nature computational science

Article

https://doi.org/10.1038/s43588-023-00572-6

Quantifying spatial under-reporting disparities in resident crowdsourcing

In the format provided by the authors and unedited

1 Detailed Related Work

This section discusses the literature on equity and efficiency in crowdsourcing systems, and briefly outlines our contribution to this literature. Given the importance of 311-like systems in allocating government services, there has been much interest in quantifying how efficient and equitable these systems are. Previous works mainly aim to answer two questions: (1) is *reporting* behavior similar across socioeconomic groups, and (2) do governments *respond* to the reporting equitably, or do they prioritize some neighborhoods for the same types of requests? If there are disparities in either stage of the process, then government services may be allocated inequitably.

Towards the second question, there has been a long line of work since the late 1970s, documenting how the allocation of government services has changed since the adoption of co-production systems. Early works warned of the potential of biases in how governments responded [12, 36]. Researchers have found response time differences between neighborhoods, e.g., during Hurricane Katrina [8]; other recent studies suggest that the practical differences induced by these differential response times are negligible [7] or explained by other factors [37].

This work considers the first question, of how reporting behavior varies across demographic and incident characteristics. The literature here focuses on two types of reporting behavior: *under-reporting*, and *misreporting*. The latter refers to when residents report problems that, upon inspection by the agency, are not found or are less severe than reported; there is some evidence that such misreporting occurs and is heterogeneous across areas [24].

On the other hand, differential *under-reporting* is when people, faced with similar problems, differentially report those problems to the government. Such differences might emerge, for example, due to access to communications technologies, familiarity with the system, or trust in government [3]. O'Brien et al. [28] and O'Brien [27] measure differences in the delay of reporting, as an indicator of the citizens' custodianship of their environment. They do so by measuring the delay from when researchers observe the incident during a street audit, to when the incidents are reported by the public.

As discussed above, the biggest challenge to studying *under-reporting* is that researchers rarely directly observe an incident *unless it is reported*. It is thus difficult to disentangle whether certain areas have fewer reports because problems truly occur less frequently there, or whether people in those areas are reporting less frequently given a similar distribution of problems.

One line of work in answering this question does not attempt to distinguish between these two possibilities. Such work entails regressing the number of reports (or the number of unique incidents, if some incidents are reported multiple times) as a function of socioeconomic and report characteristics, as well as potentially space and time [5, 6, 25]. Such works have found that the number of reports may (but does not always) differ by wealth, race, and education level. However, as we formalize in Supplementary Information 2.2, such a method cannot identify whether these fewer reports are due to fewer true incidents, or less reporting for the same incidents. Rather, these works must assume that the latter is the cause. Akpinar et al. [1] point out that such an assumption may not always hold, affecting downstream models, in the context of crime reporting and predictive police systems.

Another approach to quantifying under-reporting is to construct a proxy for the true incident rate and then compare the estimates with the observed rates. Kontokosta and Hong [16] analyze pothole complaints in Kansas City, Missouri, at a fine-grained spatial level; they leverage additional street assessment data (resulting from scheduled visual inspections) that grade the quality of roadways to construct relative predictions for the true number of potholes. They find that low-income and minority neighborhoods are less likely to report street conditions or "nuisance" issues while prioritizing more serious problems. Hacker et al. [10] use a similar metric, with the addition of

temporal trends geared more towards an epidemiological study, and Pak et al. [30] use roadway length as part of their proxy. Kontokosta et al. [15] analyze residential building problem complaints in New York City's 311 system—using building conditions as a proxy—and also find that many socioeconomic factors contribute to differential reporting behavior. O'Brien et al. [28] use administrative records from active broken streetlight inspections. While the approach can disentangle reporting behavior from true incident rates, it heavily relies on the accuracy of the 'ground-truth' proxy estimate. It may be difficult to find such proxies or to validate their accuracy, especially for different types of incidents within the same class (e.g., more vs less serious potholes), or more generally for high dimensional types: for example, estimating ground truth rates in every census tract for every kind of incident. In particular, note that each of the above examples estimates ground truth rates for a specific type of incident, in a specific time period.

The challenge between identifying occurrence rates as opposed to reporting rates is also present in the ecology literature when counting the number of animals of a given species in an area. Some works similarly use proxy methods – such as by normalizing by occurrence rates in neighboring geographies or using reference species whose occurrence rates are approximately known [11, 13, 32].

To this literature, our work contributes a statistical technique to identifying under-reporting—that does not require such external data or the construction of proxies for the ground truth incident rate. Rather, it relies only on report data already logged by many agencies and connects this task to a large literature on Poisson rate estimation. In applying our method to 311 data, we find reporting rate differences across neighborhoods associated with socioeconomic status and show that these differences are beyond what one would expect from just incident-level characteristics.

