Impact of Waterborne E. coli Outbreaks on Local Communities: Evidence from Housing Transactions in Michigan

Marcello Graziano[1], Pengfei Liu,[2], Kevin Meyer[3] and Wendong Zhang[4]

[1]Ruralis, Norwegian University of Science & Technology, Loholt alle 81, 7049 Trondheim, Norway

[2]Department of Environmental and Natural Resource Economics, University of Rhode Island. 1 Greenhouse Road, University of Rhode Island, Kingston, RI 02881, USA.

[3]Department of Economics, Saginaw Valley State University, 321 Curtiss Hall, 7400 Bay Road, Saginaw, MI, 48710, USA.

[4]Dyson School of Applied Economics and Management, Cornell University, 461 Warren Hall, Ithaca, NY, 14853, USA.

Corresponding author: Kevin Meyer kmeyer@svsu.edu

This paper was completed with no internal or external funding.

The data used for this paper was made available through a confidential agreement with Zillow and is not available for distribution.

Abstract

Waterborne Escherichia coli (E. coli) represents a pervasive water quality problem across the United States. In Michigan, the presence of E. coli has become problematic for many areas where agricultural run-off and ineffective policies have made these outbreaks endemic. Combining the universe of housing transaction dataset from 2009 to 2017 with the State of Michigan water sampling dataset, we investigate and quantify the negative impacts of E. coli outbreaks on local housing prices. Our difference-in-differences model estimates an overall impact of -8.94% for houses in the treatment group relative to the control group. This effect is only short-term, as sales prices recover after the outbreak has ended.

1 Introduction

Ecological degradation leads to the loss of valuable ecosystem services and creates many potentially negative socioeconomic consequences at various scales (Scott et al., 2011). Economists often use housing prices to capture (at least partially) the social costs of decreased ecological quality, including those associated with water quality and its preservation (Wilson and Carpenter, 1999; Spanou et al., 2020). Among the many sources of ecological pressure on U.S. water systems, waterborne pathogenic *Escherichia coli* (*E. coli*) can cause immediate and potentially lethal effects on human health (Ishii and Sadowsky, 2008). *E. coli* also constitutes a pressing issue due to an increased rate of outbreaks driven by climate change, especially in the Laurentian Great Lakes basin (GLB) (Patz et al., 2008). Although bacteria levels in water decrease over time, *E. coli* can remain in sediments along the shores or on beaches, thus continuing to affect water quality and making outbreaks more likely over time (Davies et al., 1995). In Michigan, these outbreaks have become a growing issue for both inland and coastal communities, especially in rural counties. According to the state's website, roughly half of Michigan's waterbodies exceed the EPA's water quality standard for *E. Coli* as of early 2024.[1]

Given the harmful effects of *E. Coli* and its prevalence in Michigan, it is important to quantify these effects to help design efficient mitigation policies. In this study we use hedonic pricing models and non-water-clarity-based approaches to estimate the effect of *E. coli* outbreaks in freshwater streams on housing prices across Michigan. To assess the effects of these outbreaks, we use point-level measurements conducted by Michigan's Environment, Great Lakes, and Energy Department (EGLE), combined with Zillow point-level housing transaction data from 2009 to 2017. We first establish a baseline model using a traditional hedonic framework, where we regress the log of the sales price on the distance from the nearest outbreak along with a set of covariates and fixed effects and find that houses closer to outbreaks see a decline in sales price. Following Currie et al. (2015), Haninger et al. (2017), and Tanaka and Zabel (2018), we next employ a difference-in-differences (DID) model to capture the effect of recent *E. coli* outbreaks on the sales price of nearby houses. This DID approach mitigates the omitted variable bias and endogeneity concerns associated with the distance to *E. coli* outbreaks and is widely used in quantifying the value of environmental attributes.

Our main results suggest that proximity to *E. coli* outbreaks negatively affects housing prices, and that these effects are heterogeneous across multiple distances. After controlling for housing characteristics and spatial fixed effects, our preferred DID regressions reveal that proximity to *E. coli* outbreaks leads to a 8.9% price drop for houses sold during the outbreaks,

[1] https://www.michigan.gov/egle/about/organization/water-resources/assessment-michigan-waters/e-coli-in-surface-waters

which is over $13,000 for the average house. However, these effects do not persist past an interim period.

This study contributes to the literature linking property values and water quality. First, it is the first to exploit a large dataset linked to waterbodies across the entire state of Michigan over more than a decade. Second, while there have been many studies that quantify the impacts of various water quality variables on housing prices, few focus on *E. Coli* despite its risk. Our estimates will present a baseline of comparison for future research.

The remainder of this article proceeds as follows. We first review the relevant literature on hedonic analyses related to water quality and provide background information on water quality issues in Michigan. Next, we provide a description of the data used in this study, followed by a summary of the empirical models we employ. We then present the main results along with policy implications before the concluding section.

2 Background

Food and waterborne illnesses are one of the leading causes of morbidity worldwide, with diarrheal diseases in particular accounting for approximately 1.8 million deaths each year. Although most cases are found in developing countries, there are still about 76 million cases of foodborne illness in the U.S. each year, resulting in around 5,000 deaths annually (Ishii & Sadowsky, 2008).

One of the leading causes of waterborne illnesses is the bacteria *E.Coli.* Although *E. Coli* is naturally found in the intestines of humans and animals and is typically harmless, some strands can cause severe illness or even death. Some of the most dangerous strains are thought to originate from untreated human sewage as well as animal waste (WHO, 2018). Since these sources can travel through runoff into nearby waterways, officials typically rely on water samples to detect dangerous levels of *E. Coli*. These samples are compared to water quality standards to determine if action should be taken. In Michigan, for example, the state has set a standard of a daily maximum of 300 *E. Coli* per 100 ml of water, or a geometric mean across 30 days of no more than 130 *E. Coli* per 100 ml of water. As of 2024, the state of Michigan estimates that over half of Michigan's waterbodies exceed these levels, and about 20% of monitored beaches have been closed recently due to bacterial pollution (EGLE, 2024). Regarding the Great Lakes, although Lake Michigan's *E. coli* and swimming advisories have decreased in recent years, *E. coli* still affects more than 10% of Lake Michigan beaches and almost 90% of beaches in western Lake Erie (Weiskerger and Whitman, 2018).

