spatial and temporal units due to the insufficient number of lakeshore property transactions. However, Kuminoff and Pope (2014) argued that the capitalization effect is lower bound compared to welfare measures.

Table 5: Capitalization loss estimates of waterbody impairment. Note that NLA (2012) and NLA (2017) estimates are calculated by region. A detailed calculation is provided in Appendix E: Results Table S4.

Impairment data source	**Extent**	**Percentage of lakes impaired**	**Estimated number of impaired lakes**	**Estimated number of lakeshore property of impaired lakes**	**Model estimate**	**Average value of lakeshore property**	**Total capitalization (billion USD)**
Sample	Sample (n=550)	100%	550	80,696	4.47%	366,406	1.32
Sample	All lakes (n=35,875)	29.92%	10,735	497,446	4.47%	323,820	7.20
ATTAINS	All lakes (n=35,875)	70.26%	25,206	1,167,985	4.47%	323,820	16.90
ATTAINS	All lakes (n=35,875)	70.26%	25,206	1,167,985	10.39%	323,820	39.30
NLA (2012)	All lakes (n=35,875)	25.18%	8,965	415,778	4.47%	352,039	5.96
NLA (2017)	All lakes (n=35,875)	24.51%	8,375	378,370	4.47%	352,039	5.40

We estimate capitalization effects from different perspectives and provide them in Table 5. First, we limit our estimates to our study sample for the most parsimonious estimate. In our final selection, 550 impaired lakes have 400,068 lakeshore properties with an average fair market value of $366,406.[10] As our baseline estimate is 4.47% depression in property value, we get an estimate of $3.18 billion ($2019). Our dataset contains lakeshore property information of all lakes greater than 4ha in the US. About 30% of lakes in our sample are impaired. Considering the average fair market value of $323,820 within all lakeshore properties (n=1.66 million), the total capitalization effect is a $3.5 billion decrease in total housing stock value.

USEPA's report to the congressional committee mentions that there are 70.26% of assessed lakes, impoundments, and reservoirs area are found impaired (USEPA, 2017). Using this higher percentage of impairment rate, we estimate there is a $8.21 billion decrease in property value nationwide. Using the same impairment rate from ATTAINS but changing the capitalization rate to the highest for aesthetic amenities, we find the capitalization effect as high as $29.36 billion. While this number can be contested as the selection of assessment units is not random, we look for external sources other than our sample from ATTAINS. The National Lake Assessment (NLA)

[10] Note that we are estimating for number of properties, not number of sales. Number of sales will be much less than number of properties. Here, fair market value represents value of the property even if it was not sold within 2000–2019-time window. Also note that the average fair market value is different from our average lakeshore property value. This is because this dataset consists of all the properties while our sample contain property transactions only.

database of the USEPA provides representative samples of lake water quality. Although NLA directly does not provide impairment status, they indicate the level of disturbances of lake water quality from comprehensive water quality and ecological perspectives. The 'reference lakes' in the NLA database corresponds to minimal disturbed and 'trash' lakes indicate highly disturbed lakes as the labeling suggests. The NLA uses this metric to indicate the health of waterbodies by each ecoregion. We assume the percentages of impaired lakes are the same as percentages of 'trash' lakes by ecoregions. Using fair market value estimates of lakeshore property and our baseline capitalization rate, we find the total losses in housing stock value of $2.9 billion and $2.62 billion for NLA 2012 and NLA 2017 respectively. Considering all these variations of capitalization estimates, we conclude that the range of capitalization effect is $5.40 billion to $16.90 billion, with an extreme value of $21.84 billion.

3.5 Robustness checks

3.5.1 Comparing with water clarity measure

One of our key results of this study is that housing market responses to visible and recreational use (e.g., Aesthetic) impairment are larger compared to impairment causes (e.g., Chlorine) that are less salient to housing market participants. We found a 10.39% negative effect for aesthetic use impairment where Chlorine impairment has an insignificant effect of 1.97%. We test this result further by comparing the impact of the water clarity measure (secchi disk depth) with our results. Out of 1,838 lakes in our sample, 1,222 lakes have secchi depth reading over time collected from LAGOS-NE (**LA**ke multi-scaled **G**e**OS**patial and temporal database) and the USEPA Water Quality Portal. We matched property transactions with water clarity measures using a fuzzy matching algorithm.[11] We estimate equation (1) for a smaller sample of 1,222 lakes that also has secchi depth readings and compare the results with two different models listed in equations (5) and (6).

$$\begin{aligned} ln(P_{hct}) = \beta_s \times ln\,(Secchi_{hct}) + \beta_l LakeShore_{hc} \\ + \delta ln\,(Secchi_{hct}) \times LakeShore_{hc} + \gamma X_{hct} + \sigma L_w + \eta_{ct} + \epsilon_{hct} \end{aligned} \tag{5}$$

[11] Code to clean and match water quality with property values are provided in Appendix F.

$$
\begin{aligned}
ln(P_{hct}) = \beta_i \times PostImpair_{hct} + \beta_s \times ln\ (Secchi_{hct})\ + \beta_l LakeShore_{hc} \\
+ \theta PostImpair_{hct} \times LakeShore_{hc} + \delta \\
ln\ (Secchi_{hct})\ \times LakeShore_{hc} \\
+ \mu ln(Secchi_{hct}) \times PostImpair_{hct} + \gamma X_{hct} + \sigma L_w + \eta_{ct} + \epsilon_{hct}
\end{aligned} \tag{6}
$$

In equation (5), we switch water quality variable from $PostImpair_{hct}$ to $ln\ (Secchi_{hct})$, keeping other variables are the same. In equation (6), we combine both water quality measures and include their interaction term, $ln(Secchi_{hct}) \times PostImpair_{hct}$. The results of these models are provided in Appendix E: Results Table S5. For estimates of equation (5), we find both β_s and δ significantly positive indicating water clarity increases property prices. We employed a non-nested J test proposed by Davidson and MacKinnon (1981) to find out if equation (1) or equation (5) provides a better prediction of property prices using water impairment status ($PostImpair_{hct}$) and $ln\ (Secchi_{hct})$ respectively. We find equation (5) provides better estimates. However, results combining both water quality measures in equation (6) indicate an interesting story – water clarity is not the penultimate goal in restoring water quality. The μ parameter for the interaction term, $ln(Secchi_{hct}) \times PostImpair_{hct}$ is statistically significant and negative. This result indicates that increasing 1% secchi disk depth while the waterbody becomes impaired has a 2.03% negative consequence in terms of property values. Recall our example of mussels in a lake eating up all the debris and making the lake clean, but at the same time, the lake becomes impaired due to the presence of invasive species. In such cases, our results can be useful for policy communication for aquatic invasive species control.

3.5.2 Limiting to time-invariant impairment status

Water quality literature argues that the water quality changing process is very slow. In addition, the ambient quality of the lake gives a signal of impairment well before it is listed as impaired. We formulate Tang, Heintzelman and Holsen (2018) or Papenfus (2019) type estimation where we estimate a variation of our baseline model shown in equation (7) by taking time invariant impairment status instead of post impairment.[12]

$$
\begin{aligned}
ln(P_{hct}) = \beta_i \times Impair_{hc} + \beta_l LakeShore_{hc} + \theta Impair_{hc} \times LakeShore_{hc} \\
+ \gamma X_{hct} + \sigma L_w + \eta_{ct} + \epsilon_{hct}
\end{aligned} \tag{7}
$$

[12] Notice that in equation (7), we do not have subscript t as $Impair_{hc}$ is time invariant in this equation.

The results from estimation of equation (7) show that the impact of impairment is higher (6.6%) compared to our baseline model (4.47%), which partially explains why Papenfus (2019) found a higher estimate of 11%. We used the J test to find out if time-invariant impairment status ($Impair_{hc}$) or time varying impairment status ($PostImpair_{hct}$) is a better predictor of property prices. We find that our baseline model in equation (1) ($PostImpair_{hct}$) better predicts the property prices compared to equation (7) ($Impair_{hc}$).

However, using impairment status, without considering when the waterbody is impaired requires assuming that water impairment status does not change over time.[13] We estimate equation (8) by limiting to cross-sectional data of a particular year so that the impairment status does not change. Notice that in equation (8), we dropped the time dimension and switch to census tract fixed effect (ϕ_c) from census tract by year fixed effect (η_{ct}).

$$\begin{aligned} ln(P_{hc}) = \beta_i \times Impair_{hc} + \beta_l LakeShore_{hc} + \theta Impair_{hc} \times LakeShore_{hc} + \gamma X_{hc} \\ + \sigma L_w + \phi_c + \epsilon_{hc} \end{aligned} \tag{8}$$

We estimate equation (8) for the years 2010, and 2016 where we have the most data for water impairment status.[14] Results for these estimations are shown in Appendix E: Results Table S6. We find a significant effect for 2016 data which is similar to our baseline model, but an insignificant effect for 2010 data. Again, these results explain higher estimates from Papenfus (2019) with a little caveat for aesthetic amenity impairment where we also find a large effect of 10.39%.

3.5.3 Using binary lake size

In our main model, we used lake size to control the effect of lake size on our estimation. We used lake size as a continuous variable and our main result shows that there is a 0.08% positive effect of 1km bigger lake, *ceteris paribus*. For robustness, we reestimate equation (1) by subsetting data into two distinct groups – large lakes and small lakes. If a lake is bigger than the median in size

[13] Water quality might change over time, but the change needs to be small enough such that it does not cross the threshold of water quality standards leading to listing the lake as impaired.

[14] We also have a significant number of impairment listing for 2018, but our property data ends at 2019, Q4 put limits on number of *PostImpair* transactions. Hence, we did not estimate equation (8) for 2018.

within our sample data, we define the lake as large. We also estimate equation (9) where we drop the lake size continuous variable.

$$
\begin{aligned}
ln(P_{hct}) = \beta_i \times PostImpair_{hct} + \beta_l LakeShore_{hc} \\
+ \theta PostImpair_{hct} \times LakeShore_{hc} + \gamma X_{hct} + \eta_{ct} + \epsilon_{hct}
\end{aligned} \tag{9}
$$

We found that large lakes (9.2%) impairment status has more negative effects on housing market compared to small lakes (1.26%). A large lake usually comes with more amenities compared to small lakes and hence impairment also has a higher negative effect. The results of lake size are provided in Appendix E: Results Table S7.

3.5.4 Relevant use impairment

In our baseline model, we restrict to relevant use impairments. Here we test this assumption and estimate equation (1) with non-relevant use impairments. Non-relevant use impairment includes ‘Agricultural’, ‘Aquatic life’, ‘Fishery-other’, and ‘Other’. We argue that those categories of designated use impairments do not directly affect house purchasing and selling preferences. We test this by estimating equation (1) with non-relevant use impairment data. We find an insignificant effect for non-relevant use impairments. Hence removing them from our analysis is justified both logically and empirically. The results of the model are shown in Appendix E: Result Table S8.

3.5.5 Results for reporting agencies that report impairment status by watershed

One of our identification strategies is to remove reporting agencies that report waterbody impairment by watershed because it does not reflect the condition of each waterbody in that watershed. Reporting agencies usually create their TMDL plan by watershed to control pollution, especially point source pollution. One might argue that the groundwater in a watershed is similar in quality, but surface water quality differs within a watershed. In addition, housing market participants respond to more local water quality than water quality in a watershed. Removing those reporting agencies from our baseline is a practical choice, but we conduct a robustness check regardless to see if there is any difference if we include those reporting agencies. Appendix E: Results Table S9 shows results for our estimation of equation (1) by two sets of data – only using data for five reporting agencies that report by watershed, using all the data together. We find significant results for both groups. However, we still do not include those reporting agencies’ data in our baseline model due to the aforementioned reasons.

3.5.6 Other robustness checks

We check the robustness of limiting our sample to different distance bins for control properties. In our baseline model, we use 150-1500m buffer for control properties. We found robust results for all the models by limiting data to 500m and extending up to 2000m in 250m increase bins. Results are provided in Appendix E: Results Table S10 and Figure S9. In our main model, we excluded transactions where the property is closer to a river than a lake. We checked the robustness of different cutoffs for distances from river and found estimates are similar (Appendix E: Results Table S11). We also check the robustness of inflation adjustment factors (Appendix E: Results Table S12).

4 DISCUSSION AND CONCLUSION

One of the key contributions of our study is broad spatial coverage using previously untapped nationwide water impairment and property price databases. We found a negative significant effect of waterbody impairment on lakeshore property values, leading to an estimated 4.47% decrease in property sales prices within 150m of a lake as the result of its impairment. Our results generally conform with hedonic models estimating water quality deterioration with property prices. There are only three studies that directly used water impairment data at the regional level and estimate the effect on property value. Cho, Roberts and Kim (2011) found a proximity benefit for the unimpaired part of waterbody. Papenfus (2019) found a decrease of 11% while Tang, Heintzelman and Holsen (2018) estimated an effect of 6-7% depreciation in property value. Although our results of 4.47% generally fit well within previous literature, it is slightly smaller compared to Papenfus (2019) and Tang, Heintzelman and Holsen (2018). We attribute this difference to the improved methodology by tracking impairment status over time. Our implicit price of $20,455 for lakeshore properties is also smaller compared to Papenfus (2019)[15]. Papenfus (2019) found an annualized implicit price of $1,942 which translates to $38,480.[16] We argue that the higher estimate from Papenfus (2019) is driven by higher effect size and highly valued lakeshore homes in Washington state.