2 Technical results and proofs

2.1 Non-identifiability of reporting rate with observed incidents count

Proposition 1. *Consider the simplest setting: both incident occurrence and reporting follow time-homogeneous Poisson processes, with $\Lambda_\theta(\tau) = \Lambda_\theta$ and $\lambda_\theta(\tau) = \lambda_\theta$. Further, let incident reporting duration T_i be distributed according to F. Then using just the number of observed incidents $N_\theta^{observed}(T)$ of type θ in a known time duration $[0, T]$, the reporting rate λ_θ is not identifiable. In other words, $N_\theta^{observed}(T)$ is a function of both Λ_θ and λ_θ:*

$$\lim_{T\to\infty} \frac{N_\theta^{observed}(T)}{T} = \Lambda'_\theta$$
$$\textit{where} \qquad \Lambda'_\theta = \Lambda_\theta \left[1 - \int_0^\infty \exp\left(-\lambda_\theta t\right) dF(t)\right].$$

The proof of Proposition 1 follows directly from the following Lemma.

Lemma 1. *Suppose each incident gets reported independently, and the distribution of the interval of reporting duration T_i has density function $f(\cdot) : [0, \infty) \mapsto \mathbb{R}$. Then under steady state, $N_\theta^{observed}$ follows a Poisson process with rate Λ'_θ, where*

$$\Lambda'_\theta = \Lambda_\theta \left[1 - \int_0^\infty \exp\left(-\int_0^t \lambda_\theta(u)du\right) f(t)dt\right].$$

In the simplest time-homogeneous case, this rate simplifies to:

$$\Lambda'_\theta = \Lambda_\theta \left[1 - \int_0^\infty \exp\left(-\lambda_\theta t\right) f(t)dt\right].$$

Proof of Lemma 1. Let $m(t)$ be the number of times an incident is reported in an interval of t, starting from its birth. We know from the model assumption that $m(t)$ follows a Poisson distribution:

$$m(t) \sim \text{Poisson}\left(\int_0^t \lambda_\theta(u)du\right).$$

In steady state, each unique incident gets reported with probability

$$\begin{aligned} p &= \int_0^\infty \Pr[m(t) \geq 1 | T_i = t] f(t)dt \\ &= \int_0^\infty \left[1 - \exp\left(-\int_0^t \lambda_\theta(u)du\right)\right] f(t)dt \\ &= 1 - \int_0^\infty \exp\left(-\int_0^t \lambda_\theta(u)du\right) f(t)dt, \end{aligned}$$

which simplifies under time-homogeneity to

$$p = 1 - \int_0^\infty \exp\left(-\lambda_\theta t\right) f(t)dt.$$

Over a time interval of t, the total number of incidents of type θ that happen, $N_\theta(t)$ follows a Poisson process with rate Λ_θ. Conditional on $N_\theta(t) = n, n = 0, 1, \ldots,$ under steady state, $N_\theta^{\text{observed}}(t)$ follows a binomial distribution with parameters (n, p). Thus $N_\theta^{\text{observed}}(t)$ follows a Poisson distribution with rate $\Lambda_\theta p$, which completes the proof.

□

Proof of Proposition 1. Lemma 1 establishes that the rate at which we observe unique incidents depends on a lot of various aspects. Under steady state, $N_\theta^{\text{observed}}$ follows a Poisson process with parameter Λ'_θ, where Λ'_θ is a function of the incident occurrence rate Λ_θ, the (potentially non-homogeneous) reporting rate $\lambda_\theta(\cdot)$ and the distribution of reporting duration $f(\cdot)$.

In practice, when the observation period length T is large, we can safely assume that for each period $[\tau, \tau+1), \tau = 0, \ldots, T$, the observed reports $N_\theta^{\text{observed}}([\tau, \tau+1])$ are close to steady state, and thus follow independent and identical Poisson(Λ'_θ) distribution. Following the law of large numbers we get

$$\lim_{T\to\infty} \frac{N_\theta^{\text{observed}}(T)}{T} = \lim_{T\to\infty} \frac{\sum_{\tau=0}^{T-1} N_\theta^{\text{observed}}([\tau, \tau+1])}{T} = \mathbb{E}\left[N_\theta^{\text{observed}}(1)\right] = \Lambda'_\theta.$$

Thus, if we are only using information about the unique incidents, it is impossible to determine λ_θ without having full knowledge about both Λ_θ and $f(\cdot)$.

□

2.2 Identifiability of reporting rate with duplicate reports

In this subsection, we restate Theorem 1 formally in the notation of stochastic processes. Then, we show that when an observation interval is properly defined, evaluating the likelihood of the data within this observation interval identifies the reporting rate.

The core challenge is that, because we do not observe true incident birth (and potentially true incident death), our "observation interval" – when we observe the Poisson reporting process – *is itself random and dependent on the Poisson reporting process.* For example, we can only start the counting interval at the time of the first report, and the agency responds to the incident partially as

a function of the number of reports. Thus, the distribution of the entire data, not just the number of reports within the interval (i.e., including start and stop times) could be a complex, unknown function of λ. The theorem formalizes that, as long as the start and end times depend only on the rate λ through the number of reports up to those times, we can nevertheless decompose the likelihood and reduce the task to Poisson rate estimation.