There are three main mechanisms through which the state makes outbreaks public. First, when a waterbody exceeds a water quality standard, a Total Document Daily Load (TMDL) document is required by the Federal Clean Water Act. A TMDL shows the recorded levels of *E. Coli*, its likely sources, and possible regulatory solutions, but does not require any action to be taken. *E. Coli* measurements, potentially dangerous waterbodies, and current TMDL's are made public through EGLE's website. The second way that *E. Coli* outbreaks are made public is through outdoor signs installed by the state near afflicted waterbodies. Finally, the Michigan Sellers Disclosure Act of 1993 requires sellers to notify buyers of recent environmental problems.

Although Michigan tracks *E. Coli* levels throughout the state and warns the public of outbreaks, there are currently no Michigan laws that regulate these levels. When state governments consider potential policies, they typically undertake a cost-benefit analysis. A cost-benefit analysis associated with a policy to reduce *E. Coli* levels requires damage estimates. One approach, used by the USDA Economic Research Service (ERS) in 2013 and updated in 2018, is to identify the different ways *E. Coli* has impacted society, estimate the dollar amount of each, and add them up. The USDA-ERS identified medical costs, productivity loss, and deaths as the main human damages due to *E. Coli*. The sum of the damages for the entire U.S. was close to 300 million dollars (Ahn, 2021). Another common strategy is the benefit transfer approach, which uses pre-existing estimates in one setting to predict measure of economic value in a different setting (Johnston et al., 2012). The benefit transfer approach is often used to support decision making based on cost-benefit analysis when time or funding for a new study is cost prohibitive. However, the lack of previous estimates of the economic impact *E. Coli* on housing prices makes this option unavailable to current decision makers.

A different approach commonly used in economics is to estimate damages through stated or revealed consumer preferences. The stated preference approach uses a survey to ask, for example, if a respondent would vote for a tax intended to clean up a lake (Meyer, 2020), or if a respondent would prefer one environmental over another with different attributes. The revealed preference method, on the other hand, is based on observed data of the choices people made.

One of the most common techniques in the revealed preferences literature is known as the hedonic method (Rosen, 1974). In the context of house sales, the hedonic method captures a world where buyers and sellers of new or existing houses are assumed to reach an equilibrium point where neither can be made better off without losing utility. This implies a relationship between housing prices and housing characteristics that reveal a consumer's willingness to pay for certain housing characteristics (Bishop, et al., 2020). While these characteristics include the physical characteristics of the house, such as the number of bathrooms, they also include amenities and disamenities located in the region. This makes the hedonic method one of the main tools that economists have used to estimate the costs of both air and water pollution.

The literature of hedonic studies related to water quality has grown over time as the availability of water quality data improved. One of the key decisions in these studies is the choice of water quality variable. Heberling et al (2024) conducted a meta-analysis of hedonic models that use water quality and found studies based on water clarity, nutrients, sediment, biochemical factors, and bacteria. Of the studies on the effects of bacteria on housing prices, most focused on fecal coliform. For example, one of the earliest hedonic water quality studies by Leggett and Bockstael (2000) studied the effects of fecal coliform on housing prices around Chesapeake Bay and found that an increase of 100 fecal coliform counts per 100 ml produced an approximate 1.5% decrease in property prices. However, Heberling et al (2024) only found one paper that focused on the effects of *E. Coli*. Netusil et al (2014) studied two watersheds in the U.S. Northwest and found that an increase of 100 count per 100 ml increase in *E.Coli* decreased housing prices from -.71% to -2.90% depending on the distance to the stream and the econometric model used. Our research provides an additional estimate of the effects of *E.Coli* on housing prices to help fill the void in this literature.

The welfare effects of water pollution are particularly important to Michigan, which has long relied on its freshwater as a resource to boost economic development, whether using it to harvest natural resources, to support its manufacturing sector, or as way for transporting goods to and from the Atlantic Ocean (Steinman et al., 2017). Michigan's Freshwater is also a major source of employment—a 2007 study estimates that 2.7 million Michigan jobs are linked to the Great Lakes (Allen-Burton et al., 2010). The state currently faces a highly fragmented wastewater policy landscape, which allocates most of the monitoring and implementation powers to counties and county subdivisions. As a result, decades of industrial and agricultural pollution combined with the fragmented water policy has led to several water quality issues across the state (Allen-Burton et al., 2010). Brashares (1985) studied 78 lakes in southeast Michigan and found that fecal coliform had a negative effect on the sales price of lakefront houses. Rabinovici et al. (2004) found that the closure of an average Michigan lake could create an economic loss of up to $37,000 per day based on a benefit transfer analysis. Wolf et al. (2017) found that algal blooms in Lake Erie resulted in $2.25–$5.58 million in losses to the fishing industry. To our knowledge, however, there are no papers that focus on the effects of *E. Coli* on housing prices in Michigan. By using home sales and water quality information across Michigan, our study fills a gap in the literature on valuing Michigan water quality, which is crucial for policymakers to design cost-effective regulation.

3 Data

We combine two fine-grain data sets that cover the entire state of Michigan. The first dataset we derive from Zillow residential housing transaction data for single-family homes (ZTRAX)[2] from 2009 to 2017. ZTRAX lists sales prices, latitude and longitude, and various housing characteristics including total bedrooms, total bathrooms, lot and building square footage, and number of stories for all properties posted on Zillow. The full dataset over 9 years includes almost half a million observations, but for our main results we drop a number of observations that may be considered outliers. As a first step, we drop observations where the sales price was less than $10,000 to avoid "arm's length" transactions. From this sample, we further restrict the sample to houses that have less than 10 bedrooms and/or 50 total rooms, sold for less than

[2] This database is Zillow's Assessor and Real Estate Database (ZTRAX) accessed through a contract with Zillow.

