Out-of-sight, Out-of-mind? Distance Decay in Use and Non-use Value of Water Quality

Social Cost of Water Pollution Workshop

Nathaniel Merrill, Emmett Reynier, and Xibo Wan
October 2024

[html version]

Water quality valuation

The US EPA relies on a **metaanalysis-based benefit transfer function** to estimate the benefits of water quality improvements (Corona et al. 2020).

- Limits extent of market to a given radius around water features (100mi)
- WTP is constant within that radius
- Based on 1996 National Survey on Recreation, where about **80% of all recreational uses** were within 100mi (Viscusi et al. 2008)
- Known to be a conservative assumption, leaving benefits on the table

More generally...

- **Travel cost models** can inform us about the decay of **use values**
- We know **considerably less** about how distance affects **non-use values**

Our approach

Q What is the relationship between distance and WTP for water quality? Does that relationship differ for use and non-use values?

- Nest a **travel cost random utility model** within a utility function that is also **directly affected by water quality**, the value of which is impacted by the distance between a household and the water feature
- Estimate travel cost RUM with **cell phone-based mobility data** on trips to water-based recreation sites
- Combine with several **stated preference studies** (currently Johnston et al. 2023 and Vossler et al. 2023) to estimate rest of utility parameters

Preliminary Results

- **Use values decay quickly**, over 90% of benefits captured within 100mi
- **Non-use value decays much slower**, and potentially not at all

Structural decomposition of use and non-use values

Structural Model Details

Kim and Lupi (2023) propose a method for combining stated and revealed preferences to decompose use and non-use values.

- Our goal is to take this model to survey data from **many studies**
- Fill in missing visitation data with cell phone-based mobility data

Person i's utility conditional on visiting water feature j in month t and state of the world s is

$$u_{ijts} = v_{ijts}^{\text{use}} + v_{its}^{\text{nonuse}} + v_{ijts}^{o} + \varepsilon_{ijts}$$

Numeraire is linear in income net of travel cost p_{ij} and program cost ${c_s}^{\dagger}$

$$v_{ijts}^{o} = \gamma(y_{it} - p_{ij} - c_s)$$

† The baseline case has $c_0 = 0$.

Structural Model Details

Use Value If person i decides to visit site j, the utility they get is an alternative specific constant α_{jt} plus a vector representing site quality q_{js}

$$v_{ijts}^{\text{use}} = \alpha_{jt} + q_{js}'\beta$$

Thus, β reflects the marginal use utility of quality.

Non-use Value Non-use utility is the distance weighted sum across all sites,

$$v_{its}^{\text{nonuse}} = \sum_j (d_{ij} + 1)^\lambda (a_{jt} + q_{js}'b)$$

where d_{ij} is distance between i and j, λ is the rate of decay, and b is the marginal non-use utility of quality. **Does not depend on actual site choice!**

People **visit the sites** with the highest indirect utility. They then vote on the proposed policy from the sum of their **expected use and non-use utility**.

Structural Model Details

Welfare measures

We can calculate WTP for a given water quality change $\boldsymbol{\Delta q}$ as

$$WTP_i^{\text{use}} = \frac{1}{\gamma} \sum_t \log \left[\frac{\sum_j \exp\left(\tilde{\alpha}_{jt} + \Delta q_j' \beta - \gamma p_{ij}\right)}{\sum_j \exp\left(\tilde{\alpha}_{jt} - \gamma p_{ij}\right)} \right]$$

$$WTP_i^{\text{nonuse}} = \frac{1}{\gamma} \sum_t \sum_j (d_{ij} + 1)^{\lambda} \Delta q_j' b$$

both of which vary as respondents get further from a given water feature.†

† Travel costs p_{ij} are a non-linear function of driving distance

Estimation Strategy

Estimation Strategy

We can make distributional assumptions to derive a likelihood function and estimate the parameters via MLE---this requires four key pieces of data

1. Characteristics of the proposed policy (i.e. cost and Δ water quality)
2. **Distance** between respondents and water quality improvements
3. Respondent's **vote yes/no for the proposed program**
4. Respondent's **site visitation behavior**

Authors of stated preference studies should have 1-3 with **relatively low effort**, but most do not have **specific site visitation data**.