[15] Cho, Roberts and Kim (2011) and Tang, Heintzelman and Holsen (2018) do not provide any implicit price estimates.

[16] Papenfus (2019) multiplied by 0.05 to get annualized implicit price.

Our study is important in providing an overall assessment of capitalization effect of lake in the contiguous United States. In the absence of national benefit of lake impairment, a policymaker or an policy analyst at the UPSEPA relies on local or regional level studies that are not broadly applicable across space. In this case, policy analysts use a techinique known as “benefits transfer”. Benefit transfer combines results from existing local or regional studies and “transfer” to regions where studies have not been undertaken to develop national benefits estimates (Corona et al., 2020; Newbold et al., 2018). However, the validity of benefits transfer can be compromised by the dearth of the empirical literature, its limited spatial representation, and differences in regional preferences and lake characteristics across the US (Guignet et al., 2020; Johnston et al., 2017; Moeltner et al., 2019; Rolfe et al., 2015).

We also found significant heterogeneity of estimates based on designated uses impairment and causes of impairments. Results show that home buyers have significant responses to aesthetic and recreational use impairments. However, we argue that an aesthetically pleasing lake does not necessarily reflect a healthy lake. We substantiate this finding by modeling water impairment and water clarity measures together. Our results show that clear water with impairment still has a negative capitalization effect on property values. Recall our example of the presence of mussels in a waterbody can make the waterbody clear but at the same time it is impaired due to the presence of invasive species. Our result can be a strong driver in formulating water quality policies, invasive species control policies in particular.

One of the main criticisms of hedonic studies is their failure to provide welfare estimates that can lead to policy implications. None of the water impairment studies in the literature provide any welfare estimates. We find nationwide \$5.4 to \$16.9 billion property value reduction due to impairment of waterbody. In order to interpret capitalization effects as welfare measures, the gradient of the hedonic price function needs to be the same across the study area (Banzhaf, 2021; Kuminoff & Pope, 2014). Given our study area is the continental US and we did not allow our estimate to vary by year and by spatial unit, we interpret our results as total capitalization loss. Despite this limitation, our losses in capitalization measures can be used to guide water policy at the national level by partially justifying the cost of improving water quality.

The capitalization effect we found is a lower bound as it only relates to property prices. There are other implications of waterbody impairment that we do not explicitly capture through a hedonic model. Downing *et al.* (2021) pointed out that there exists a global level effect of deteriorating water quality through emissions of methane into the atmosphere. The climate effect of water quality deterioration is 10 times more compared to the local estimate for recreation in Lake Erie (Downing et al., 2021). Compared to air quality studies, the estimate we found is very small, which is often the case for water quality studies overall (Keiser, Kling, & Shapiro, 2019). Generally, air quality studies derive estimates through the value of statistical life and morbidity, which corresponds to higher estimates compared to property prices. Although water quality has secondary health consequences from pollutants (e.g., PFAS, mercury), the average citizen in the US does not necessarily directly get exposed to subpar water quality as water comes as filtered to them. Average American thinks quality water resources are abundant and spatially extensive. Contrary to the subjective perceptions of the general population, the reporting agencies are assessing new waterbodies over time, and the number of impaired waterbodies is continuously increasing. In addition, the process of waterbody restoration requires time while so many human-induced factors contribute to water quality deterioration.

ACKNOWLEDGEMENT

The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper. URL: http://www.msi.umn.edu. We thank Zillow, Inc. for having made ZTRAX available free of charge for U.S. academic, nonprofit, and governmental researchers. Christoph Nolte acknowledges support from the Department of Earth & Environment at Boston University, the Junior Faculty Fellows program of Boston University's Hariri Institute for Computing and Computational Science, and The Nature Conservancy. We acknowledge the useful feedback of different conferences' participants. We acknowledge the limnological and hydrochemistry insights of Chris Filstrup, Bridget Ulrich, and Chanlan Chun, with gratitude. We also acknowledge time to time feedback from Catherine Kling, Daniel Phaneuf, Kevin Boyle, and Dave Keiser. The results and interpretations are those of the authors and do not reflect the position or policies of the University of Minnesota or the Boston University, or the Zillow Group.

REFERENCES

Banzhaf, H. S. (2021). Difference-in-Differences Hedonics. *Journal of Political Economy, 129*(8), 2385–2414. https://doi.org/10.1086/714442

Bishop, K. C., Kuminoff, N. V., Banzhaf, H. S., Boyle, K. J., Gravenitz, K. von, Pope, J. C., … Timmins, C. D. (2020). Best Practices for Using Hedonic Property Value Models to Measure Willingness to Pay for Environmental Quality. *Https://Doi.Org/10.1093/Reep/Reaa001*, *14*(2), 260–281. https://doi.org/10.1093/REEP/REAA001

Cho, S. H., Roberts, R. K., & Kim, S. G. (2011). Negative externalities on property values resulting from water impairment: The case of the Pigeon River Watershed. *Ecological Economics*, *70*(12), 2390–2399. https://doi.org/10.1016/j.ecolecon.2011.07.021

Davidson, R., & MacKinnon, J. G. (1981). Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Econometrica, 49*(3), 781. https://doi.org/10.2307/1911522

Corona, J., Doley, T., Griffiths, C., Massey, M., Moore, C., Muela, S., Rashleigh, B., Wheeler, W., Whitlock, S.D., Hewitt, J., 2020. An Integrated Assessment Model for Valuing Water Quality Changes in the United States. Land Econ. 96, 478–492. https://doi.org/10.3368/WPLE.96.4.478

Downing, J. A., Polasky, S., Olmstead, S. M., & Newbold, S. C. (2021). Protecting local water quality has global benefits. *Nature Communications 2021 12:1*, *12*(1), 1–6. https://doi.org/10.1038/s41467-021-22836-3

Federal Housing Finance Agency, 2021. House Price Index datasets Available at https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_PO_monthly_hist.xls.

Gao, X., Song, R., & Timmins, C. D. (2021). The Role of Information in the Rosen-Roback Framework. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3866100

Gibbs, J. P., Halstead, J. M., Boyle, K. J., & Huang, J.-C. (2002). An Hedonic Analysis of the Effects of Lake Water Clarity on New Hampshire Lakefront Properties. *Agricultural and Resource Economics Review, 31*(1), 39–46. https://doi.org/10.1017/s1068280500003464

Gindelsky, M., Moulton, J. G., Wentland, S. A., Abraham, K., Diewert, E., Fixler, D., … Verbrugge, R. (2019). Valuing Housing Services in the Era of Big Data: A User Cost Approach Leveraging Zillow Microdata. *Big Data for 21st Century Economic Statistics*. Retrieved from http://www.zillow.com/ztrax.

Guignet, D., Heberling, M.T., Papenfus, M., Griot, O., Holland, B., 2020. Property values, water quality, and benefit transfer: A nationwide meta-analysis, Appalachian State University, Department of Economics Working Paper.

Guignet, D., & Nolte, C. (2021). *Hazardous Waste and Home Values: An Analysis of Treatment and Disposal Sites in the U.S.* (21–07).

Horsch, E. J., & Lewis, D. J. (2009). The Effects of Aquatic Invasive Species on Property Values: Evidence from a Quasi-Experiment. *Land Economics*, *85*(3), 391–409. https://doi.org/10.3368/LE.85.3.391

Johnston, R.J., Besedin, E.Y., Stapler, R., 2017. Enhanced Geospatial Validity for Meta-analysis and Environmental Benefit Transfer: An Application to Water Quality Improvements. Environ. Resour. Econ. 68, 343–375. https://doi.org/10.1007/S10640-016-0021-7/TABLES/7

Kashian, R., Eiswerth, M. E., & Skidmore, M. (2006). Lake rehabilitation and the value of shoreline real estate: Evidence from Delavan, Wisconsin. *Review of Regional Studies*, *36*(2), 221--238.

Keiser, D. A., Kling, C. L., & Shapiro, J. S. (2019). The low but uncertain measured benefits of US water quality policy. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(12), 5262–5269. https://doi.org/10.1073/PNAS.1802870115/-/DCSUPPLEMENTAL

Kuminoff, N. V., & Pope, J. C. (2014). Do "capitalization effect" for public goods reveal the public's willingness to pay? *International Economic Review*, *55*(4), 1227–1250. https://doi.org/10.1111/IERE.12088

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157. https://doi.org/10.1086/259131

Liao, F., Wilhelm, F., & Solomon, M. (2016). The Effects of Ambient Water Quality and Eurasian Watermilfoil on Lakefront Property Values in the Coeur d'Alene Area of Northern Idaho, USA. *Sustainability*, *8*(1), 44. https://doi.org/10.3390/su8010044

Liu, T., Opaluch, J. J., & Uchida, E. (2017). The impact of water quality in Narragansett Bay on housing prices. *Water Resources Research*, *53*(8), 6454–6471. https://doi.org/10.1002/2016WR019606

Microsoft. (2018). U.S. building footprints. (2018). Retrieved February 2, 2020, from https://github.com/microsoft/USBuildingFootprints.

Moeltner, K., Balukas, J.A., Besedin, E., Holland, B., 2019. Waters of the United States: Upgrading wetland valuation via benefit transfer. Ecol. Econ. 164, 106336. https://doi.org/10.1016/J.ECOLECON.2019.05.016

Moore, M. R., Doubek, J. P., Xu, H., & Cardinale, B. J. (2020). Hedonic Price Estimates of Lake Water Quality: Valued Attribute, Instrumental Variables, and Ecological-Economic Benefits. *Ecological Economics*, *176*, 106692. https://doi.org/10.1016/j.ecolecon.2020.106692

Netusil, N. R., Kincaid, M., & Chang, H. (2014). Valuing water quality in urban watersheds: A comparative analysis of Johnson Creek, Oregon, and Burnt Bridge Creek, Washington. *Water Resources Research*, *50*(5), 4254–4268. https://doi.org/10.1002/2013WR014546

Newbold, S., David Simpson, R., Matthew Massey, D., Heberling, M.T., Wheeler, W., Corona, J., Hewitt, J., 2018. Benefit Transfer Challenges: Perspectives from U.S. Practitioners. Environ. Resour. Econ. 69, 467–481. https://doi.org/10.1007/S10640-017-0207-7/TABLES/1

Nolte, C. (2020). High-resolution land value maps reveal underestimation of conservation costs in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(47), 29577–29583. https://doi.org/10.1073/PNAS.2012865117/-/DCSUPPLEMENTAL

Nolte, C., Boyle, K. J., Chaudhry, A., Clapp, C., Guignet, D., Hennighausen, H., … Uhl, J. H. (2021). Studying the impacts of environmental amenities and hazards with nationwide property data: best data practices for interpretable and reproducible analyses. *SSRN*.

Papenfus, M. (2019). Do housing prices reflect water quality impairments? Evidence from the Puget Sound. *Water Resources and Economics*, *27*(December 2018), 1–10. https://doi.org/10.1016/j.wre.2018.12.001

Poor, P. J., Boyle, K. J., Taylor, L. O., & Bouchard, R. (2001). Objective versus Subjective Measures of Water Clarity in Hedonic Property Value Models. *Land Economics*, *77*(4), 482–493.

Poor, P. J., Pessagno, K. L., & Paul, R. W. (2007). Exploring the hedonic value of ambient water quality: A local watershed-based study. *Ecological Economics*, *60*(4), 797–806. https://doi.org/10.1016/j.ecolecon.2006.02.013

Pope, J. C. (2008). Buyer information and the hedonic: The impact of a seller disclosure on the implicit price for airport noise. *Journal of Urban Economics*, *63*(2), 498–516. https://doi.org/10.1016/j.jue.2007.03.003

Rolfe, J., Windle, J., Johnston, R.J., 2015. Applying Benefit Transfer with Limited Data: Unit Value Transfers in Practice, in: The Economics of Non-Market Goods and Resources. Springer, Dordrecht, pp. 141–162. https://doi.org/10.1007/978-94-017-9930-0_8

Sprague, L. A., Oelsner, G. P., & Argue, D. M. (2017). Challenges with secondary use of multi-source water-quality data in the United States. *Water Research*, *110*, 252–261. https://doi.org/10.1016/J.WATRES.2016.12.024

Tang, C., Heintzelman, M. D., & Holsen, T. M. (2018). Mercury pollution, information, and property values. *Journal of Environmental Economics and Management*, *92*(April 2016), 418–432. https://doi.org/10.1016/j.jeem.2018.10.009

U.S. Geological Survey. (2017a). 3D Elevation Program (3DEP). Retrieved from https://www.usgs.gov/core-science-systems/ngp/3dep/about-3dep-products-services

U.S. Geological Survey. (2017b). USGS National Hydrography Dataset (NHD) Downloadable Data Collection - National Geospatial Data Asset (NGDA) National Hydrography Dataset (NHD). Retrieved from https://www.sciencebase.gov/catalog/item/4f5545cce4b018de15819ca9

United Nation Environmental Program (UNEP). (2021). *Global status of Indicator 6.3.2 Proportion of bodies of water with good ambient water quality (2017-2020)*. Retrieved from https://sdg6data.org/indicator/6.3.2

United States Environmental Protection Agency (USEPA). *2017 National Water Quality Inventory Report to Congress*. , (2017).