$1,000,000, and have less than 1,000,000 square feet. We also remove houses that were labeled as having zero bedrooms. The remaining subset contains 195,331 observations[3]. Finally, we control for housing inflation by converting sales price to 2017 dollars using the Case-Shiller home price index.

The second dataset we use is EGLE's publicly available data on *E. coli* point-level measurements over the same period as the housing data. The *E. coli* data set is geo-coded, allowing us to estimate the precise proximity of each house to an *E. coli* outbreak. Within the EGLE database, we drop all points whose samples were below the risk levels considered by the states. These risk thresholds are:

1. 30-day geometric mean across three sample points greater than 100 *E. coli*/ml: OR
2. Daily maximum geometric mean of greater than 300 *E. coli*/ml

These values are slightly more restrictive than those imposed by Michigan, but in line with those used by neighboring states.

E. coli outbreaks that occur through water are typically the result of increases in rainfall or melting snow (Griffith et al., 2003; Roslev and Bukh, 2011). With a relatively low die off rate, once *E. coli* enters a waterway it can travel large distances away from the source through the available network of rivers and streams (Foppen et al., 2006). Between the randomness of precipitation events and the ability to travel long distances from the source, we consider *E. coli* outbreaks as exogenous to the sales price of a house. Figure 1 displays the correlation between house sales prices and *E. Coli* outbreaks throughout Michigan from 2009-2017. The highest

[3] The results for the full sample are in the appendix and do not qualitatively differ from the main results.

correlation is in the lower half of the lower peninsula of Michigan, particularly near the middle. In total, there were 3,763 samples above limits throughout our ten-year study period.

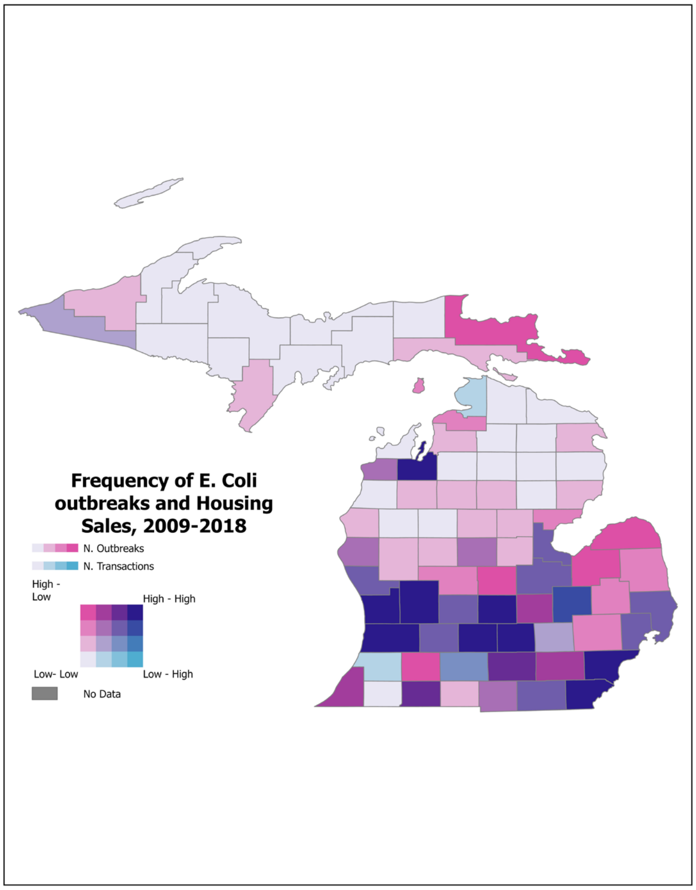

Figure 1: *E. coli* outbreaks in Michigan, 2009–2017. Note: location of houses is not displayed to protect privacy and data confidentiality.

We merge the Zillow housing transaction data with *E. coli* outbreak data and use ArcGIS Pro and PyCharm to calculate the nearest distance in miles from a particular house to an *E. coli* outbreak above the state risk thresholds that occurred after the sales date of the house. We also calculate the number of days since the outbreaks occurred. Table 1 shows the summary statistics of our variables.

Table 1 Summary Statistics

	Mean	Std. Dev.	Minimum	Maximum
Sales Price (2017 dollars)	151,923.2	109,119.1	10,003.7	996,604.4
Distance from outbreak (miles)	2.0	0.7	0.0	3.0
Days Since Outbreak	-423.8	1071.2	-3534.0	3118.0
Lot Size (sq. ft.)	23,241.4	59,241.7	1011.2	991,861.0
Rooms	3.3	3.4	0.0	37.0
Bedrooms	3.1	0.8	1.0	10.0
Full Bathrooms	1.5	0.7	0.0	12.0
Stories	1.5	0.6	1.0	10.0
Building Area (sq. ft.)	1353.2	611.6	300.0	10,199.0
Distance to waterbody (meters)	3546.4	2481.4	0.0	13,323.1
Distance to waterbody^2	1.9e+07	2.5e+07	0.0	1.8e+08
treat	0.1	0.3	0.0	1.0
after	0.3	0.5	0.0	1.0
interim	0.0	0.2	0.0	1.0
Observations	195,331			

4 Empirical Models

Since Rosen's seminal work (Rosen, 1974), the hedonic property value model has become the workhorse model to reveal the marginal values for non-market characteristics. The main idea of a hedonic model is that a product's price represents a package of attributes- in the case of houses, this includes not only the square footage, number of bedrooms, etc., but also location specific attributes such as environmental quality. By regressing the observed sales prices on these attributes, we can recover an estimate for the marginal values for being located at various distances from an *E. coli* outbreak.

A criticism of hedonic property values is the potential for omitted variable bias. Specifically, we are only able to control for a limited number of characteristics- even highly detailed housing datasets will inevitably leave out features that are valued by potential homeowners. These omitted features will be captured by the error term, and if they are correlated with *E. coli* outbreaks the coefficient of interest will be biased.