$\implies$ We want to use **cell data** to fill in missing visitation behavior

[RP details] [SP details]

Our Approach

We'd like to jointly estimate all of the parameters in the model. However, we have first done things in two steps to get preliminary results:

1. Estimate mean utilities, $\delta_{jt} = \alpha_{jt} + q_j'\beta$ and the marginal utility of income, γ, using **travel cost RUM with cell data**
2. Take $\hat{\delta}$'s and $\hat{\gamma}$ as given in **MLE with stated preference data** to estimate rest of utility parameters $\theta = \{\beta, b, \lambda, \sigma^{SP}\}$

$$\tilde{v}_{is} = \kappa + \left(v_{is}^{nonuse} + E\left[v_{is}^{use} + v_{is}^{o}\right]\right)\frac{1}{\sigma^{SP}}$$

where $E\left[v_{is}^{use} + v_{is}^{o}\right] = \sum_t \log \sum_j \exp\left(\hat{\alpha}_{jt} + q_{js}'\beta + \hat{\gamma}\left(y_i - p_{ij} - c_s\right)\right)$

Data

Cell phone mobility data

Patterns data from Advan (formerly SafeGraph)

- Aggregated monthly visits from CBGs to POIs
- Can access data between 2019 to 2022--focus on 2022 to avoid Covid
- Privacy protection measures may cause econometric issues [details]
- Intersect Advan park POIs with river and lake shapes from NHDv2 [details]
- Gives us **30,950 lake-based** sites and **36,920 river-based** sites [map]
- Aggregate sites to HUC12 level

Dealing with zero share issues

1. **Unbalanced Panel**: Drop CBG-HUC12-month's with zero visits
2. **Balanced Panel**: Use empirical Bayes estimator as in Li (2023) or add a small number of visits to zero-visit sites

We calculate **60 million** driving distances between CBG and HUC12 centroids using the Open Source Routing Machine (OSRM) [details] [travel cost function]

Stated preference data

We are greatful to the authors of two studies who have shared microdata from their studies with us, Johnston et al. (2023) and Vossler et al. (2023).

Johnston et al. (2023) [details]

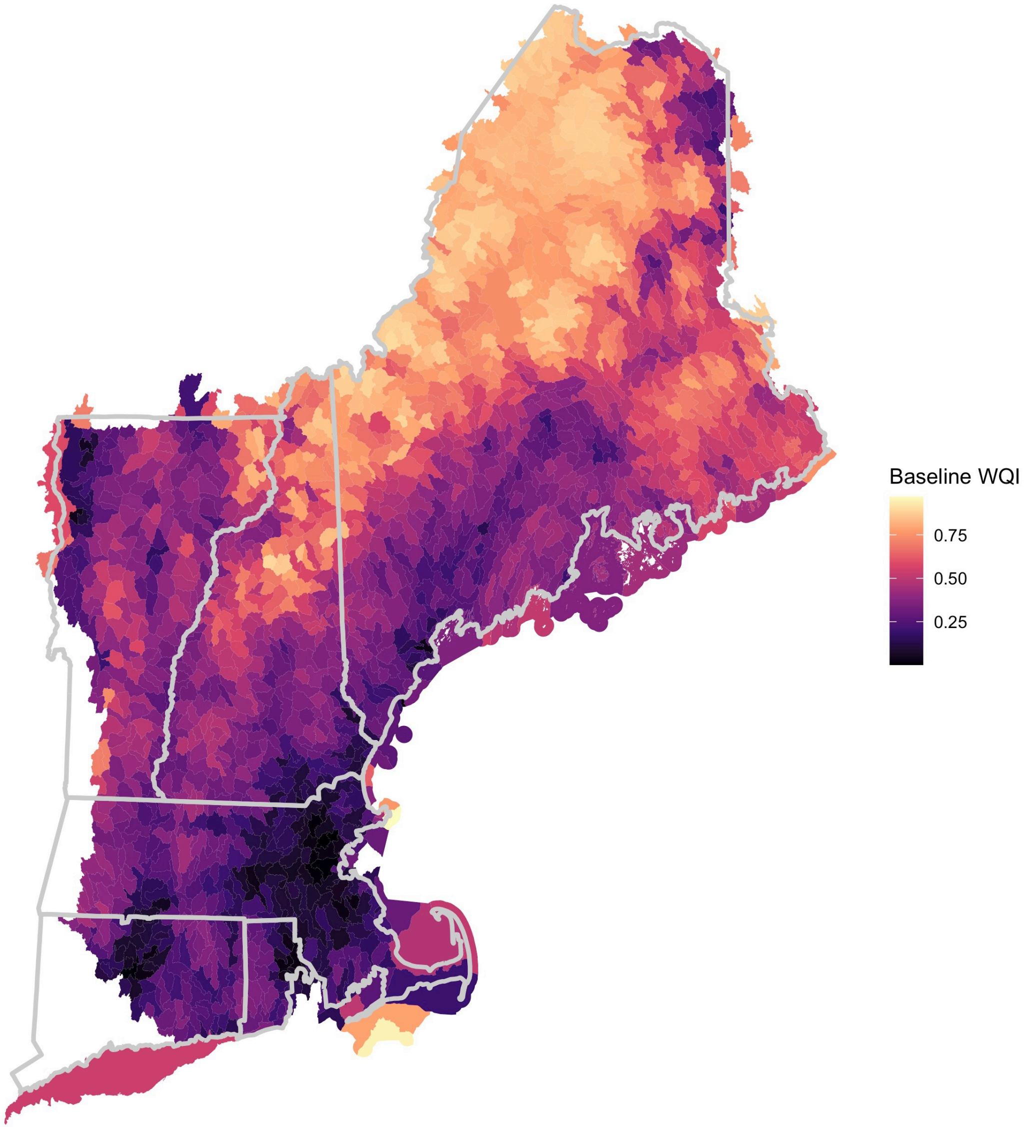
- Focused on New England
- Survey instrument compares BAU to alternative policy scenarios in 2025
- Model three different water quality metrics using FRAMES
- Innovative analysis using data on where people zoomed in on maps