United States Environmental Protection Agency (USEPA). (2021). The Assessment, Total Maximum Daily Load (TMDL) Tracking and Implementation System (ATTAINS). Retrieved from https://www.epa.gov/waterdata/get-data-access-public-attains-data

Walsh, P., Griffiths, C., Guignet, D., & Klemick, H. (2017). Modeling the Property Price Impact of Water Quality in 14 Chesapeake Bay Counties. *Ecological Economics*, *135*, 103–113. https://doi.org/10.1016/j.ecolecon.2016.12.014

Weng, W., Boyle, K. J., Farrell, K. J., Carey, C. C., Cobourn, K. M., Dugan, H. A., … Weathers, K. C. (2020). Coupling Natural and Human Models in the Context of a Lake Ecosystem: Lake Mendota, Wisconsin, USA. *Ecological Economics*, *169*, 106556. https://doi.org/10.1016/J.ECOLECON.2019.106556

Wolf, D., & Kemp, T. (2021). Convergent Validity of Satellite and Secchi Disk Measures of Water Clarity in Hedonic Models. *Land Economics*, *97*(1), 39–58. https://doi.org/10.3368/WPLE.97.1.050819-0062R1

Wolf, D., & Klaiber, H. A. (2017). Bloom and bust: Toxic algae's impact on nearby property values. *Ecological Economics*, *135*, 209–221. https://doi.org/10.1016/J.ECOLECON.2016.12.007

Zhang, C., & Boyle, K. J. (2010). The effect of an aquatic invasive species (Eurasian watermilfoil) on lakefront property values. *Ecological Economics*, *70*(2), 394–404. https://doi.org/10.1016/J.ECOLECON.2010.09.011

Zhang, J., Phaneuf, D., & Schaeffer, B. (2021). Property values and cyanobacterial algal blooms: Evidence from satellite monitoring of inland lakes. *Ecological Economics*, 199, 107481.

Zillow. (2019). Zillow Transaction and Assessor Dataset, 2019 Q4. Retrieved from http://www.zillow.com/ztrax

Supplementary information for

Nationwide Lake Impairment and Property Value

Saleh Mamun[1,2,3,*], Stephen Polasky[1,2,4], and Christoph Nolte[5,6]

July 15, 2022

[1] *Department of Applied Economics, University of Minnesota, St. Paul, MN*
[2] *The Natural Capital Project, University of Minnesota, St. Paul, MN*
[3] *Natural Resources Research Institute, University of Minnesota – Duluth, Duluth, MN*
[4] *Department of Ecology, Evolution & Behavior, University of Minnesota, St. Paul, MN*
[5] *Department of Earth & Environment, Boston University, Boston, MA*
[6] *Faculty of Computing & Data Sciences, Boston University, Boston, MA*

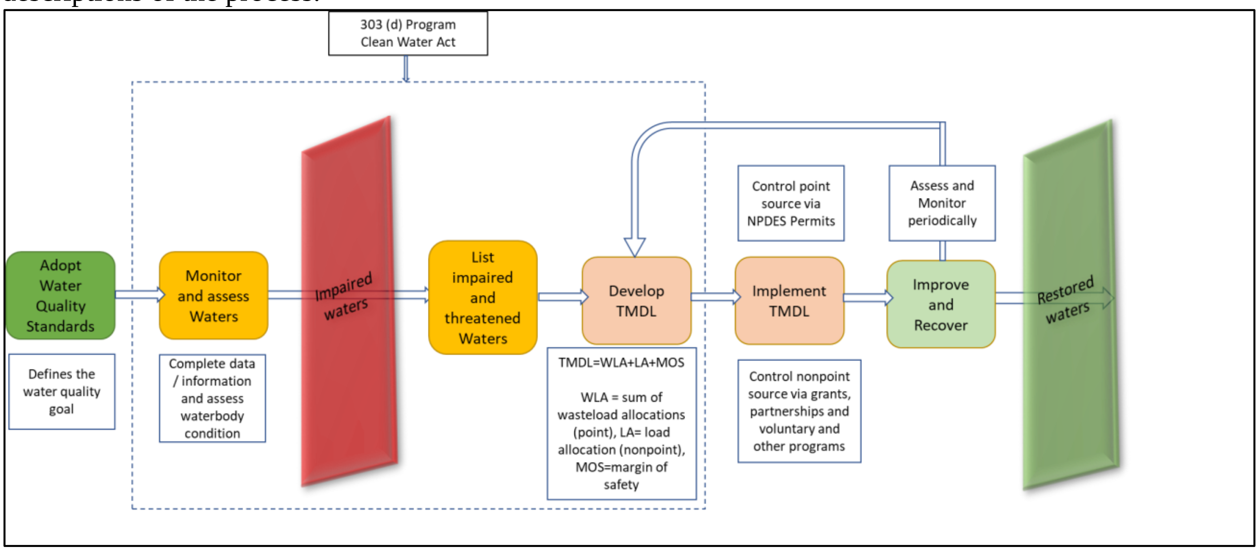
**corresponding author, email: salmamun@d.umn.edu*

This document contains:

APPENDIX A: IMPAIRED WATER RESTORATION PROCESS

As a part of the Clean Water Act 1972 (CWA), the states, territories, and tribal authorities are required to take actions with the help of USEPA to improve the water quality such that it meets its designated uses. This process is called the impaired water restoration process is shown in Figure S1.

Figure S1: The impaired water restoration process steps. The figure is recreated based on USEPA descriptions of the process.



Adapt water quality standards:

Reporting agencies first adapt water quality standards for different water quality parameters and pollutants as per Section 131. The water quality standards are thresholds of magnitude, duration, and frequency for each of the water quality parameters and pollutants. As an overseeing body, USEPA review, approve or disapprove, and promulgate the water quality standards set by the reporting agencies.

Monitor and assess waters:

Water monitoring is required as per several Sections of CWA to evaluate the effectiveness of different water quality improvement programs. The reporting agencies are responsible for designing the monitoring project, collecting and managing, and interpreting data, and conveying results. The reporting agencies do not monitor all waters but use target monitoring and statistical probability sampling to select which waterbodies to monitor. Instead of monitoring by waterbody, the reporting agencies select assessment units. Assessment units are the area of interest that can be the whole waterbody or part of the waterbody or a group of waterbodies. Nationally 32% of rivers and streams length and 44% of lake area are monitored against a variety of water quality parameters.

Listing impaired and threatened waters

The monitored assessment units are evaluated using state adapted and USEPA approved water quality standards. Based on the expert evaluation, assessment units are listed as five distinct categories:

Category 1: Assessment unit supports all designated uses and there are no uses threatened. Reporting agencies can define a designated use as threatened when the water quality is currently meeting the standard, but it is expected to not attain water quality standards by the next reporting cycle.

Category 2: Assessment unit supports some of the designated uses but there is limited information to determine if the waterbody support other designated uses that the assessment unit is supposed to support.

Category 3: There is insufficient available data and/or information to make a use support determination.

Category 4: Based on available information, the assessment unit is impaired, but it does not require a TMDL because a TMDL is previously completed (4A), or water quality standards can be attained within a reasonable period (4B), or impairment is not caused by a pollutant (4C).

Category 5: At least one designated use is not being supported by the assessment unit, and a TMDL is needed.

Category 5-alt: The assessment unit is impaired and a TMDL is needed, but there is an alternative plan to achieve water quality standards that have been associated with the assessment unit.

Develop TMDL/clean-up plan

Although TMDL development is not required by the CWA, most of the sites that are impaired due to pollutants need a TMDL plan to clean up. Some assessment units without pollutants (e.g., biological habitat alteration) can have an alternative clean-up plan approved by the USEPA. A

TMDL consists of a maximum allowance for pollution to the assessment unit or watershed as a whole by point sources, non-point sources, and with an allowance of margin of safety.

Implement TMDL plan

The next phase is to implement the TMDL plan which requires monitoring and regulations. Point sources of pollution (such as industry effluent) can be regulated through National Pollutant Discharge Elimination System (NPDES) permits issued by the USEPA. Non-point sources (such as nitrogen from farmlands) can be regulated through grants, partnerships, and voluntary programs.

Improve and recover

Continuous monitoring of the assessment unit (waterbody as a whole) is required for the successful implementation of the clean-up plan. Based on updated monitoring and assessment TMDL plan and clean-up plan can be revised such that it efficiently achieves water quality standards. Finally, if the assessment unit meets water quality standards, then the restoration is complete and the assessment unit is delisted from the impairment list.

APPENDIX B: HEDONIC ANALYSIS PUBLICATIONS OF WATER QUALITY

We have searched methodically Google Scholar for relevant literature. Table S1 summarizes search terms and corresponding search results. Altogether, google scholar indexed 232 documents including duplicates. After careful consideration (relevance, restriction, and availability), we read all 21 water quality hedonics journal articles. A summary of these studies is presented in Table S2 (next page).

Table S1: Search terms and corresponding results in google scholar.

Search term	**Results (number of documents)**
allintitle: "hedonic" water OR lake OR stream OR ocean	105
allintitle: "property value" water OR lake OR stream OR ocean	9
allintitle: "property values" water OR lake OR stream OR ocean	53
allintitle: "real estate" water OR lake OR stream OR ocean	43
allintitle: "property price" water OR lake OR stream OR ocean	7
allintitle: "property prices" water OR lake OR stream OR ocean	6
allintitle: "housing market" water OR lake OR stream OR ocean	9

Table S2: Summary of water quality hedonic literature

Year	**Authors**	**Location**	**Method**	**Data**	**Environmental Measures**	**Environmental data source**	**Main Results**
1974	Barrager (1974)	Willamette River (Clackamas County, Oregon)	hedonic analysis	98 single-family residences within 4000 feet of the shoreline which sold between 1969 and 1971	Change of water quality due to pollution abatement project. No objective parameters	NA	The percentage increase in property value due to pollution abatement is expressed as a function of the distance between properties and the riverbank.
1983	Willis and Foster (1983)	1500 feet from Housatonic River, MA and Winoski River, Montpelier, VT	Multiple hedonic models	81 or 40 properties from 1962-1980	perceived water quality from a survey in the scale of drinkable to not boatable	authors own data	Hedonic results were not supportive; surveys suggested homeowners had little awareness of water quality.
1995	Lansford and Jones (1995)	2.5 miles from Travis Lake, Texas	standard hedonic model	609 SFH transactions 1988-1990 within 2.5 miles of the lake.	Water level at the time of sale: lake levels: mean, +/- 1std dev, +/- 2 std. dev. distance from lake to prop in feet; LDIST distance from lake to prop in feet up to 4,000 feet	Lower Colorado River Authority.	At the waterfront 1 foot decline in water level is associated with $56 decrease in sales price. At 1500 feet from waterfront 1 foot decline in water level is associated with $12 decrease in sales price. At 3000 feet from waterfront 1-foot decline in water level is associated with a $5.41 decrease in sales price.
2001	Poor et al (2001)	Freshwater lakes and ponds in Maine	hedonic models	348 lakefront properties, 1990-1995	water clarity (subjective and objective)	secchi depth from the Maine Department of Environmental Protect. Subjective clarity is from survey results.	The objective measures of water clarity better describe the amenity levels that property owners were considering at the time of purchase than retrospective self-reports of the amenity level.
2002	Gibbs et al (2002)	69 lakes in New Hampshire	standard hedonic model	Lakefront housing sales data,1990 - 1995	secchi disk readings	New Hampshire Department of Environmental Services (DES).	1m decrease in water clarity is associated with 0.9% to over 6% decreases in sales price

Year	Authors	Location	Method	Data	Environmental Measures	Environmental data source	Main Results
2003	Loomis and Feldman (2003)	Lake Almanor, California	linear hedonic property regression	964 SFH	lake level	data collected for each day (1987-2001)	Property prices were negatively and significantly related to the number of linear feet of exposed lake shoreline. Each additional one foot of exposed shoreline reduces the property price by \$108–\$119. A view of the lake added nearly \$31,000 to house prices, while lakefront properties sold for \$209,000 more than non-lake front properties
2004	Parsons and Noailly (2004)	Delaware’s Ocean Beaches	standard hedonic model	266 transactions from 1991-1992	Sea level rise scenarios and projected property damage		Projected property losses over 2000-2050 are \$291 million in 2000 dollars.
2006	Kashian et al (2006)	Three lakes in Walworth County, WI	hedonic model	314 SFH assessed value on 1987, 1995, and 2003 (panel)	Secchi disk depth	Wisconsin DNR and Lauderdale Lakes Management District	Effect of water clarity was significant (positive). A one-foot increase in clarity was associated with a \$5207 increase in price of an average property.
2007	Poor et al (2007)	St. Mary's River watershed, Maryland	hedonic model	1377 residential property sales from June 1, 1999 and May 31, 2003	dissolved inorganic nitrogen (DIN) and total suspended solids (TSS)	Biology Department at St. Mary's College of Maryland	Ambient water quality within a small local watershed can significantly influence residential property values regardless of whether they are waterfront properties.
2011	Cho et al (2011)	Pigeon River watershed, NC and TN	spatial hedonic model	595 and 497 SFH from 2001-2004	impairment status of the watershed	STORET	North Carolina residents residing in sub-watersheds with impaired portions of the Pigeon River, who experience economic benefit from the paper mill in addition to its harmful effects on water quality, do perceive the pollution as a negative externality

Year	Authors	Location	Method	Data	Environmental Measures	Environmental data source	Main Results
2013	Bin and Czajkowski (2013)	Waterfront homes in Martin County, Florida	spatial error hedonic models with district-level fixed effect	510 SFH transactions from 2000-2004	Objective: temperature, pH, water clarity, salinity, and (DO) Composite: Florida Oceanographic Society's grade index	Florida Oceanographic Society	A 1% increase in water clarity, evaluated at the mean value, associated with $36,070 increase in mean property value. A .10-point increase in pH is associated with a $7,531 increase in mean property value A 1% increase in salinity is associated with a $31,938 increase in mean property values.
2015	Tuttle and Heintzelman (2015)	52 lakes within Adirondack Park, NY	spatial lag model with waterbody-level fixed effects	12,001 SFH transactions from 2001-2009	Eurasian watermilfoil, pH, birds distance to the nearest lake waterfront indicator variable	Adirondack Lakes Survey National Lake Assessment Program	A pH level below 6.5 is associated with a 20% decrease in sales values A pH level below 6.5 is associated with a 23% decrease for waterfront parcels relative to lakes with known good pH.
2015	Dickes and Crouch (2015)	Lake Thurmond, GA		1,030 properties from 2000-2009	water level, average temperature	for the years 1998 through 2009 was provided by the USACE	Results confirm that declining Lake Thurmond water levels have an impact on real estate values within some ranges below full pool. As climate variability places increasing pressure on communities, future research would benefit from further exploration into the relationship between economic activity and changing lake levels.