Given the potential for omitted variable bias in a hedonic model, we next employ a difference-in-difference (DID) model. A DID model assigns houses to a control group- in our case, houses that were not affected by an *E. coli* outbreak- and a treatment group for houses that were affected. Although their average prices may differ before an outbreak, we assume that they are trending in a similar direction. After an *E. coli* outbreak, average prices for untreated houses should continue along the same trend, while treated houses will follow a different trend if *E. coli* outbreaks influence the price. While we prefer the DID model, we include the hedonic estimates as a point of comparison between two popular approaches in the literature. Differences between the two estimates may also point to the effect of omitted variable bias in the hedonic approach.

4.1 Hedonic Property Value Model

In our baseline model specification, the dependent variable y_{it} is the log-transformed housing transaction price, adjusted by the S&P/Case-Shiller U.S. National Home Price Index to 2017 dollars. We can parsimoniously write our model as:

$$y_{ite} = \alpha + \beta_1 dist_{ite} + \beta_2 days_{ite} + \ \theta X_i + \eta_t + \eta_e + \epsilon_{it},$$

where $dist_{ite}$ is the distance between house i sold in year t and the nearest *E. coli* outbreak e occurring before the sales date; $days_{ite}$ is the number of days since the outbreak was reported; X_{it} is a vector containing the set of housing characteristics; α is the intercept; and, ϵ_{ite} is the idiosyncratic error term. We use η_t and η_l to represent year and month fixed effects and school district fixed effects, respectively, which account for the time- and location-invariant unobserved characteristics. Finally, we use the outbreak-level fixed effect η_e to capture idiosyncratic factors related to a particular outbreak. We cluster standard errors at the outbreak level.

Our main parameter of interest is β_1, which measures the buyers' marginal willingness to pay for being away from an *E. coli* outbreak. Since we measure the distance to an *E. coli* outbreak as distance from (rather than the proximity to) an outbreak, we hypothesize that β is positive, suggesting that houses further away from an *E. coli* outbreak would sell for a higher price. We also estimate a model that includes a quadratic functional form for distance to capture possible non-linear effects. We next employ a DID model to mitigate potential omitted variable concerns.

4.2 Difference-in-Differences Model

Our strategy follows the approach of Currie et al. (2015), Haninger et al. (2017), and Tanaka and Zabel (2018) in employing a DID model based on distance from the treatment, in our case an *E. coli* outbreak.[4] Within a certain distance, houses are similar enough that we can consider them a local market, but we can still separate the market into a treatment and control group. Most *E. coli* outbreaks are local and we use a radius of 3 km from an *E. coli* outbreak to determine local neighborhoods, which is similar to Hanninger et al. (2017). The treatment group consists of houses that are particularly close to the outbreak, while houses farther away are in the control group.

[4] We attempted to use STATA to run models that use recent advances in the DID literature, including a CSDID and DRDID estimator, and a Goodman Bacon decomposition. However, our dataset did not match the specifications required for those models.

Let $TREAT_{it}$ be a dummy variable that equals 1 if house i is within this boundary and belongs to the treatment group, and 0 if the house belongs to the control group. Let $POST_{it}$ equal 1 if a house is sold after an *E. coli* outbreak, and 0 if it is sold before the outbreak. We can then state the DID model as:

$$y_{it} = \beta_0 + \beta_1 TREAT_{it} + \beta_2 POST_{it} + \beta_3 TREAT_{it} * POST_{it} + \epsilon_{it}, \qquad (2)$$

where y_{it} is the log of the sale price of home i at time t, and ϵ_{it} is the error term that contains unobserved factors. The variable of interest, β_3, captures the difference in the expected value of y_{it} for houses in the treatment group versus the expected value of y_{it} in the control group. Specifically:

$$\beta_3 = \left(E[y_{it}^1 | TREAT = 1] - E[y_{it}^0 | TREAT = 0]\right) \qquad (3)$$
$$-(E[y_{it}^0 | TREAT_{it} = 0] - E[P_{it}^0 | TREAT_{it} = 0),$$

where the superscripts equal 1 if the house is in the *E. coli* treatment group in the counterfactual state, and 0 if the house is in the *E. coli* control group in the counterfactual state.

The main identifying assumption of equation (3) is the parallel trend assumption, where the sales prices leading up to an *E. coli* outbreak follow the same trend in both the treatment and control groups. In figure 2 we visualize the results of an event study to investigate the validity of this assumption. This event study uses 360 day intervals to track the evolution of the coefficient of interest, before and after an *E. coli* outbreak.[5] Ideally, the coefficient would be zero before an *E.coli* outbreak, which is satisfied as the 95% confidence intervals overlap the x-axis. Following

[5] We also tested this assumption for 180- and 270-day increments, which produced fairly similar results.

an *E. coli* outbreak, we would expect negative, statistically significant coefficients. This appears to be the case for the near term, with the effect dissipating over time.

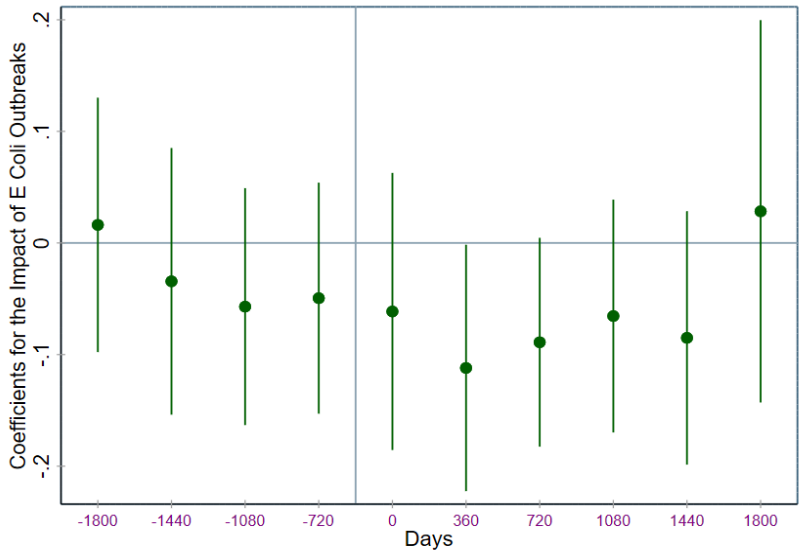

Figure 2: Coefficients of the impact of E.Coli outbreaks at 360 day intervals. Day zero indicates the initial outbreak of E.Coli. Vertical lines represent 95% confident intervals.