Vossler et al. (2023) [details]

- Focused on the upper-midwest
- Values improvements in the Biological Condition Gradient (BCG)
- Explore value for local vs non-local improvements

Today's results only use the Johnston et al. data.

Stated preference data



Estimation Results

Estimation Results: Travel Cost RUM

Estimates for the **marginal utility of income** γ using cell data:

	(1)	**(2)**	(3)	(4)	(5)	(6)
Travel Cost	-0.015*	**-0.021***	-0.025*	-0.018*	-0.011*	-0.004*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	493,413	1,520,726	1,520,726	1,520,726	1,520,726	1,520,726
N (Zipcode)	727	727	727	727	727	727
Unbalanced	Yes	No	No	No	No	No
Share	-	Bayes	1-e6	1-e5	1-e4	1-e3

- Use GMM estimator to solve the market share system
- Instruments for travel costs with interaction between state and crude oil price, and travel distance
- Get similar results modeling γ_i as a random coefficient

* significant at 1% level

Estimation Results: Stated Preferences

Estimate rest of parameters via maximum likelihood taking results from travel cost model as given. Restrict $\lambda <= 0$

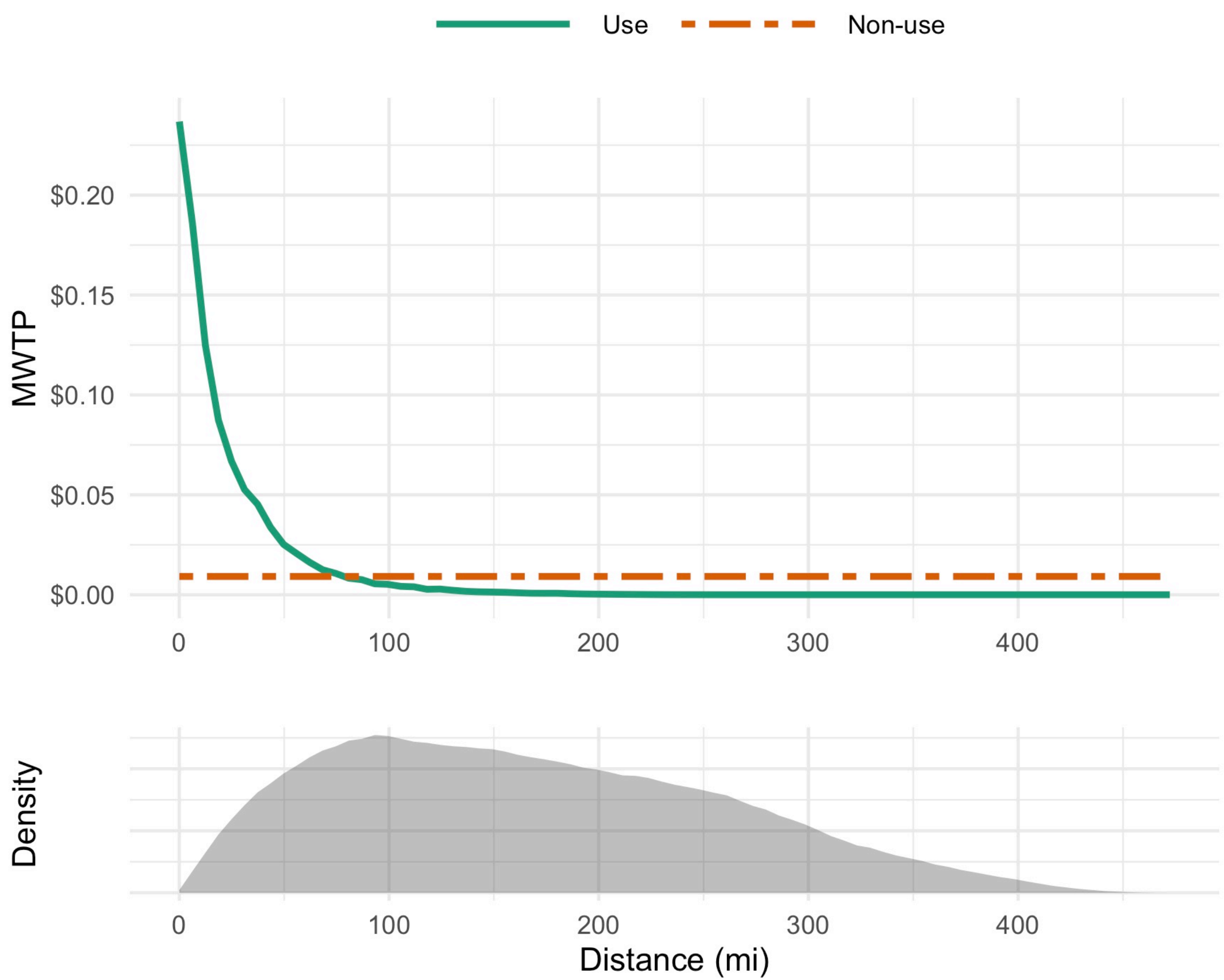
Parameter	Estimate
β	93.68
b	0.09
λ	-3.3e-9
σ^{RP}	55.90