Year	**Authors**	**Location**	**Method**	**Data**	**Environmental Measures**	**Environmental data source**	**Main Results**
2016	Seidel et al (2016)	1500 m buffer from Lower St Johns Basin in Florida	hedonic analysis	23,494 records period 2003 through 2013	The 'Trophic State Index' (TSI), an indicator of biomass within the water column provides one such measure. TSI(SD) = 60 – 14.41* ln(SD); SD = Secchi depth. Used Secchi depth finally.	Florida Department of Environmental Protection	Waterfront properties with the highest clarity enjoyed an increased value premium of close to 24% for river frontage, while properties with the lowest clarity saw this premium reduced to only 6% of the sales price.
2016	Liao et al (2016)	Coeur d'Alene lakefront, Idaho	hedonic analysis	614 lakefront property from Kootenai County's Assessor's office for 2010–2014	secchi depth and presence of milfoil (invasive species)	Idaho Department of Environmental Quality (IDEQ),	property values are positively associated with Secchi depth (a proxy of water quality or clarity), and negatively related to the presence of watermilfoil.
2017	Walsh et al (2017)	14 Chesapeake Bay counties, Maryland	spatial hedonic models	229,513 properties from 1996-2008	The water-column light attenuation coefficient (KD). Light attenuation can be converted to SDM based on the following statistical relationship: KD = 1.45/SDM	EPA's Chesapeake Bay Program Office (CBP)	water quality improvements in the Bay, such as those required by EPA's Total Maximum Daily Load, could yield significant benefits to waterfront and near-waterfront homeowners.
2017	Bin et al (2017)	The Northeastern portion of Martin County, FL	hedonic model	2243 waterfront properties 2001-2010	Calculated water grade using: water temp, PH, visibility (secchi depth/water depth), salinity, dissolved oxygen	Florida Oceanographic Society (FOS 2011)	water quality improvement is associated with higher property values

Year	Authors	Location	Method	Data	Environmental Measures	Environmental data source	Main Results
2018	Singh et al (2018)	Salton sea lake, California	spatial autoregressive model with autoregressive disturbances (SARAR)	1,140 sales 2009-mid2013	water quality: Secchi disk readings; and air quality: PM10	water clarity: Secchi disk readings from US Bureau of Reclamation (USGS 2013). PM10 data from the California Air Resource Board (CARB)	Secchi disk reading has positive relation while PM10 has negative relation with housing prices
2018	Klemick et al (2018)	Meta-analysis: 14 counties on Chesapeake Bay, Maryland; benefit transfer: DC, Delaware, Virginia, four additional Maryland counties	meta-analysis	Meta-analysis of 70 estimates of water clarity (1996-2008)	Log of water-column light attenuation coefficient (KD)	different studies	Importance of water clarity increased with proximity to the bay. Ten percent improvement in one-year light attenuation led to a statistically significant property value increase of 0.6% for waterfront properties, and 0.1% for non-bayfront homes extending out to 500 m. Aggregate near-waterfront property values were projected to increase by $400–$700 million in response to water clarity improvements.
2020	Liu (2020)	1-mile buffer of White Bear Lake	spatial autoregressive model with autoregressive disturbances (SARAR)	4,611 properties sales from 1995 to 2015 from MetroGIS Regional Parcel Dataset	water level	Minnesota Department of Natural Resources	When the lake elevation drops to 6 feet below the ordinary high water level, the marginal implicit price of an additional foot of water loss is estimated to be $52,473.11 for an average property in the White Bear Lake sample

Year	Authors	Location	Method	Data	Environmental Measures	Environmental data source	Main Results
2020	Moore et al (2020)	32 states	Hedonic regression	1462 house data around 113 lakes in 2007 (cross-section)	Secchi disk depth; temperature; P concentration, N concentration	2007 National Lakes Assessment (NLA)	The estimated water-clarity effect shows that a one-tenth of a meter change in water clarity leads to a one-percent change in housing price or elasticity of 0.20 at mean clarity.

Appendix B: Reference

Barrager, S. M. (1974). The impact of water resource quality changes on surrounding property values. *Journal of the American Water Resources Association*, *10*(4), 759–765. https://doi.org/10.1111/j.1752-1688.1974.tb05636.x

Bin, O., Czajkowski, · Jeffrey, Li, J., Villarini, G., & Czajkowski, J. (2017). Housing Market Fluctuations and the Implicit Price of Water Quality: Empirical Evidence from a South Florida Housing Market. *Environ Resource Econ*, *68*, 319–341. https://doi.org/10.1007/s10640-016-0020-8

Bin, O., & Czajkowski, J. (2013). The impact of technical and non-technical measures of water quality on coastal waterfront property values in South Florida. *Marine Resource Economics*, *28*(1), 43–63. https://doi.org/10.5950/0738-1360-28.1.43

Cho, S. H., Roberts, R. K., & Kim, S. G. (2011). Negative externalities on property values resulting from water impairment: The case of the Pigeon River Watershed. *Ecological Economics*, *70*(12), 2390–2399. https://doi.org/10.1016/j.ecolecon.2011.07.021

Dickes, Lori; Crouch, E. (2015). The impact of changing Lake levels on property values: A hedonic model of Lake Thurmond. *Review of Regional Studies*, *45*(3), 221--235.

Gibbs, J. P., Halstead, J. M., Boyle, K. J., & Huang, J.-C. (2002). An Hedonic Analysis of the Effects of Lake Water Clarity on New Hampshire Lakefront Properties. *Agricultural and Resource Economics Review*, *31*(1), 39–46. https://doi.org/10.1017/s1068280500003464

Kashian, Russ; Eiswerth, Mark E; Skidmore, M. (2006). Lake rehabilitation and the value of shoreline real estate: Evidence from Delavan, Wisconsin. *Review of Regional Studies*, *36*(2), 221--238.

Klemick, H., Griffiths, C., Guignet, D., & Walsh, P. (2018). Improving Water Quality in an Iconic Estuary: An Internal Meta-analysis of Property Value Impacts Around the Chesapeake Bay. *Environmental and Resource Economics*, *69*(2), 265–292. https://doi.org/10.1007/s10640-016-0078-3

Lansford, N. H., & Jones, L. L. (1995). Recreational and Aesthetic Value of Water Using Hedonic Price Analysis. In *Journal of Agricultural and Resource Economics* (Vol. 20).

Liao, F., Wilhelm, F., & Solomon, M. (2016). The Effects of Ambient Water Quality and Eurasian Watermilfoil on Lakefront Property Values in the Coeur d'Alene Area of Northern Idaho, USA. *Sustainability*, *8*(1), 44. https://doi.org/10.3390/su8010044

Liu, W. (2020). Valuation of Water Level: A Spatial Hedonic Analysis on Lakeshore Properties. *Journal of Agricultural and Resource Economics*, *45*(1), 20--37.

Loomis, J., & Feldman, M. (2003). Estimating the benefits of maintaining adequate lake levels to homeowners using the hedonic property method. *Water Resources Research*, *39*(9). https://doi.org/10.1029/2002WR001799

Moore, M. R., Doubek, J. P., Xu, H., & Cardinale, B. J. (2020). Hedonic Price Estimates of Lake Water Quality: Valued Attribute, Instrumental Variables, and Ecological-Economic Benefits. *Ecological Economics*, *176*, 106692. https://doi.org/10.1016/j.ecolecon.2020.106692

Parsons, G. R., & Noailly, J. (2004). A value capture property tax for financing beach nourishment projects: An application to Delaware's ocean beaches. *Ocean and Coastal Management*, *47*(1–2), 49–61. https://doi.org/10.1016/j.ocecoaman.2004.03.003

Poor, P. J., Boyle, K. J., Taylor, L. O., & Bouchard, R. (2001). Objective versus Subjective Measures of Water Clarity in Hedonic Property Value Models. *Land Economics*, *77*(4), 482–493.

Poor, P. J., Pessagno, K. L., & Paul, R. W. (2007). Exploring the hedonic value of ambient water quality: A local watershed-based study. *Ecological Economics*, *60*(4), 797–806. https://doi.org/10.1016/j.ecolecon.2006.02.013

Seidel, V., Diamond, C., Yacobellis, P., Barker, A., & Cortez, C. (2016). The economic impact of the St. Johns river water quality on property values. *Water E-Journal*, *1*(3), 1–12. https://doi.org/10.21139/wej.2016.028

Singh, A., Saphores, J. D., & Bruckner, T. (2018). A spatial hedonic analysis of the housing market around a large, failing desert lake: the case of the Salton Sea in California. *Journal of Environmental Planning and Management*, *61*(14), 2549–2569. https://doi.org/10.1080/09640568.2017.1405799

Tuttle, C. M., & Heintzelman, M. D. (2015). A loon on every lake: A hedonic analysis of lake water quality in the Adirondacks. *Resource and Energy Economics*, *39*, 1–15. https://doi.org/10.1016/j.reseneeco.2014.11.001

Walsh, P., Griffiths, C., Guignet, D., & Klemick, H. (2017). Modeling the Property Price Impact of Water Quality in 14 Chesapeake Bay Counties. *Ecological Economics*, *135*, 103–113. https://doi.org/10.1016/j.ecolecon.2016.12.014

Willis, C. E., & Foster, J. H. (1983). The Hedonic Approach: No Panacea for Valuing Water Quality Changes. *Journal of the Northeastern Agricultural Economics Council*, *12*(1), 53–57. https://doi.org/10.1017/s0163548400003642

APPENDIX C: DATA CLEANING AND DATABASE CONSTRUCTION

Property data and control variables

In hedonic analysis, identification of the location of the property and its fair market value is critical. We used the Private-Land Conservation Evidence System (PLACES) dataset which is primarily based on digital geospatial parcel maps from publicly available and licensed data sources (namely Regrid). PLACES addresses the identification of the location of the property and its fair market value issue of Zillow. PLACES contains data for 133 million parcels in 2,738 counties in the United States. PLACES combines parcel boundary data with Zillow's Transaction and Assessment Database (ZTRAX, version: Oct 09, 2019) (Zillow, 2019). Although ZTRAX contains geographic coordinates of the property, it does not have parcel boundary maps. Using Zillow-provided coordinates without correcting for coordinate system projection, duplicates, missing data, and building locations can lead to misleading estimates of location-based variables. Furthermore, transaction prices in ZTRAX dataset often do not reflect fair market value: they can be affected by intra-family transfers, foreclosures, as well as underpriced or discounted public transactions. Including non-fair market transactions in an analysis can lead to a misrepresentative conclusion. We ensure fair market value and exact location of the property by adopting some best practices described by Nolte et al., 2021.

We improve the geographic information of properties by linking parcel boundaries with ZTRAX using the assessor's parcel number and a customized pattern matching algorithm (Nolte, 2020). We ignored the transactions if parcel subdivision and consolidation resulted in unsuccessful or partial linking of parcels with the ZTRAX dataset.

To increase the likelihood that sales prices reflect fair market values, we remove transactions that are inter-family transfer, or foreclosure, or undervalued public transactions. Inter-family transfers are identified based on Zillow’s inter-family transfer flag and a published algorithm to identify similarity in buyer and seller names (Nolte et al., 2021). Foreclosures and other document code analyses are used to flag and remove transactions that do not reflect fair market value. We also removed transactions where either seller or buyer is a public entity. We identify these public entities from seller and buyer names using a published repository of string patterns (Nolte et al., 2021). We also removed transactions if the sale price is less than $10,000 (Gindelsky et al., 2019) and the top 1 percentile of the sale price to address remaining outliers.

We calculated the lot area from digital parcel boundary maps. The National Elevation dataset is spatially joined with parcel boundary maps to extract the average slope and elevation of the parcel (U.S. Geological Survey, 2017a). We substantiate building area information from various sources as ZTRAX-provided building area attributes has many missing values. Some counties even do not report building areas. For some counties, ZTRAX only reports gross building area. If a parcel does not report building area, then we use building area from gross building area, which is a very good proxy. If both building area and gross building area are missing, we used the estimated building area from Microsoft's building footprint (Microsoft, 2018). Building age is calculated at the time of sale based on the built year or effective year. If the effective year is missing, then the built year is used. We removed transaction records that occurred before the built year or effective year.