Although figure 2 offers suggestive evidence that the parallel trends assumption is satisfied, without additional covariates to control for observable characteristics, this assumption may be too strong. As a result, we include additional covariates including housing characteristics, time and location fixed effects, and outbreak fixed effects:

$$y_{it} = \beta_0 + \beta_1 TREAT_{it} + \beta_2 POST_{it} + \beta_3 TREAT_{it} * POST_{it} + \theta X_{it} + \eta_i + \eta_t + \eta_k + \epsilon_{it}. \quad (4)$$

To estimate the impact of *E. coli* outbreaks, we need to identify the treatment buffer zone in which an outbreak influences houses that are sufficiently close (*TREAT*= 1), while not influencing those that are sufficiently far away (*TREAT* = 0). Following Haninger et al. (2017), we first control parametrically for housing attributes, and then nonparametrically estimate housing price gradients over distance for houses close to *E. coli* outbreaks before and after the *E. coli* outbreaks separately. We then determine the distance threshold by identifying the point where these two price gradients converge using a nonparametric approach, meaning that beyond this distance, housing prices before the *E. coli* outbreak are not statistically different from those sold after the outbreak. If home buyers disliked the *E. coli* outbreaks, we should observe a lower price for houses within a distance threshold of the outbreak following that event. Moreover, we expect to see no differential patterns in housing prices outside this treatment buffer.

Figure (2) plots the estimated price gradients over distance to the nearest *E. coli* outbreak for houses sold before and after an outbreak. As expected, the prices of houses sold after the outbreaks are noticeably lower than those before the outbreaks, up to approximately one mile away, where the 95% confidence intervals of the two price gradients start to overlap. Figure 2(b) further breaks down the sale timing as before, during, and after the outbreak, and it reveals that beyond one mile, the three confidence intervals largely overlap, especially for houses sold during and before the outbreaks. This graphical evidence suggests that *E. coli* outbreaks dampen nearby housing prices and provides support to our DID approach of classifying houses within and outside one mile of an outbreak as the treatment and control groups, respectively.

Figure 3. Nonparametric estimates of housing price gradient with 95% confidence intervals for houses in Michigan sold before and after *E. coli* outbreaks.

Our data on *E. Coli* outbreaks is unique in that the outbreak has a recorded start and end date, which we label the *interim* period. This period might also be thought of as the short-term effect of *E. Coli* on housing prices once an outbreak has been made public. In our dataset, the average interim period is about three months (86 days). Therefore, to further quantify the potential differential impacts on houses sold during versus after *E. coli* outbreaks we include the dummy variable *INTERIM*. If a house is sold in between the *E. coli* outbreak starts and end date for the nearest outbreak, we denote this sale as sold *during* the outbreak, and thus INTERIM equals 1. Our variable of interest for this scenario is an interaction term between *INTERIM* and *TREAT*, which measure the impact of outbreaks on houses within the 1 mile buffer during the interim period. In addition, we use POST = 1 to denote the houses sold *after* the end date of the nearest *E. coli* outbreak.

$$y_{it} = \beta_0 + \beta_1 TREAT_{it} + \beta_2 POST_{it} + \beta_3 TREAT_{it} * POST_{it} + \beta_4 INTERIM_{it} + \beta_5 TREAT_{it} * INTERIM_{it} + \theta X_{it} + \eta_i + \eta_t + \eta_k + \epsilon_{it}. \quad (4)$$

4 Results

4.1 Hedonic Model Results

We first present the results of the baseline hedonic property value model. Note that, for comparison, we use the same estimating sample in the hedonic regressions and DID regressions. Table (2) shows the regression coefficients for the distance to the nearest *E. coli* outbreak after controlling for a set of housing characteristics and various fixed effects. All three specifications use the log of housing sales prices as the dependent variable. Column (1) is a simple regression analysis using only the variable of interest, distance, as a control variable. Column (2) adds the distance to the waterbody and its square, days since the outbreak, housing characteristics and year by month fixed effects. Column (3) further adds school district and outbreak fixed effects.

The estimated coefficient on *distance from outbreak* is positive in all three models, which aligns with our intuition—houses that are farther away from an outbreak sell for a higher price relative to houses that are closer. With the semi-log functional form, the magnitude of the coefficient varies from 1.39% to 3.06%, although the model with the full complement of fixed effects was not significant at the 5% level. The signs on the housing characteristics are intuitive, as increases in lot size, house size, number of bathrooms, and number of stories lead to increases in sales prices.

Although the sign on *distance from outbreak* is intuitive, the mixed results for statistical significance, as well as the possibility of omitted variable bias, cast some doubt on the reliability of the results. In the next section we show the results for the DID model, which uses a quasi-experimental method to better control for omitted variable bias.