To put these results into context,

- Johnston et al. (2023) find **$12.2** WTP for 1pt improvement in region-wide WQ, our model suggests **$26.4** the same region
- Corona et al. (2020) find **$3.1** WTP for 1pt improvement in a HUC4, median of 1pt improvements across HUC4's in our model is **$2.4**

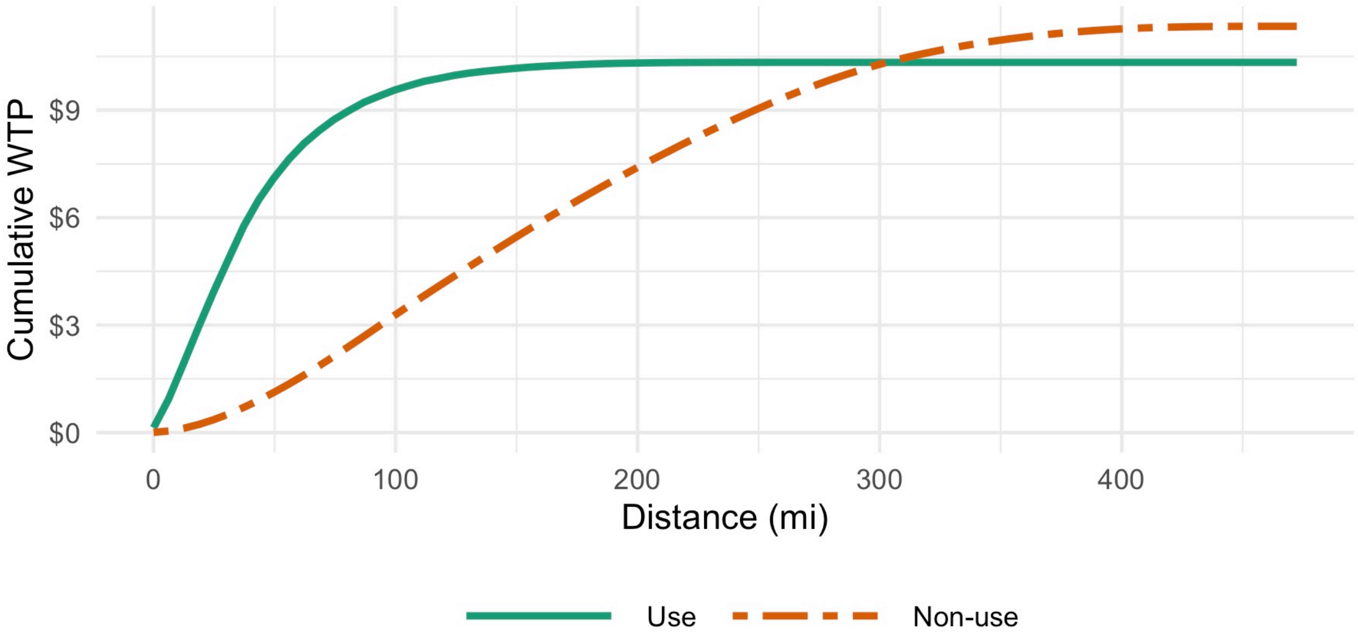
Results

MWTP for 1pt WQ in a HUC12



Cumulative WTP for 1pt WQ by distance

We can plot the cumulative WTP for survey respondents as we get further from the HUC12 seeing a water quality improvement,



At 100 miles, we caputure **92%** of the total use value, but only **29%** of the non-use value, or **59%** of total WTP.