We used the National Hydrography Dataset (NHD) to calculate the distance from a lake or river to the centroid of a parcel (U.S. Geological Survey, 2017b). Our final property dataset contains 660,707 single-family residential houses within 1500m from lakes greater than 4 ha for the years 2000-2019. We exclude the five great lakes from our analysis as great lakes play completely different recreational, ecological, and amenities roles as other inland lakes. As we are interested in lake water quality, we also omit parcels that are closer to a river than a lake. We also check the robustness of alternative definitions of this filter. We identified residential developed single-family homes if there is any building area and Zillow reported a building code of 'RR101' (single-family home).

Water impairment data

Water impairment status data comes from USEPA's Assessment, Total Maximum Daily Load (TMDL) Tracking, and Implementation System (ATTAINS) database (USEPA, 2021). The states, territories, and tribal authorities are required to assess and report[17] the condition of waterbodies in their jurisdiction every two years. Instead of monitoring and listing by waterbodies, the reporting agencies report by assessment units. An assessment unit is an area of interest that can be part of a waterbody (e.g., bay of lake, or part of lake that is within the state's jurisdiction), or whole waterbody, or multiple waterbodies (e.g., watersheds). Based on their assessment, the reporting agencies suggest if an assessment unit is impaired or not. The USEPA verifies the information and

[17] This report is called integrated report by the reporting agencies. It includes CWA section 303 (d) and section 305(b) listing of waterbody in reporting agencies' jurisdiction.

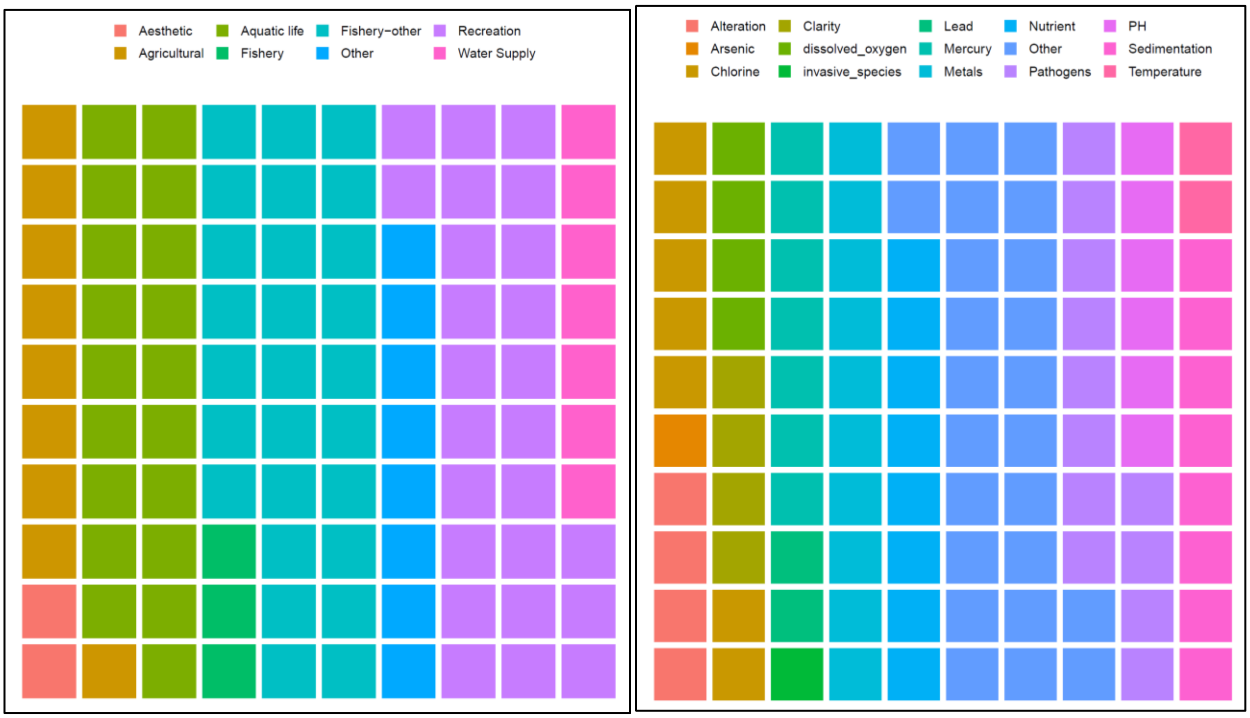
stores the data in two databases that can be accessed through Representational State Transfer Application Programming Interfaces (REST APIs):[18]. The ATTAINS geospatial database provides the latest condition of assessment units along with its geographic information, while the ATTAINS parameter database provides the condition of assessment units over time. The former contains three separate layers of polygons (n=68,088), lines (n=352,479) and points (n=3,999). Lakes, impoundments, and reservoirs are represented in all three shape forms.

Although the geospatial database lists several causes of impairment and several designated impaired uses, it does not track the impairment status of the assessment unit over time. We used the parameter REST API that provides information over time for each of the parameters assessed in each of the assessment units. Parameter REST API provides information on if the assessment unit is assessed, what are the parameters that are assessed (*parameterName*) for this assessment unit, for each parameter if the assessment unit is impaired (*parameterAttainmentCode*), for each parameter what associated use is potentially impaired (*associatedUseName*), when a waterbody is first listed as impaired (*cycleFirstListed*), and when the information is provided to and documented in ATTAINS (*cycleParm*).

Note that reporting agencies are not required to submit information in ATTAINS database under CWA, but they are required to submit integrated reports of waterbodies in their jurisdiction. This contributes to the poor structure of ATTAINS database. Nomenclature of *parameterName* and *associatedUseName* varies substantially by state and ends up with 857 different *parameterName* and 407 different *associatedUseName*. We consolidated all these differences in *parameterName* and *associatedUseName* with the help of limnologists and hydrochemists. The *parameterName* and *associatedUseName* groupings are provided in the supplementary document list and online link in Appendix F.

[18] The code and access to data can be found in anonymous repository: https://osf.io/gyd6f/?view_only=3968d945107846dcb0b198b055c7856d

Figure S2: The waffle plot in Panel A presents frequencies of different parameter impairments in ATTAINS database. The waffle plot in Panel B shows frequencies of number of designated uses impaired. Pathogens are the most frequent parameter listed as the cause of impairment. Fishery-other that includes commercial fishing is the most frequent use impairment.



Waffle plots in Figure S2 show impairment status by parameter and by designated use respectively. A careful consolidation of ATTAINS reported *parameterName* and *associatedUseName* leads us to 15 *parameterGroups* and 8 designated *useGroups*. Recreation, fishery/fish consumption, and aquatic life impairment are the most frequently impaired designated uses while pathogens, nutrients, and different metals and chemicals are the most frequently named sources of impairment. We exclude '*Fishery-other*', '*Aquatic life*', '*agricultural*', and '*other*' categories as those are least likely to be relevant information to the housing market participants.

Although reporting agencies are submitting an integrated list of waterbody impairment, the ATTAINS database has started storing data very recently. There are no integrated reports stored in the database before 2002. In addition, a lot of integrated reports are missing in the database. We compiled a report of data availability queried on April 15, 2021, in the supplementary document list and online link in Appendix F.

APPENDIX D: TREATMENT AND CONTROL IDENTIFICATION STRATEGY

The 'treatment' properties in this analysis are located within 150m of a waterbody that is listed as impaired at the time of sale. We choose 150m buffer as treatment as the average distance from waterfront properties[19] centroid to waterbody boundary in our baseline dataset is about 150m. The control properties are either within 150-1500m away from the waterbody or the waterbody is fully supporting its designated uses. Challenges and ways to resolve the identification of '*treatment*' and '*control*' property issues are described below:

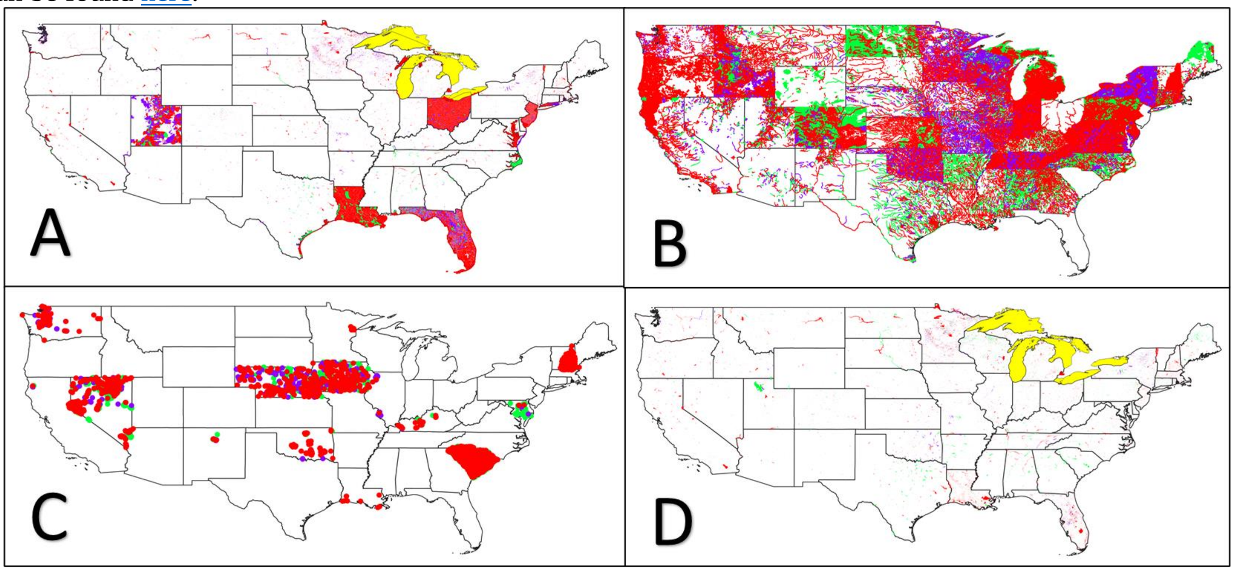
Identification of assessed waterbody

The ATTAINS geospatial database provides geographic information of assessment units in polygon, line, or point shape form. As lake, impoundment, and reservoirs are represented as points or lines, it will be misrepresentative if we calculate spatial variables (e.g., distance from lake to property) using mixed spatial shape form. We used spatial join to connect assessment units with NHD waterbody layer that provides polygon. However, the assessment units do not match one-to-one with NHD waterbody. An assessment unit can be part of a waterbody, or whole waterbody, or multiple waterbodies. We only use it when an assessment unit represents a whole or nearly whole waterbody.

Panel A, B, and C of Figure S3 show polygon, line, and point layers of assessment unit status. Notice that five state reporting agencies (FL, UT, LA, NJ, and OH) report their impairment status by watersheds instead of surface waterbody locations (panel A). The impairment status of one assessment unit (or waterbody) can be used to determine the impairment status of a whole watershed that has multiple waterbodies. It might make sense to report as a watershed as most of the clean-up plans (TMDL) are based on watersheds. However, we argue that TMDL plans are usually binding for other users (agricultural or industry users) rather than house owners. For this reason, a house owner is less likely to be directly impacted by the water impairment status of a watershed. For these reasons, we do not include these states in our main analysis. We check this in our robustness section.

[19] A property will be waterfront property if the property parcel polygon has shared boundary with lake boundary.

Figure S3: The figure shows ATTAINS geospatial data (panel A, B, and C) and matched NHD waterbodies (panel D). The red, green, and purple colors indicate the overall status of assessment units or waterbodies as impaired, fully supporting, and not assessed or not enough information respectively. The yellow color indicates great lakes (when applicable). Panel A, B, and C shows polygon (n=68,088), line (n=352,479) and point (n=3,999) layers from ATTAINS database respectively. After matching with NHD 7 million waterbodies, we got a database of 797,071 waterbodies shown in panel D. A better resolution of this figure can be found here.



Panel D of Figure S3 shows spatial join results of all these assessment units with NHD waterbody dataset. While spatially joining assessment units with NHD waterbody polygon layer, we estimate the percent of waterbody is assessed. For the polygon ATTAINS layer, we calculate the area of intersection of the assessment polygon layer and NHD waterbody polygon layer. Comparing the intersection area with NHD waterbody area, we find the percent of the waterbody assessed. If the assessment area is less than 80% of the total area of the waterbody, we exclude those waterbodies. One example can be Boga Lake, Minnesota where one of the bays is assessed and found to be fully supporting the designated uses (Figure S4). As the assessment area is only 20.28% of the total area of Boga lake, we exclude Boga lake from the analysis.

Figure S4: Waterbody area is partially assessed. The green color indicates the area of Boga lake, MN that is assessed and found fully supporting designated uses. The summation yellow and green area is the total area of the lake.

Similar to the polygon layer, we also calculate the percentage of line length assessed compared to the maximum length of a waterbody according to NHD dataset. We estimated the length of the waterbody in two methods. First, we estimate the maximum distance between any two vertices of NHD waterbody polygon. For some of the lakes, the number of vertices is quite high which makes calculating distances between vertices computationally intensive. In order to save computational resources at Minnesota Supercomputing Institute, we calculate distances among 40 randomly chosen vertices at maximum. Second, we estimate the diagonal distance of polygon encompassing the rectangle. The maximum length attained from these two methods is the total length of the waterbody. We compare the total length with assessment line length and estimate the percent assessed. Again, similar to the polygon layer, we use an 80% cut-off value.