Table 2 Hedonic Model

	(1)	(3)	(3)
Distance from outbreak (miles)	0.0304***	0.0139***	0.0306
	(0.00235)	(0.00231)	(0.0279)
Distance to waterbody (meters)		0.00000420**	0.00000293*
		(0.00000187)	(0.00000166)
Distance to waterbody^2		-2.54e-10	-8.85e-11
		(1.85e-10)	(1.60e-10)
Days Since Outbreak		-0.0000220***	-0.00000104
		(0.00000250)	(0.00000269)

Lot Size (sq. ft.)		0.000000840***	0.000000908***
		(3.11e-08)	(8.32e-08)
Rooms		-0.00174***	0.00560***
		(0.000522)	(0.00198)
Bedrooms		0.00976***	0.0105
		(0.00272)	(0.0120)
Full Bathrooms		0.240***	0.177***
		(0.00313)	(0.0136)
Stories		0.0220***	0.0298***
		(0.00278)	(0.0108)
Building Area (sq. ft.)		0.000388***	0.000285***
		(0.00000372)	(0.0000190)
Observations	195331	160324	160304
Adjusted R^2	0.001	0.276	0.446
Year/Month FE	No	Yes	Yes
School District FE	No	No	Yes
Outbreak FE			Yes

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parenthesis.

4.2 Difference-in-Differences Results

Table (3) shows the regression results for three specifications of the DID model. In this model, we keep all observations before and after an outbreak, as opposed to the hedonic model in which we only keep post-outbreak observations. However, we subset the data for the DID model to a radius of three miles around each outbreak to keep the neighborhoods and houses relatively homogeneous between treatment and control groups. We define the treatment observations as houses within a one-mile radius from the outbreak, while the control observations are between one and three miles from the outbreak.

We again start with a parsimonious baseline model to evaluate the evolution of the coefficient of interest (*Treat * Post*) (as we control for additional factors). Since the *E. coli* dataset includes a "start" and "end" date for the *E. coli* outbreak, we also include an interaction between *Treat* and a dummy variable *interim* that equals 1 if the house sale was between the start and end date, and 0 if it was not.

Table 3 Difference-in-Difference Results

	(1)	(2)	(3)
treat	-0.0621***	-0.0265***	-0.0215***
	(0.00676)	(0.00701)	(0.00609)
post	0.00912**	0.0157**	0.0433***
	(0.00367)	(0.00612)	(0.00547)
treat*post	0.0468***	0.00847	-0.0140
	(0.0108)	(0.0107)	(0.00960)
interim	0.0700***	-0.00688	0.0447***
	(0.00910)	(0.00899)	(0.00841)
treat*interim	0.0263	-0.0165	-0.0894***
	(0.0275)	(0.0262)	(0.0250)
Days Since Outbreak		-0.0000272***	-0.0000177***
		(0.00000331)	(0.00000550)
Distance to waterbody (m)		0.00000422**	0.00000287*
		(0.00000187)	(0.00000163)
Distance to waterbody squared (m)		-2.55e-10	-8.10e-11
		(1.85e-10)	(1.63e-10)
Constant	11.69***	10.45***	10.91***
	(0.00237)	(0.0137)	(0.0783)
Observations	195,331	160,324	160,304

Housing Characteristics	Yes	Yes	Yes
Adjusted R^2	0.001	0.276	0.446
Year/Month FE	No	Yes	Yes
School District FE	No	No	Yes
Outbreak FE			Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

In Table 3, Column (1) shows the results for the simple DID model. The coefficient on *treat* is negative and statistically significant ($p < .01$). Column (2) further controls for housing characteristics, days since the outbreak, and year and month fixed effects. Column (3) adds fixed effects for school districts and outbreaks. Based on our preferred specification in column (3), the coefficient on *(Treat * Post)* is -0.0104 but is not statistically significant. However, the coefficient on (*Treat*Interim)*, at -0.0894 and is statistically significant ($p < .01$). These two results imply that housing first prices see an immediate decrease of 8.94% following an *E. Coli* outbreak, but the decrease in prices is not permanent (since *treat*post* is not significant). In our sample the average house sells for $151,923, therefore an estimate of the average decrease in housing prices during the interim period is $13,582.

4.3 Robustness Checks

Recall that our results from table (3) used a cutoff radius of 1 mile. That is, every house within a mile of the outbreak was considered "treatment", while every house between 1 and 3 miles was considered "non-treatment". Table (4) shows the results for models that varies this cutoff. Column (1) uses a slightly smaller radius of 0.9 miles, while columns (2) and (3) use a radius of 1.1 and 1.2 miles, respectively. With the smaller radius, the model does not pick up the long-term effects of an *E. coli* outbreak, with the coefficient on *TREAT*POST* not significant.

However, models with a 1.1- and 1.2-mile cutoff showed a statistically significant coefficient on *TREAT*POST* at the 10% and 5% significance level, respectively. On the other hand, the short-term effect of *E.coli* on housing prices, captured by the variable *TREAT*INTERIM,* are significant regardless across all models ($p < .01$). The value of the coefficients imply a decrease in house sales prices of 7.46% to 8.65%, which are close to our estimate of 8.94% in table (3).

Table 4 Robustness Checks

	(1)	(2)	(3)
Treated group definition	Dist. to E Coli outbreak < 0.9 mile	Dist. to E Coli outbreak < 1.1 mile	Dist. to E Coli outbreak < 1.2 mile
treat	-0.0418***	-0.00952*	-0.00808
	(0.00665)	(0.00564)	(0.00528)
post	0.0419***	0.0438***	0.0449***
	(0.00545)	(0.00550)	(0.00553)
treat*post	-0.00153	-0.0151*	-0.0203**
	(0.0106)	(0.00884)	(0.00827)
interim	0.0421***	0.0462***	0.0466***
	(0.00834)	(0.00850)	(0.00860)
treat*interim	-0.0764***	-0.0865***	-0.0746***
	(0.0274)	(0.0229)	(0.0212)
Days Since Outbreak	-0.0000177***	-0.0000176***	-0.0000176***
	(0.00000550)	(0.00000550)	(0.00000550)
Distance to waterbody (m)	0.00000284*	0.00000286*	0.00000287*

	(0.00000163)	(0.00000163)	(0.00000163)
Distance to waterbody^2 (m)	-8.02e-11	-8.03e-11	-8.01e-11
	(1.63e-10)	(1.63e-10)	(1.63e-10)
Constant	10.91***	10.91***	10.91***
	(0.0783)	(0.0783)	(0.0783)
Observations	160,304	160,304	160,304
Housing Characteristics	Yes	Yes	Yes
Adjusted *R*2	0.446	0.446	0.446
Year/Month FE	Yes	Yes	Yes
School District FE	Yes	Yes	Yes
Outbreak FE	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.