Conclusions

Implications for benefits estimation

- Non-use values are low, may not decay much with distance
- The radius-based extent of market would severely understate benefits

Next steps

- Estimate all parameters jointly
- Integrate more stated preference studies, and figure out how to handle the host of estimation complexities that brings
- Specify richer, more flexible utility model

Appendix

Model Details: Revealed Preferences

Assume that trips occur in the baseline period, $s = 0$ before quality change. Then the probability i visits j is

$$\begin{aligned} Pr^{visit}_{ijt0} &= Pr(u_{ijt0} - u_{ikt0} > 0) \quad \forall\, k \neq j \\ &= Pr\left[(v^{\text{use}}_{ijt0} + v^{\text{nonuse}}_{it0} + v^{o}_{ijt0}) - (v^{\text{use}}_{ikt0} + v^{\text{nonuse}}_{it0} + v^{o}_{ikt0}) > \varepsilon_{ikt0} - \varepsilon_{ijt0}\right] \\ &= Pr\left[(v^{\text{use}}_{ijt0} + v^{o}_{ijt0}) - (v^{\text{use}}_{ikt0} + v^{o}_{ikt0}) > \varepsilon_{ikt0} - \varepsilon_{ijt0}\right] \quad \forall\, k \neq j \end{aligned}$$

The non-use values cancel out since they do not depend on choice j. Given T1EV errors, we have

$$Pr^{visit}_{ijts} = \frac{\exp\left(\alpha_{jt} + q'_{js}\beta - \gamma(y_{it} - p_{ij} - c_s)\right)}{\sum_k \exp\left(\alpha_{kt} + q'_{ks}\beta - \gamma(y_{it} - p_{ik} - c_s)\right)}$$

[return]

Model Details: Stated Preferences

Stated preference portion

The utility i receives from state of the world s is

$$u_{is} = \tilde{v}_{is} + \tilde{\varepsilon}_{is} = T\tilde{v}_{is} + \sum_t \tilde{\varepsilon}_{its}$$

where $\tilde{v}_{is}$ is the deterministic portion of utility,

$$\tilde{v}_{is} = \frac{1}{\sigma^{SP}} \left(v_{is}^{\text{nonuse}} + \sum_j \exp\left(v_{ijs}^{\text{use}} + v_{ijs}^{o} \right) \right) + \kappa$$

σ^{SP} is the variance of T1EV shock and the left terms are expected use-utility in state s with constant of integration κ. Household i will choose the change scenario $s = 1$ with probability

$$Pr_{i1}^{vote}(u_{i1} > u_{i0}) = Pr\left(\tilde{v}_{i1} - \tilde{v}_{i0} > \tilde{\varepsilon}_{i1} - \tilde{\varepsilon}_{i0}\right)$$

[return]

Likelihood Function

Calculate log-likelihood as

$$ll_i = \sum_t \sum_j Y_{ijt} \log Pr_{ijt}^{\text{visit}} + \sum_s Y_{is} \log Pr_{is}^{\text{vote}},$$

where Y_{ijt} and Y_{is} are the actual visit choices and vote choices respectively.

[return]

Advan privacy protection measures

Advan implements the following procedure to protect user's privacy:

1. Differential privacy: add a Laplane noise with $\boldsymbol{N(0, 10/9)}$
2. Discretization: round the number to integer
3. Integer Floor: present a visitor count if $\boldsymbol{floor(V_{ijt})}$ is larger than $\mathbf{2}$
4. Censoring: $\max\{4, \text{floor}(V_{ijt})\}$

In-progress work suggests that these techniques may lead to bias using standard estimation techniques.

[return]

Identifying Water POIs

We intersect the Advan parks polygon (NAICS code 712190) with the NHDv2:

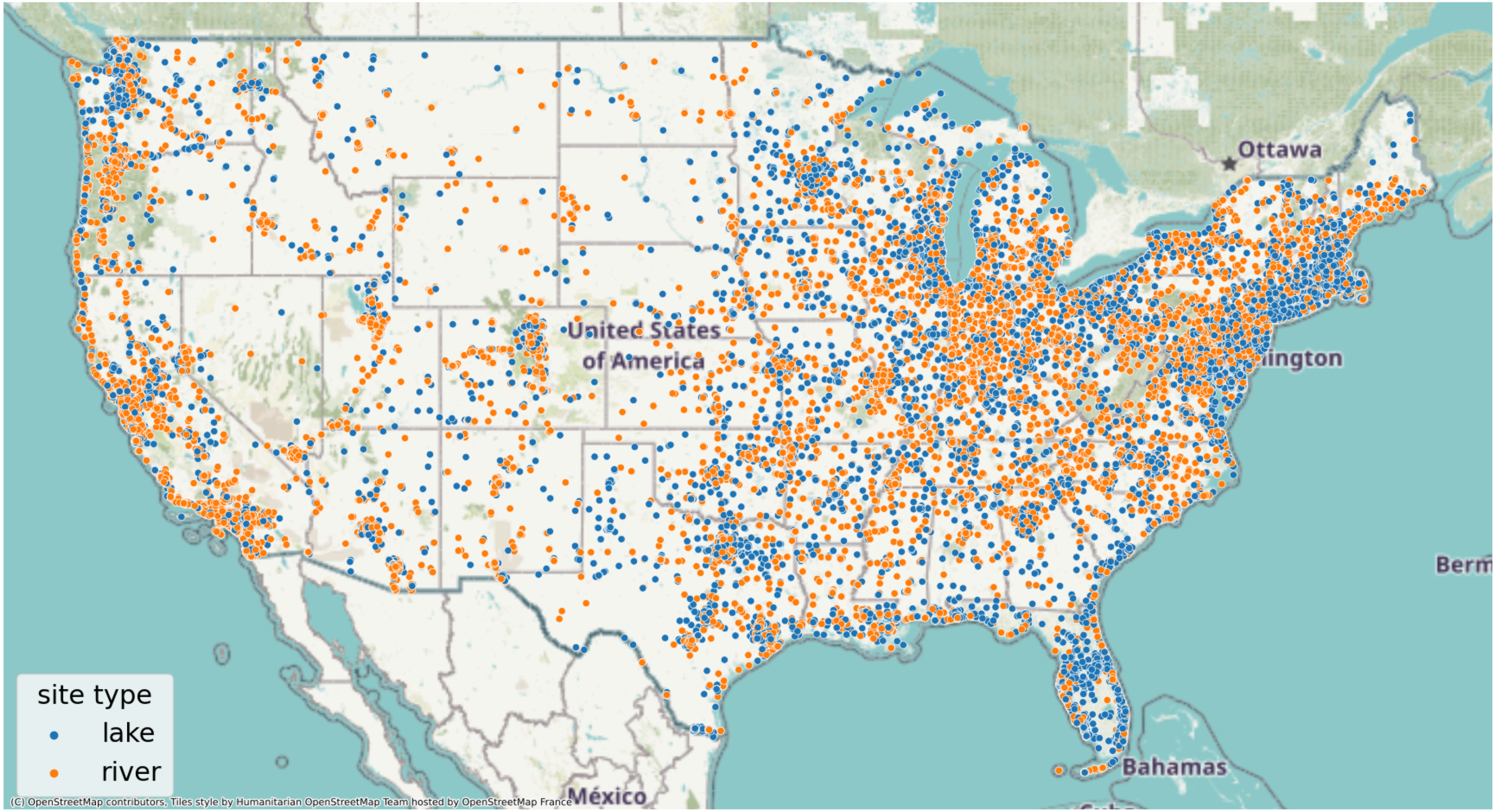
- Create 50 meter buffer around park polygons
- Create 20 meter buffer around river flowlines
- Identify lake-based sites as those adjacent to a water body

A few important caveats

1. Difficult to identify river visits (esp. river sites across states)
2. Aggregation of sites to HUC12 may be a concern: the partial aggregation approach reasonably approximates the results of a disaggregate model (Lupi and Feather (1988), Daniel Phaneuf's recent work with EPA grants)
3. Represents approximately 10% of all smartphones
4. Tends to represent a younger demographic than the general population
5. Mobile GPS positioning may have ~5 meter errors

[return]