Waterbodies with multiple bays or jurisdictions can be assessed and found to have conflicting impairment statuses. We exclude if a waterbody has conflicting impairment status. Figure S5 shows examples of each of these cases. Navajo Reservoir (panel A) located at the border of Colorado and New Mexico has two different impairment statuses by jurisdiction. It might be a result of adapting different water quality standards, or having different methods of measuring water quality, or measuring different water quality parameters altogether. Panel B shows the three bays of Gilbert lake in Minnesota where one bay is supporting designated uses, the second bay's impairment status is inconclusive due to lack of information, and the third bay is not assessed.

Figure S5: Multiple assessment units with differing impairment statuses. Panel A (left) shows the impairment status of the Navajo Reservoir. The Colorado part of the waterbody is listed as fully supporting while the New Mexico part is not supporting designated uses. In panel B (right), one bay of the Gilbert lake of Minnesota is listed as fully supporting, another bay as inconclusive, while the third bay is not assessed.

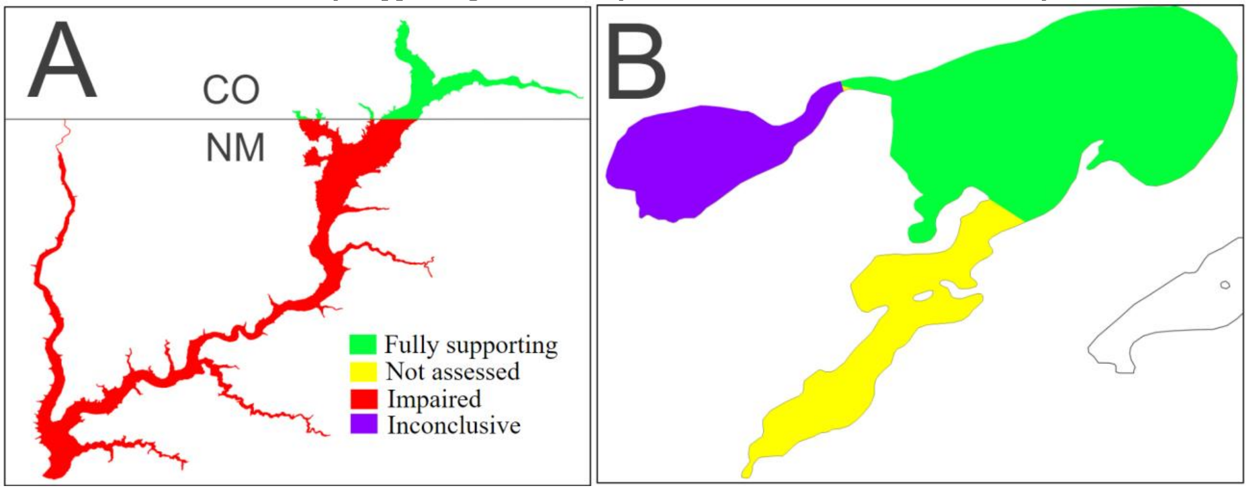
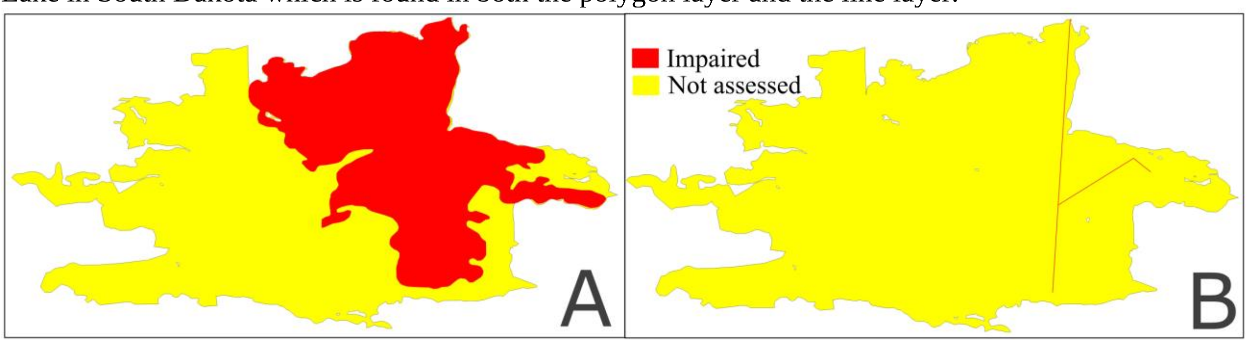


Figure S6: Same assessment unit can be listed in two different shape layers. Panel A and B are for Bitter Lake in South Dakota which is found in both the polygon layer and the line layer.



The reporting agencies represented some of the assessment units by different forms of shapes. We implemented an order system such that an assessment unit is used once. If an assessment unit is listed in the polygon layer, we do exclude the assessment unit from other layers. The precedence order is watershed, polygon, line, and point. An illustrative example is shown in Figure S6 where Bitter Lake is listed in both polygon and line layers. All other information other than the shape representation is the same for this assessment unit. We only include the assessment unit once in the polygon layer. Note that Bitter Lake is finally excluded as only 39.61% of the area is assessed.

Identification of impairment listing

The ATTAINS database lists waterbodies into several categories. The listing process depends on comparing water quality parameters of the waterbody with state-adopted and the USEPA approved water quality standards. Category 1 belongs to lakes with good water quality that fully support all the designated uses. Category 2 and 3 list waterbodies that lack information to determine if the waterbodies are impaired or not. All the impaired waterbodies that have a clean-up plan through TMDL approach are in category 4 and its sub-categories (4A, 4B, or 4C). Category 5 lists impaired waterbodies without a TMDL plan. Category 5A is also an impaired waterbody list but these waterbodies have alternative clean-up plans. A detailed description of categories is in Appendix A (Impaired water restoration process). In this analysis, we used category 4 and category 5 as impaired waterbodies where category 1 is fully supporting waterbodies. We do not include categories 2 and 3, and water bodies that are not assessed and listed in ATTAINS database as we do not know if they support designated uses or not.

Identification of the timing of impairment

In order to find if the nearest waterbody to a property is listed as impaired or not at the time of sale, we compare the sale year with *cycleFirstListed* variable that indicates when any water quality parameter is listed as impaired for the first time. Whenever *cycleFisrtListed* has missing data, we use the *cycleParm* variable, which indicates when the waterbody is listed in the ATTAINS database. However, we exclude if the year of sale is after *cycleFirstListed* but before min *cycleParm* reported in ATTAINS database.

Figure S7: Identification of the timing of waterbody impairment. We exclude observations if the year of sale is in-between cycle first listed and minimum cycle time. Minimum cycle time refers to the year when a waterbody is first documented in the ATTAINS database. Excluded observations are shaded in red.

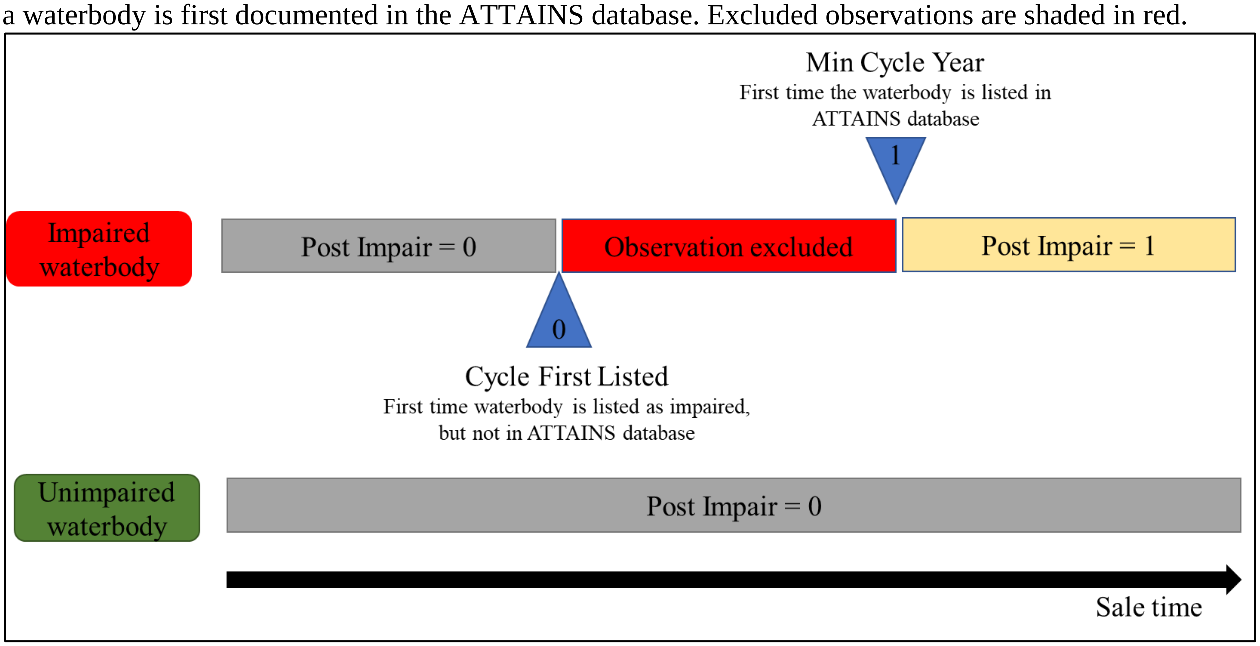
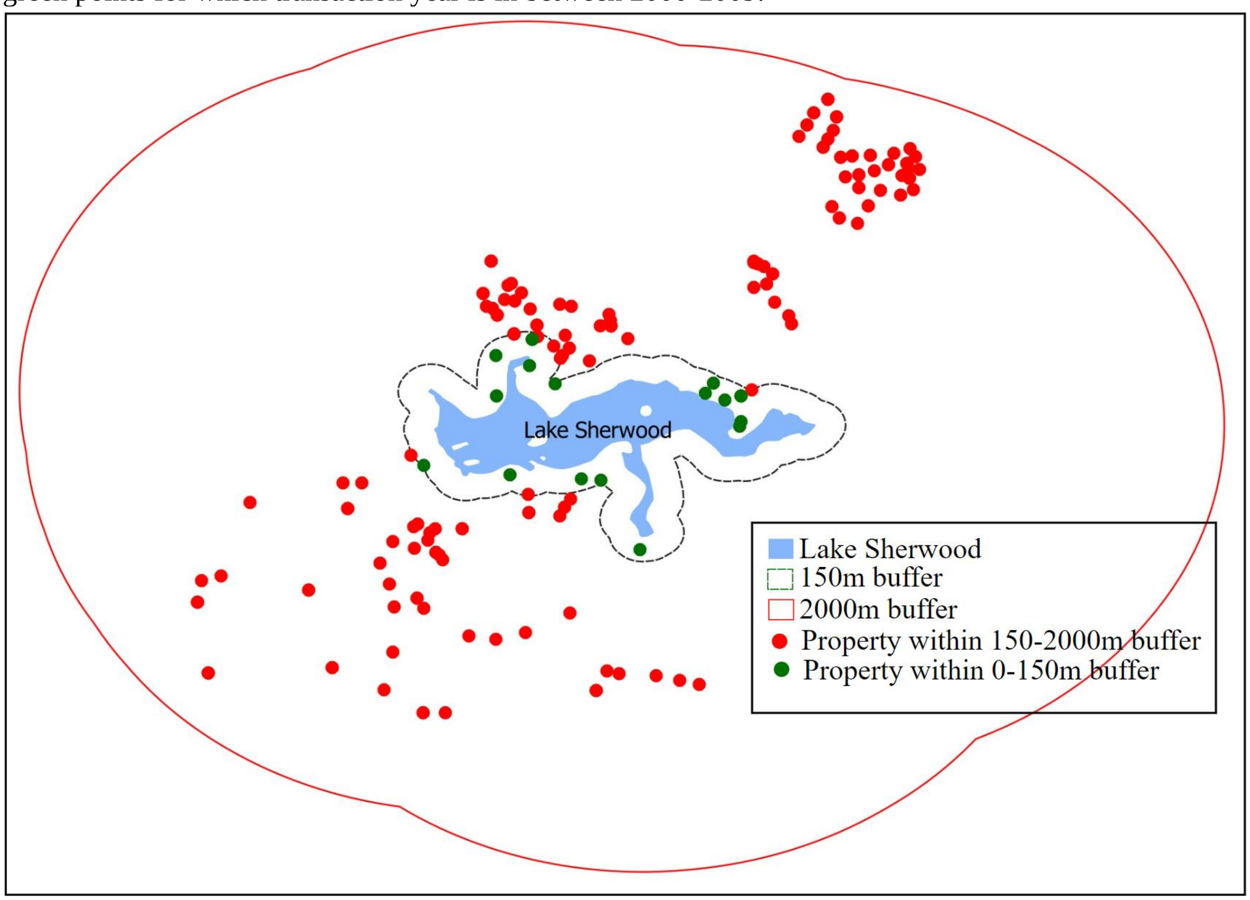


Figure S7 shows an illustrative example of treatment timing. A property sale whose nearest lake is impaired will be considered impaired waterbody is the year of sale is after it is documented in ATTAINS database. The same transaction will be pre-impaired if the sale year is before it is listed in the integrated report provided by the reporting agencies.

Identification of lakeshore properties

Lakeshore properties (properties within a buffer distance from the lake) have different responses to water impairment than properties distant from the lake. We used a LakeShore dummy variable (0-150m buffer) to capture the heterogeneous effect of distance to waterbodies. Our treatment property transactions are within 0-150m from the waterbody that is indicated as impaired waterbody at the time of transaction. Figure S8 shows an illustrative example of treatment and control property transactions.