An additional consideration is the source of water for each house. Although our dataset did not include this information, we attempted to proxy for it using several methods. First, we used a subset of houses that were within 1 mile of the Great Lakes. With the DID framework, we found a long-term impact of approximately -11%. We also ran a model with county fixed effects instead of school district fixed effects, which gave similar results to the main model.[6]

5 Conclusions and Discussion

In this article, we use point-level data and a quasi-experimental design along with a standard linear hedonic model to estimate the effects that proximity to an *E. coli* outbreak

[6] These results are available upon request.

exercises on housing prices. Our results show that *E. coli* outbreaks negatively affect housing prices by at least 8.9% for houses within one mile of the outbreak. However, these effects are very sensitive to time, and they decrease as days pass from the last outbreak affecting the transaction. The combination of these two factors possibly signals a lack of ‘memory’ by market actors, who tend to discount the effects of the outbreaks and/or their recursiveness once levels return to safe parameters. These results have several other implications for future state and local governments, both in Michigan and across the Great Lakes region.

Given the size of the damages by these recurring outbreaks, it may be cost-effective to design and implement policies addressing future outbreaks, especially in a state like Michigan, where water-related activities play a major social and economic role. Currently, there is no statewide TMDL for *E. coli*.[7] Instead, it is up to local municipalities to set regulations, which may result in conflicting policies and inefficient societal outcomes. Estimates such as ours can be beneficial for policymakers in assessing the costs and benefits of a statewide approach.

The policies of neighboring Midwest states demonstrate a range of policy options. In 2016, Ohio set revised statewide *E. coli* standards for wastewater discharge permits. These standards vary by recreational use of the receiving stream—bathing, swimming, and other primary uses cannot exceed a 90-day geometric mean of 126, while other uses that involve minimal contact cannot exceed a 90-day geometric mean of 1,030. Wisconsin seems to be moving in a similar direction as Ohio (Kaeding and The Associated Press, 2019). Indiana monitors water bodies for *E. coli*, but only provides information to local entities to develop

[7] See "https://www.michigan.gov/egle/0,9429,7-135-3313_3681_3686_3728-376271--,00.htmlfor more information on statewide efforts to implement a TMDL.

pollution reduction plans. Unfortunately, there is little research into the effectiveness of each policy, providing opportunities for future research.

Although we controlled as many factors as the data allowed, our results may be influenced by omitted variable bias. As the first estimate of the effect of *E. Coli* outbreaks on housing in Michigan, it is difficult to gauge how realistic the estimates are. Research using alternative statistical methods or updated data will help place our estimates in context. One possibility is for future researchers to separate waterbodies by type. Waterfront properties on the Great Lakes (Lake Michigan, Lake Huron, Lake Superior) may capture further benefits despite being subject to the effects of *E. coli* (see e.g., Colwell and Dehring, 2005; Wyman et al., 2020). In addition, communication of outbreaks to the public varies substantially across the state; thus, potential buyers may have been privy to information about the outbreaks prior to purchasing the property, while those buyers who are less inclined to purchase houses near outbreaks may simply look elsewhere ex-ante.

Open Research

The housing data information was obtained through a private agreement with Zillow and must remain confidential. Data on *E. Coli* outbreaks was obtained through the *Egle* department of Michigan and is available upon request.

References

Allen-Burton, G., Scavia, D., Luooma, S.N., Love, N.G., Austin, J. 2010. Leveraging the Great Lakes Region's water assets for economic growth. Metropolitan Policy Program, New York: Brookings Institute.

Bishop, K. C., Kuminoff, N. V., Banzhaf, H. S., Boyle, K. J., von Gravenitz, K., Pope, J. C., ... & Timmins, C. D. (2020). Best practices for using hedonic property value models to measure willingness to pay for environmental quality. *Review of Environmental Economics and Policy*.

Brashares, E.N. 1985. Estimating the Instream Value of Lake Water Quality in Southeast Michigan (Hedonics, Prices, Property, Demand, Benefits). University of Michigan Press, Ann Arbor.

Currie, J., Davis, L., Greenstone, M., Walker, R., 2015. Environmental health risks and housing values: Evidence from 1,600 toxic plant openings and closings. Am. Econ. Rev., 105(2), 678–709.

Colwell, P.F., Dehring, C.A. The Pricing of Lake Lots. J. Real Estate Finan Econ 30, 267–283 (2005). https://doi.org/10.1007/s11146-005-6407-x

Davies, C.M., Long, J.A., Donald, M., Ashbolt, N.J. 1995. Survival of fecal microorganisms in marine and freshwater sediments. Appl. Environ. Microbiol. 61(5), 1888-96.

Foppen, J. W. A., & Schijven, J. F. (2006). Evaluation of data from the literature on the transport and survival of Escherichia coli and thermotolerant coliforms in aquifers under saturated conditions. *Water Research*, *40*(3), 401-426.

Griffith, John F., Stephen B. Weisberg, and Charles D. McGee. "Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples." Journal of Water and Health 1.4 (2003): 141-151.

Haninger, K., Ma, L., Timmins, C., 2017. The value of brownfield remediation. J. Assoc. Environ. Resour. Econ, 4(1), 197– 241.

Heberling, M. T., Guignet, D., & Papenfus, M. (2024). Hedonic property values and water quality: A meta-analysis of commodity, market, and methodological choices. *Journal of Environmental Management*, *352*, 119829.

Ishii, S., Sadowsky, M.J. 2008. Escherichia coli in the environment: implications for water quality and human health. Microbes Env. 23(2), 101-108.