Map of Water-Based POIs

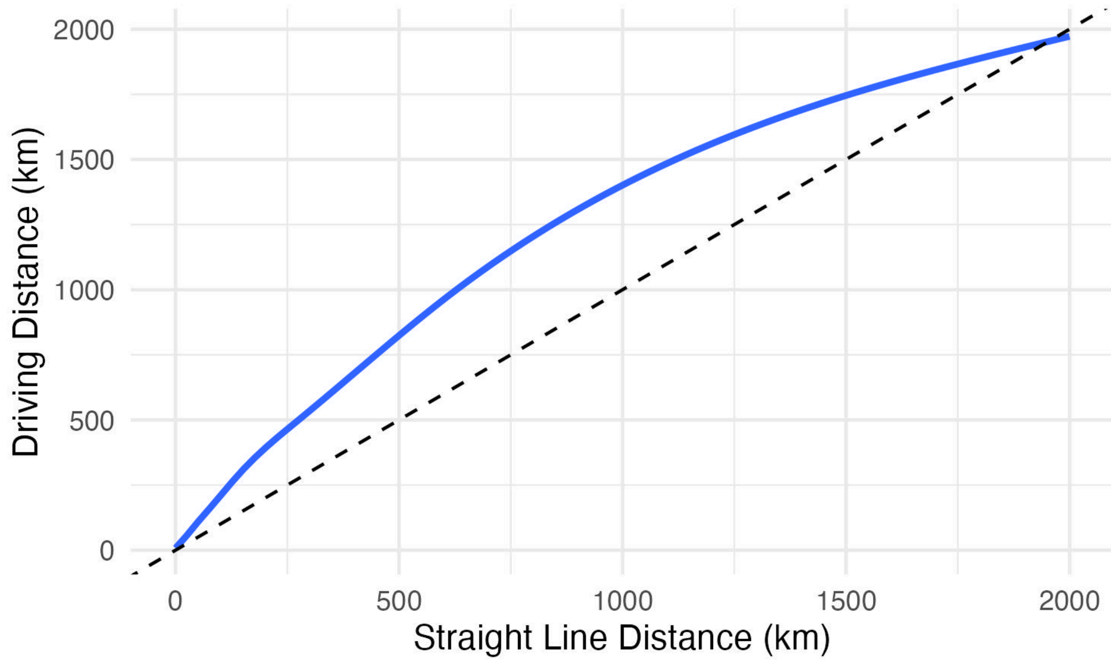


[return]

Driving Distances with OSRM

We use the Open Source Routing Machine (OSRM) to calculate the shortest driving distance between CBG and water-based POI site centroids.

- Getting OSRM to run nationally requires ~100gb of RAM in inital setup (used virtual machine on Google Compute Engine)
- Fed it 60 million lat/lon pairs and waited for 3 days...



[return]

Calculating Travel Costs

We use the following formula to calculate travel costs:

$$TC_{ijt} = 2\left(gs_{it} + f_t\right)\mathrm{Dist}_{ijt} + 2\gamma\frac{\mathrm{Medinc}_i}{2080}\mathrm{Time}_{ijt}$$

where...

- $\mathbf{Dist}_{ijt}$ and $\mathbf{time}_{ijt}$ are one-way travel distance and time between the centroid of CBG i and HUC12 j in period t
- gs_{it} is state-level average gas cost in year t
- f_t is marginal vehicle maintenance cost, repair cost, and depreciation
- $\frac{\mathrm{Medinc}_i}{2080}$ is median annual income CBG i, converted to hourly wages
- $\gamma = \frac{1}{3}$ for central estimate, with robustness checks
- Travel time: replace missing values with travel distance/38.5

[return]

Reduced form results with cell data

We begin with some simple regressions measuring distance decay in use values with the cell-phone data:

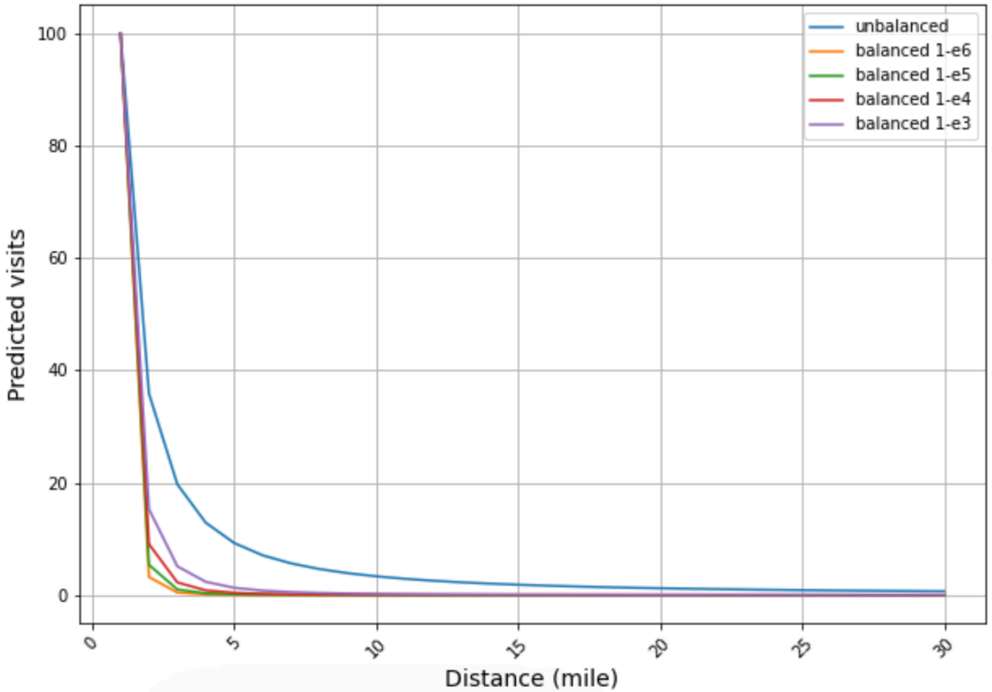
$$\log \mathrm{Visit}_{ijt} = \beta \log d_{ij} + \mathbf{zip}_i + \mathbf{site}_j + \mathbf{ym}_t + \varepsilon_{ijt}$$