Figure S8: An illustrative example of treatment and control selection. Figure shows Lake Sherwood in California which is listed as impaired in 2010. The green points indicate properties that are within 150m buffer of the lake. The red points are within 1500m buffer of the lake. Our treatment sales are green points for which transaction year is in between 2010-2019. Our control properties are all red points and some green points for which transaction year is in between 2000-2009.



Identification of relevant use impairments

Our choice of using water impairment status over other measures of water quality hinges on the assumption that housing market participants respond to relevant use impairment of the waterbody. ATTAINS dataset contains 857 different *associatedUseName* impairments and with the help of limnologists and water chemists we put them in 8 use impairment groups. The list of these grouping can be found in the supplementary document list and online link in Appendix F. Out of these 8 groups of designated use impairments we classify ‘Aesthetic’, ‘Recreation’, ‘Fishery’, and ‘Water Supply’ as relevant use impairments that the housing market participants will have a response. The rest 4 categories (‘Agricultural’, ‘Other’, ‘Aquatic life’, and ‘Fishery-other’) do not have a direct relation to house buyers and sellers. Take, for example, agricultural use impairment that includes ‘Irrigation’ and ‘Livestock and Wildlife Watering’ does not have a direct relationship with the housing market. Residents might care about the local economy, but that does not reflect their preference for house purchases. We test this in one of the robustness checks.

APPENDIX E: RESULT TABLES AND FIGURES

Table S3: Full results of the baseline model in equation (1). We used 'lfe' package in R to estimate the fixed effect. Fixed effect coefficients are not shown.

	Description	Baseline Model
PostImpair	=1, if the nearest waterbody is impaired at the time of sale	0.0127 (0.0106)
LakeShore (0-150m)	=1, if the property is located within 150m of the waterbody	0.2490*** (0.0095)
Lake size (km^2)	Size of lake measured in square km	0.0008*** (0.0002)
Lot area (m^2)	Lot size of the parcel in square meters	0.0668*** (0.0029)
Average Slope of Parcel	Average slope of the parcel measured from Digital Elevation Model	0.0036*** (0.0006)
Average Elevation of Parcel	Average elevation of the parcel measured from Digital Elevation Model	-0.0005*** (0.0002)
ln(distance to highway in m)	Natural log of distance to the nearest highway in TIGER dataset in meters	0.0141*** (0.0013)
ln(building age in years)	Natural log of building age in years	-0.0442*** (0.0023)
ln(building area in m^2)	Natural log of building area in sqm	0.3851*** (0.0086)
PostImpair x LakeShore	Interaction term of impairment indicator and lake proximity indicator	-0.0346** (0.0157)
N		660,707
adj. R-squared		0.9584
Log-likelihood		-207,412
BIC		929,907
AIC		491,695
Cluster		census tract
Fixed Effect		census tract x year
Lake		1,838
State		39

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

Table S4: Capitalization loss estimate for waterbody impairments.

Impairment data source	Extent	Region	percentage of lakes impaired	Estimated number of impaired lakes	Model estimate	Estimated number of lakeshore property of impaired lakes	Average fair market value of property	Capitalization estimate (billion USD)	Total capitalization loss (billion USD)
Sample	Impaired lakes in sample (550)	NA	100%	550	4.47%	80,696	366,406	1.321	1.32
Sample	All lakes with lakeshore properties (n=35,875)	NA	29.92%	10,735	4.47%	497,446	323,820	7.196	7.20
ATTAINS	All lakes with lakeshore properties (n=35,875)	NA	70.26%	25,206	4.47%	1,167,985	323,820	16.896	16.90
NLA (2012)	All lakes with lakeshore properties (n=35,875)	CPL	25%	1,912	4.47%	78,342	314,871	1.102	5.96
		SAP	25%	867		68,544	334,437	1.024	
		XER	28%	228		14,013	321,380	0.201	
		WMT	24%	312		23,992	649,258	0.696	
		SPL	24%	211		7,718	470,998	0.162	
		NAP	21%	1,177		49,078	295,946	0.649	
		TPL	27%	1,320		52,360	287,444	0.672	
		UMW	26%	2,902		121,239	267,437	1.448	
		NPL	26%	38		492	226,581	0.005	
NLA (2017)	All lakes with lakeshore properties (n=35,875)	CPL	29%	2,220	4.47%	90,978	314,871	1.280	5.40
		SAP	18%	619		48,926	334,437	0.731	
		XER	22%	177		10,885	321,380	0.156	
		WMT	21%	265		20,437	649,258	0.593	
		SPL	22%	196		7,171	470,998	0.151	
		NAP	19%	1,035		43,168	295,946	0.571	
		TPL	28%	1,358		53,851	287,444	0.691	
		UMW	22%	2,446		102,185	267,437	1.221	
		NPL	41%	59		768	226,581	0.008	
ATTAINS	All lakes with lakeshore properties (n=35,875)	NA	70.26%	25,206	10.39%	1,167,985	323,820	39.296	39.30

Table S5: Comparing lake impairment effect with water clarity measures on property values. Equation (1) is our baseline model for lake impairment.

	Equation (1)	Equation (5)	Equation (6)	Equation (1): full sample
PostImpair	0.0202 (-0.0139)		0.0455*** (-0.0158)	0.019 (-0.0132)
LakeShore (0-150m)	0.3573*** (-0.0145)	0.2577*** (-0.0129)	0.2706*** (-0.0146)	0.3301*** (-0.0126)
PostImpair x LakeShore	-0.0883*** (-0.0203)		-0.0491*** (-0.0186)	-0.0647*** (-0.0194)
ln(Secchi)		0.0267*** (-0.01)	0.0380*** (-0.0114)	
ln(Secchi) x LakeShore		0.1165*** (-0.0152)	0.1117*** (-0.0151)	
ln(Secchi) x PostImpair			-0.0279** (-0.013)	
N	448,578	448,578	448,578	660,707
adj. R-squared	0.6348	0.6362	0.6363	0.6594
Log-likelihood	-266,196	-265,377	-265,318	-388,043
BIC	867,447	865,808	865,729	1,291,169
AIC	583,885	582,246	582,134	852,957
Cluster	census tract	census tract	census tract	census tract
Fixed Effect	census tract x year	census tract x year	census tract x year	census tract x year
Lake	1,222	1,222	1,222	1,838
State	39	39	39	39

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

Table S6: Results showing time-invariant impairment status. Using time-invariant impairment status, the estimate is higher compared to time-variant impairment status. However, limiting to cross-sectional data, the model estimate is similar. For 2010, the Chi-square test for joint significance is not significant.

	Time-invariant	Time-invariant (limited to 2010)	Time-invariant (limited to 2016)	Time-variant (baseline Model)
Impair	0.0108 (-0.0116)	-0.0134 (-0.0368)	0.0494** (-0.0249)	
Impair x LakeShore	-0.0791*** (-0.0212)	-0.0365 (-0.0366)	-0.0767** (-0.03)	
PostImpair				0.019 (-0.0132)
PostImpair x LakeShore				-0.0647*** (-0.0194)
N	660,707	24,285	35,939	660,707
adj. R-squared	0.6595	0.6513	0.6563	0.6594
Cluster	census tract	census tract	census tract	census tract
Fixed Effect	census tract x year	census tract	census tract	census tract x year
Lake	1,315	336	882	1,838
State	39	29	34	39

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

Table S7: Results using binary waterbody size. The left-side panel A estimates equation (1) and the right-side panel B estimates equation (9) by dropping the continuous Lake Size variable.

	Panel A estimates equation (1)			Panel B estimates equation (9)		
	Large waterbody (>median)	Small waterbody (<=median)	Baseline model (include all waterbody)	Baseline model (include all waterbody)	Large waterbody (>median)	Small waterbody (<=median)
PostImpair	0.0008 (-0.0328)	0.0452*** (-0.0152)	0.019 (-0.0132)	0.0104 (-0.014)	-0.0395 (-0.0393)	0.0443*** (-0.0149)
LakeShore (0-150m)	0.4057*** (-0.0157)	0.1558*** (-0.0119)	0.3301*** (-0.0126)	0.3303*** (-0.0126)	0.4055*** (-0.0156)	0.1571*** (-0.0121)
Lake Size (km^2)	0.0010*** (-0.0002)	0.0865 (-0.068)	0.0011*** (-0.0003)			
PostImpair x LakeShore	-0.0571*** (-0.0208)	-0.0565*** (-0.0209)	-0.0647*** (-0.0194)	-0.0661*** (-0.0196)	-0.0570*** (-0.0209)	-0.0570*** (-0.0211)
N	328,023	332,684	660,707	660,707	328,023	332,684
adj. R-squared	0.6493	0.6869	0.6594	0.6592	0.6491	0.6869
Cluster	census tract	census tract	census tract	census tract	census tract	census tract
Fixed Effect	census tract x year	census tract x year	census tract x year	census tract x year	census tract x year	census tract x year
Lake	659	1,179	1,838	1,838	659	1,179
State	38	32	39	39	38	32

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

Table S8: Results showing results for non-relevant use impairments. Non-relevant use includes ‘Agricultural’, ‘Other’, ‘Aquatic life’, and ‘Fishery-other’.

	Non-relevant use impaired	Relevant use impaired (Baseline Model)
PostImpair	0.0295* (-0.0179)	0.019 (-0.0132)
PostImpair x LakeShore	0.3260*** (-0.0132)	0.3301*** (-0.0126)
N	592,173	660,707
adj. R-squared	0.6582	0.6594
Cluster	census tract	census tract
Fixed Effect	census tract	census tract x year
Lake	1,795	1,838
State	39	39

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

Table S9: Results for states that report impairment status by watershed. We ran two models: (1) 5 states that report by watershed, and (2) combine those data with baseline model and run altogether. The results are similar to our baseline model but slightly higher.

	watershed states only	Include watershed states	Baseline model
PostImpair	0.0338 (-0.0325)	0.0235* (-0.0124)	0.019 (-0.0132)
LakeShore (0-150m)	0.1745*** (-0.0182)	0.2895*** (-0.0109)	0.3301*** (-0.0126)
PostImpair x LakeShore	-0.0953*** (-0.0262)	-0.0784*** (-0.0177)	-0.0647*** (-0.0194)
N	292,810	953,517	660,707
adj. R-squared	0.7179	0.6814	0.6594
Cluster	census tract	census tract	census tract
Fixed Effect	census tract x year	census tract x year	census tract x year
Lake	752	2,589	1838
State	5	44	39

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

Table S10: Parameter estimates showing robustness when we limit our control dataset.

Distance buffer	**500**	**750**	**1000**	**1250**	**1500**	**1750**	**2000**
PostImpair	0.0049 (-0.0218)	0.0077 (-0.0162)	0.0084 (-0.0142)	0.0171 (-0.0132)	0.019 (-0.0132)	0.0111 (-0.0133)	0.0087 (-0.0124)
LakeShore	0.2930*** (-0.0116)	0.3070*** (-0.012)	0.3175*** (-0.0123)	0.3245*** (-0.0125)	0.3301*** (-0.0126)	0.3341*** (-0.0127)	0.3374*** (-0.0127)
PostImpair x LakeShore	-0.0533*** (-0.018)	-0.0593*** (-0.0184)	-0.0618*** (-0.0188)	-0.0646*** (-0.0191)	-0.0647*** (-0.0194)	-0.0645*** (-0.0197)	-0.0661*** (-0.02)
N	252,712	369,115	459,640	571,680	660,707	739,031	810,073
adj. R-squared	0.6153	0.6364	0.6487	0.655	0.9584	0.6636	0.6689
Cluster	census tract	census tract	census tract	census tract	census tract	census tract	census tract
Fixed Effect	census tract x year	census tract x year	census tract x year	census tract x year	census tract x year	census tract x year	census tract x year
Lake	1746	1799	1818	1831	1838	1838	1843
State	39	39	39	39	39	39	39

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

Figure S9: The effect of changing the size of the control buffer starting from 500m away from the lake and up to 2000m away from the lake with an increment of 250m.