Johnston, R. J., Boyle, K. J., Loureiro, M. L., Navrud, S., & Rolfe, J. (2021). Guidance to enhance the validity and credibility of environmental benefit transfers. *Environmental and Resource Economics*, *79*(3), 575-624.

Kaeding, D., and The Associated Press, 2019. DNR Board Approves E. Coli Water Regulations. Wisconsin Public Radio. At https://www.wpr.org/dnr-board-approves-e-coli-water-regulations [04/15/2022].

Leggett, C., Bockstael, N. 2000. Evidence of the effects of water quality on residential land prices. J. Environ. Econ. Manag. 39, 121-44.

Meyer, K. (2020). Immersive virtual reality and willingness to pay. *Journal of Agricultural and Resource Economics*, *45*(3), 484-495.

Michigan Department of Environment, Great Lakes, and Energy. 2019. “Michigan’s statewide E. Coli Total Maximum Daily Load.” Available at https://www.michigan.gov/documents/deq/wrd-swas-ecoli-factsheet_507238_7.pdf.

Moore, M. R., Doubek, J.P., Hu, X., Cardinale, B.J., 2020. Hedonic price estimates of lake water quality: valued attribute, instrumental variables, and ecological-economic benefits. Ecol. Econ., 176, 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.

Papenfus, M. 2019. Do housing prices reflect water quality impairments? Evidence from the Puget Sound. Water Resour. Econ. 27, 100133.

Patz, J.A., Vavrus, S.J., Uejio, C.K., McLellan, S.L. 2008. Climate change and waterborne disease risk in the Great Lakes Region of the U.S. Am. J. Med. 35(5), 451-458.

Rabinovici, S.J.M, Bernknopf, R.L. Wein, A.M. Coursey, D.L., Whitman, R.L. 2004. Economic and health risk trade-offs of swim closures at a Lake Michigan Beach. ACS Publications. https://doi.org/10.1012/es034905z.

Rosen, S. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *J.* Political Econ. 82(1), 34-55.

Roslev, Peter, and Annette S. Bukh. "State of the art molecular markers for fecal pollution source tracking in water." Applied microbiology and biotechnology 89 (2011): 1341-1355.

Scott., C.A., Pierce, S.A., Pasqualetti, M.J., Jones, A.L., Montz, B.E., Hoover, J.H. 2011. Policy and institutional dimensions of the water-energy nexus. Energy Policy 39, 6622-6630.

Spanou, E.,vKenter, J.O., Graziano, M., 2020. The Effects of Aquaculture and Marine Conservation on Cultural Ecosystem Services: An Integrated Hedonic – Eudaemonic Approach. Ecol. Economics 176, 106757.

Steinman, A.D., Cardinale, B.J., Munns, W.R., Jr., Ogdahl, M.E., Allan, J.D., Angaldi, T., Bartlett, S., Brauman, K., Byappanahalli, M., Doss, M., Dupont, D., Johns, A., Kashian, D., Lupi., F., McIntyre, P., Miller, T., Moore, M., Muenich, R.L., Poudel, R., Price, J., Provencher, B., Rea, A., Read, J., Renzetti, S., Sohngen, B., Washburn, E. 2017. Ecosystem services in the Great Lakes. J. Gt. Lakes Res. 43(3), 161-168.

Tanaka, S., Zabel, J., 2018. Valuing nuclear energy risk: Evidence from the impact of the Fukushima crisis on U.S. house prices. J. Environ. Econ. Manage., 88, 411–426.

Weiskerger, C.J., Whitman, R.L. 2018. Monitoring *E. coli* in a changing beachscape. Sci. Total Environ. 619-620, 1236-1246. https://doi.org/10.1016/j.scitotenv.2017.11.167.

Whitman, R.L, Nevers, M.B. 2003. Foreshore sand as a source of Escherichia coli in nearshore water of a Lake Michigan Beach. Appl. Environ. Microbiol. 69(9), 5555-62.

Wilson, M.A., Carpenter, S.R. 1999. Economic valuation of freshwater ecosystem services in the United States: 1971-1997. Ecol. Appl. 9(3), 772-783.

Wolf, D., Klaiber, H.A. 2017. Bloom and bust: Toxic algae's impact on nearby property values. Ecol. Econ. 135, 209-221.

Wyman, D., Hutchinson, N., and Tiwari, P., 2020. Testing the waters: a spatial econometric pricing model of different waterfronts views. J. Real Estate Res., 36(3), 363-3682.

Appendix

Full sample results

The main results of the paper come from a subsample of Zillow housing prices. Table (5) shows the results using the full sample.

Table 5 DID results

	(1)	(2)	(3)
Treat=1	0.0476***	-0.0194**	-0.0470***
	(0.00634)	(0.00652)	(0.00513)
Post=1	0.0672***	0.00661	0.0330***
	(0.00270)	(0.00519)	(0.00451)
Treat*Post	-0.0629***	-0.0382***	-0.0203*
	(0.00896)	(0.00924)	(0.00788)
Interim=1	0.107***	-0.0172*	0.0344***
	(0.00799)	(0.00811)	(0.00718)
Treat*Interim	0.0963**	-0.0540*	-0.0492*
	(0.0333)	(0.0261)	(0.0209)
Days Since Outbreak		0.0000294***	-0.0000120**
		(0.00000276)	(0.00000461)
Lot Size (sq. ft.)		6.01e-09	3.87e-09***
		(3.17e-09)	(8.09e-10)

Bedrooms		0.0293^{***}	0.00770^{***}
		(0.000968)	(0.000866)
Bathrooms		0.311^{***}	0.158^{***}
		(0.00324)	(0.00181)
Stories		0.0805^{***}	0.0529^{***}
		(0.00332)	(0.00196)
Building Area (sq. ft.)		0.0000337^{***}	0.0000279^{***}
		(0.00000523)	(0.000000196)
Observations	471,535	341,584	341,546
Adjusted R^2	0.002	0.160	0.377
Year/Month FE	No	Yes	Yes
School District FE	No	No	Yes
Outbreak FE	No	No	Yes

Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$