where

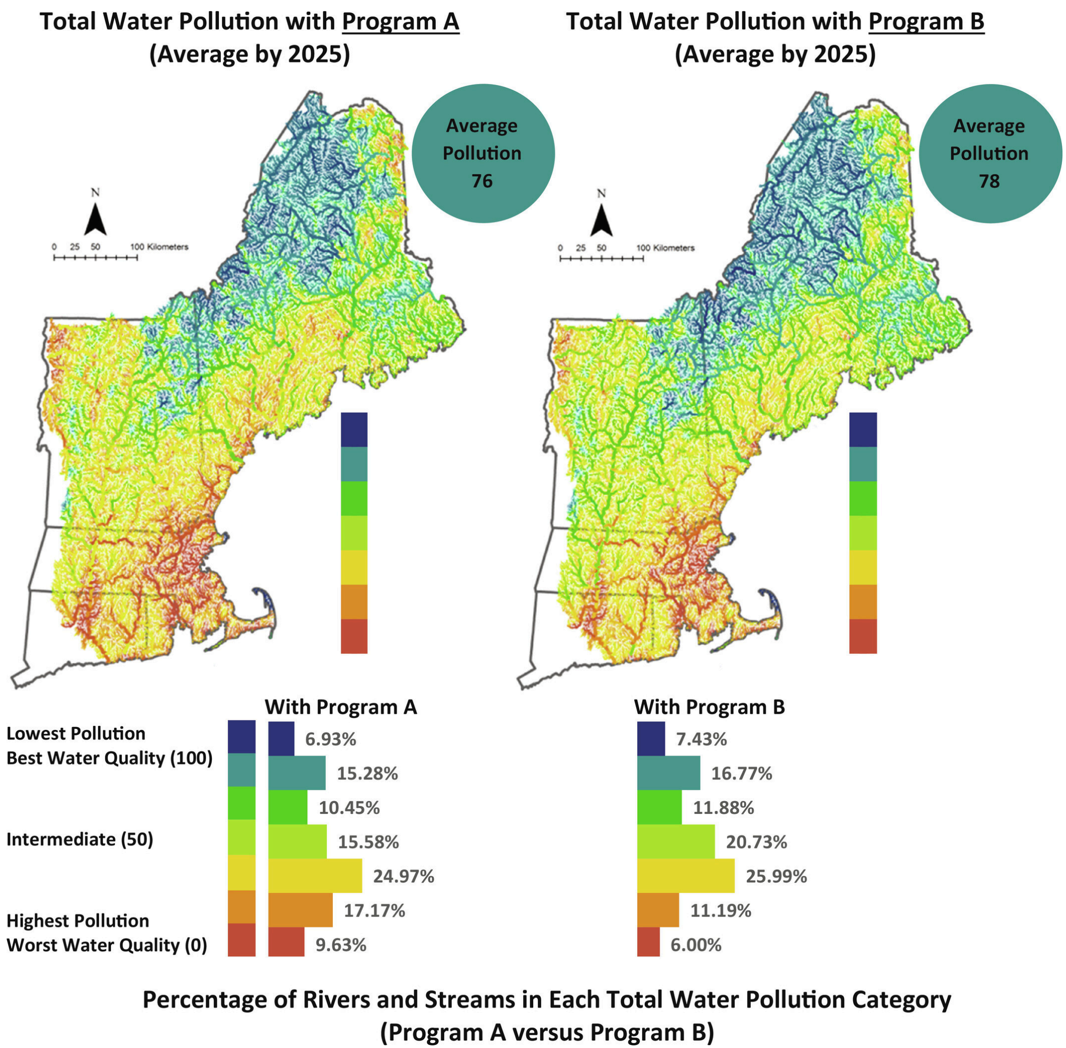
- $\mathbf{Visit}_{ijt}$ are monthly visits (normalized by number of devices in origin)
- d_{ij} is driving distance from zip code i to a HUC12 j
- $\mathbf{zip}_i$ and $\mathbf{site}_j$ are origin and destination fixed effects
- $\mathbf{ym}_t$ are year-month fixed effects

Reduced form results

Elasticity	-1.479	-4.973	-4.219	-3.464	-2.711
observation	493,413	1,520,726	1,520,726	1,520,726	1,520,726
unbalanced	Yes	No	No	No	No
number	-	1-e6	1-e5	1-e4	1-e3



Johnston et al. PNAS 2023 details



[return]

Vossler et al. PNAS 2023 details

[return]