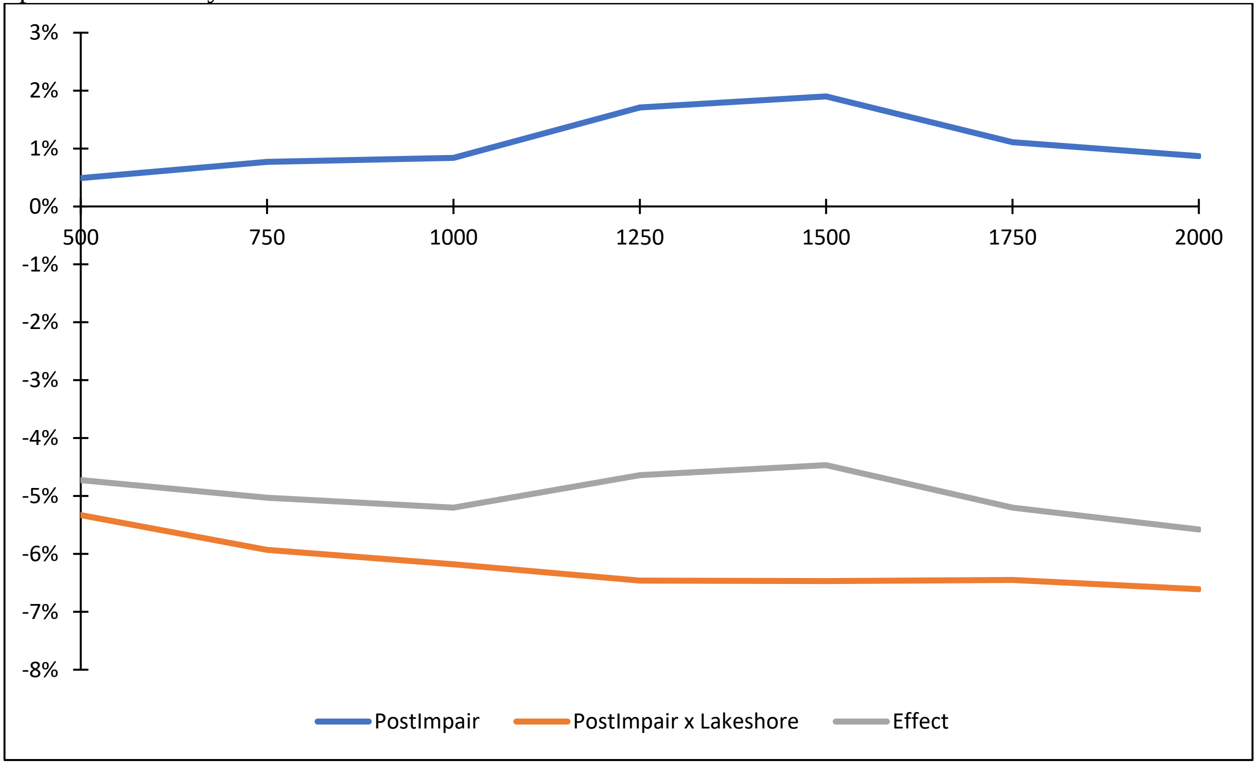


Table S11: Results for removing data based on distance from river. As we did not consider river water quality, houses that are nearby to a river than a lake are removed in the baseline. Here we relax this assumption by using specific distances to river – 50m, 100m, 200m, and 300m. Our estimations are robust to distance from river assumption.

	River distance 50m or more	River distance 100m or more	River distance 200m or more	River distance 300m or more	Baseline (lake distance < river distance)
PostImpair	0.0215 (-0.0138)	0.0207 (-0.0135)	0.0199 (-0.0132)	0.0198 (-0.0129)	0.019 (-0.0132)
LakeShore (0-150m)	0.3333*** (-0.0125)	0.3337*** (-0.0125)	0.3332*** (-0.0126)	0.3331*** (-0.0126)	0.3301*** (-0.0126)
PostImpair x LakeShore	-0.0615*** (-0.02)	-0.0613*** (-0.0199)	-0.0620*** (-0.0199)	-0.0628*** (-0.0198)	-0.0647*** (-0.0194)
N	731,589	728,272	719,531	707,495	660,707
adj. R-squared	0.6652	0.6655	0.6649	0.6644	0.6594
Cluster	census tract	census tract	census tract	census tract	census tract
Fixed Effect	census tract x year	census tract x year	census tract x year	census tract x year	census tract x year
Lake	1,846	1,846	1,832	1,823	1,838
State	39	39	39	39	39

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

Table S12: Results for using different inflation adjustment factors. We used three inflation adjustment factors. Our base case is seasonally adjusted housing price index from Federal Housing Finance Agency. (2021). Both consumer price index and not-seasonally adjusted housing price index as inflation adjustment provide similar results. The results do not vary within 4 decimal points that we report.

	CPI adjusted	HPI not-seasonally adjusted	Baseline model
PostImpair	0.019 (-0.0132)	0.019 (-0.0132)	0.019 (-0.0132)
LakeShore (0-150m)	0.3301*** (-0.0126)	0.3301*** (-0.0126)	0.3301*** (-0.0126)
PostImpair x LakeShore	-0.0647*** (-0.0194)	-0.0647*** (-0.0194)	-0.0647*** (-0.0194)
N	660,707	660,707	660,707
adj. R-squared	0.6642	0.6597	0.6594
Cluster	census tract	census tract	census tract
Fixed Effect	census tract x year	census tract x year	census tract x year
Lake	1,838	1,838	1,838
State	39	39	39

Significance: *** = p < 0.01; ** = p < 0.05; * = p < 0.1. Clustered standard errors are in parenthesis.

APPENDIX F: LIST OF DATA, CODE, AND SPREADSHEET

Name	Description	Anonymous link [osf]
Code and data folder	This folder provides source code to replicate the study. The folder also contains a high-resolution version of the images.	https://osf.io/gyd6f/?view_only=3968d945107846dcb0b198b055c7856d
parameterName.csv	List of different parameters and their corresponding parameter groups.	https://osf.io/gyd6f/files/osfstorage/6333615c0db48e0079e1113e
associatedUseName.csv	List of different *associatedUseName* and their corresponding designated use impairments.	https://osf.io/gyd6f/files/osfstorage/6333615cec7f3f007df5ffea
availability.csv	Data availability in the ATTAINS parameter REST API	https://osf.io/gyd6f/files/osfstorage/6333615e0db48e0078e10ecf
WIPV_all_models_2022_07_15.xlsx	Complete results of all the model runs.	https://osf.io/gyd6f/files/osfstorage/63370e6e31d653045f2ddd7c

Dataset S1 (Inputs and Key Results):

Housing data: We cannot provide the housing data used in this analysis. We used Zillow's ZTRAX (https://www.zillow.com/research/ztrax/) data processed through PLACES (https://placeslab.org/).

Water Impairment Data: Water impairment data can be obtained from ATTAINS. We used REST API requests using python. Codes are in the open science forum repository: https://osf.io/gyd6f/?view_only=3968d945107846dcb0b198b055c7856d

Water Quality Data: We used two sources of Water Quality data. They can be obtained using the following scripts:

1. *Water Quality Portal*: data can be downloaded using the portal link (*https://www.waterqualitydata.us*) and then processed through automated script (https://osf.io/gyd6f/files/osfstorage/6335c7210db48e0328e11539)
2. *LAGOS*-NE: Automated script (https://osf.io/gyd6f/files/osfstorage/6335c68cd5b010034e1f8025) LAGOS-NE can be used to download and clean LAGOS-NE data.

National Hydrography Dataset: National Hydrography dataset can be downloaded from USGS website (https://www.sciencebase.gov/catalog/item/4f5545cce4b018de15819ca9).

Direct link: https://prd-tnm.s3.amazonaws.com/StagedProducts/Hydrography/NHD/National/HighResolution/GDB/NHD_H_National_GDB.zip

Key results: The full regression results of various models can be downloaded from Open Science Forum: https://osf.io/gyd6f/files/osfstorage/63370e6e31d653045f2ddd7c

These regression results are estimated by varying different model specifications and parameters. These specifications and parameters are provided at the end of the dataset.

Code S1 (osf repository). The codes for cleaning water quality data, matching with housing data, and regression analysis can be viewed from an anonymous open science forum (osf) repository: https://osf.io/gyd6f/?view_only=3968d945107846dcb0b198b055c7856d

The project used the following dataset at the contagious US scale to find out what is the property value implications of water quality:

1. **Housing data:** Zillow's ZTRAX data processed through PLACES.
2. **Water quality data:** LAke multi-scaled GeOSpatial and temporal database (LAGOS-NE) and USEPA's Water Quality Portal for water quality measures.

Get Housing data

1. **collect_data.R:** This script collect fips level housing data that is processed through PLACES. It only contains housing data for houses located within 2km of all the lakes greater than 4ha. Latest data from PLACES is from May 19, 2021. We wont be able to able to provide proprietary data.

Get Water Quality Data

1. **EPA_data.py:** Get EPA-WQP water quality data and spatially match with NHD to get NHD ID. Data collected on March 10, 2021.
2. **LAGOS-NE.r:** Get LAGOS-NE Data including NHD ID (merge epi_nutr and lakes.geo data frames). Data collected on March 10, 2021.

Get ATTAINS Data

1. **attains_api_IR_parallel.py**: Get ATTAINS parameter data for all OrganizationParm except Pensylvania (21PA). Data collected on Apr 15, 2021.
2. **attains_api_IR_parallel_PA.py**: Get ATTAINS parameter data for Pensylvania (21PA). Data collected on Apr 15, 2021. We have to run Pennsylvania separately as this is huge dataset.
3. **attains_api_geo.py**: get ATTAINS geospatial polygon data. Data collected on Apr 15, 2021.
4. **attains_api_geo_line.py**: get ATTAINS geospatial line data. Data collected on Apr 15, 2021.
5. **attains_api_geo_point.py**: get ATTAINS geospatial point data. Data collected on Apr 15, 2021.

Combine Housing and Water Quality data
(not really needed for impairment project base case, but one of the robustness checks require this)

1. **Fuzzy_date_match.R:** This algorithm matches "exactly" between housing data and water quality data using NHDID and "fuzzily" between housing data and water quality data using housing transaction year and water quality sample year. If it cannot find a match still keeps the data. The outputs do not contain all housing attributes as it will slow

down the matching algorithm. It saves two csv files for chla and secchi. Similarly other water quality parameters can also be matched.

2. **regression_data.R:** Based on the sale_id identified in the Fuzzy_date_match.R, this algorithm matches housing data with each water quality data. It also creates several variables that we need to use in the regression. The output file sale_water_quality_data.pqt is basis of regression analyses.

Combine and Clean ATTAINS Data

1. **collect_impairment_data.R**: Combine attains parameters data together. It combines more than 1000 files obtained from ATTAINS parameter REST API.
2. **impair_nhd_geo_match.py**: combine impair geospatial data with NHD dataset. It has 6 outputs:
   i. *nhd_impair_merge_all.csv*: contains all match results in csv format
   ii. *nhd_match.pqt*: contains all match results in pqt format - spatial data.
   iii. *nhd_impair_merge_area.shp*: spatial data for area polygons only. It is further processed to get what is the intersecting area between the lake and assessment part of the lake.
   iv. *nhd_impair_merge_watershed.shp*: spatial data for watershed polygons only. Note that both area and watersheds are polygons, but we only include 5 states that report impairment by watersheds.
   v. *nhd_impair_merge_line.pqt*: spatial data for nhd lake polygons that are represented as lines in ATTAINS geo database. Note that the shape is polygon lakes that corresponds to lines in impair waterbody database. It is further processed to get what part of the lake is assessed. It is advantageous to make it pqt, otherwise huge file size.
   vi. *nhd_impair_merge_point.shp*: spatial data for nhd lake polygons that are represented as point in ATTAINS geo database. Note that the shape is polygon lakes that corresponds to point in impair waterbody database.
3. **geo_lakes_area.py**: The nhd lake polygons that are represented as area in ATTAINS geo database are further processed to get intersection area. It is used to find what percent of the lake is assessed.
4. **geo_lakes_line.py**: The nhd lake polygons that are represented as lines in ATTAINS geo database are further processed to get maximum length of the lake, total intersected line. It is used to find what percent of the lake is assessed.
5. **merge_impairment_data.R**: merge impairment data to get parameter group and use groups that is edited with limnologists inputs to make manageable number of groups.

Analyses

1. **DID_WIPV_base.R**: This is baseline analysis with multiple variations. The running model names are:
   i. baseline: run the model as it is.
   ii. waterbody_size: run the model but save waterbody_table variable.

iii. waterbody_size_drop: run the model where waterbody_eqn drops lake_size_sqkm from non_fe_control and save waterbody table

iv. adjust_cpi: run the model with dep_var='price_updated'

v. adjust_hpi_nsa: run the model with dep_var='price_updated_hpi_nsa'

vi. treat_shore_waterfront: redefine lake_shore as "sale_subset$lake_shore <- ifelse(sale_subset$lake_frontage > 0 | sale_subset$lake_dist==0, 1,0)"

vii. treat_shore_only_50 too 500 : change lake_shore definition from 150 to respective values

viii. control 2000 - 500: redefine "good_all <- good_all[good_all$lake_dist <= 1500,]". Instead of 1500 (baseline) use others [2000, 1750, 1250, 1000, 750, 500]

ix. watershed_with_baseline: do not delete watershed_states

x. watershed_only: delete other states. keep only watershed_states

xi. baseline_all: include all uses [baseline = relevant use only].

xii. baseline_non_relevant: use only non-relevant uses; relevant_uses = ['Aesthetic', 'Recreation', 'Fishery', 'Water Supply']; non_relevant_uses =['Agricultural', 'Other', 'Aquatic life', 'Fishery-other']

xiii. river_15000: control for river distances, baseline is to delete if the distance from lake is greater than the distance from river. other distances = [50, 100, 200, 300]

xiv. change the fixed effects: baseline = fe_year_by_tract; others = [fe_year_state, fe_year_by_state, fe_year_tract, fe_year_fips, fe_year_by_fips]

2. **DID_WIPV_parameter_updated.R**: analysis to see if the effects vary by parameters.
3. **DID_WIPV_use_updated.R**: analysis to see if the effects vary by designated uses impaired.
4. **DID_WIPV_distance_decay.R**: Analyses to see if there are any distance decay effects using alternative definitions of distance (dummy, continuous, inverse continuous).
5. **DID_WIPV_lake_shore.R**: analysis to see if alternative definitions of lakeshore have an impact on the results.
6. **DID_WIPV_secchi.R**: compare impairment results with measured water clarity data. **This is why we cleaned water quality data**.
7. **DID_WIPV_summary.R**: provide summary statistics.
8. **lake_maps.py**: visualization of key coefficient results for models.