Formally, we separate terms involving λ from terms independent of λ, which requires the two conditions; then, we arrange the likelihood function to exhibit the structure of a Poisson distribution likelihood.

The following Theorem 2 is a restated version of Theorem 1, where we use standard Poisson process notation, and state mathematically the two conditions we introduce in Theorem 1.

Theorem 2. *Consider a Poisson process $\{X(t), T \geq t \geq 0\}$ with rate λ (where $X(t)$ denotes the number of jumps up to time t), and denote the jumping times $\{T_1, T_2, \dots\}$. Denote the random variables of the start and end of an observation period S and E, and suppose S and E satisfy the following conditions, respectively:*
Condition 1: *Given $T_1 = t_1$, S is defined on $[t_1, T]$, and*

$$\mathbb{P}(S = t | T_1 = t_1) = g(t), \forall T \geq t \geq t_1,$$

where $g(t)$ is not dependent on λ or the sample path of the Poisson process after t_1.
Condition 2: *Given the jumping times of the Poisson process up to any time t, that is, for some $m \in \mathbb{Z}+$, given $\{T_1 = t_1, \dots, T_m = t_m, T_{m+1} > t\}$, the distribution of E is*

$$\mathbb{P}[E = t | \{T_1 = t_1, \dots, T_m = t_m, T_{m+1} > t\}] = h_m(t), \text{ for some } m \in \mathbb{Z}+, \forall T \geq t \geq t_m,$$

where $h_m(t)$ is a function independent of λ. Furthermore, given $T_1 = t_1$, the distribution of E is independent of the realization of S:

$$\mathbb{P}[E = t | T_1 = t_1, S = s] = \mathbb{P}[E = t | T_1 = t_1], \forall T \geq t \geq t_1, \forall s \geq t_1.$$

Now, suppose we observe data which contains for each incident, $S = s, E = e, X(e) - X(s) = M$, where $X(s) = m \geq 1$, and all the jumping times within interval $(s, e]$: $\mathcal{J}_M = \{T_{m+1} = t_{m+1}, \dots, T_{m+M} = t_{m+M}\}$, where $t_{m+1} > s$ and $T_{m+M} \leq e$. Then, conditional on the observed jumping times before s: $\mathcal{H} := \{T_1 = t_1, \dots, T_m = t_m \leq s\}$, the likelihood of the data can be decomposed as follows:

$$\mathbb{P}[X(e) - X(s) = M, E = e, S = s, \mathcal{J}_M | \mathcal{H}] = f(e, s, \mathcal{H}, \mathcal{J}_M) p(M | \lambda(e - s)), \tag{8}$$

where $p(M|\lambda(e-s))$ denotes the likelihood of a Poisson distribution with M occurrences and rate parameter $\lambda(e-s)$, and $f(e, s, \mathcal{H}, \mathcal{J}_M)$ is a function that does not depend on λ.

Perhaps surprisingly, we could not find rate estimation literature in which $E_i - S_i$ also depends on the sample path. Note that we could leverage classical Poisson process results (see e.g., [33]) if we had considered a non-data-dependent end time E_i, e.g., $E_i - S_i$ is a constant. However, this choice would be data inefficient. Some incidents are inspected and addressed within a day of being first reported, while others may take a year to be inspected. As incident inspection (and death time $t_i + T_i$) may be a function of the number of reports, we would be forced to pick a fixed interval end time of less than a single day, for all incidents.

The proof of the theorem relies on the following lemma, which establishes the distribution of the (random) time between S and the first jump time after S.

Lemma 2. *If the distribution of S satisfies **condition 1**, then the distribution of the time between any realization of S and the next jump time $T_{X(S)+1}$ is an exponential random variable with rate λ.*

Proof. Following the definition of jump times, we have for any $\epsilon > 0$,

$$\mathbb{P}[T_{X(S)+1} - S > \epsilon | T_1 = t_1] \tag{9}$$

$$= \sum_{m=1}^{\infty} \mathbb{P}[X(S+\epsilon) = m, X(S) = m | T_1 = t_1] \tag{10}$$

$$= \sum_{m=1}^{\infty} \int_{s \geq t_1} \mathbb{P}[X(s+\epsilon) = m, X(s) = m | S = s, T_1 = t_1] \mathbb{P}[S = s | T_1 = t_1] ds \tag{11}$$

$$= \sum_{m=1}^{\infty} \int_{s \geq t_1} \mathbb{P}[X(s+\epsilon) = m | X(s) = m, S = s, T_1 = t_1] \mathbb{P}[X(s) = m | S = s, T_1 = t_1] g(s) ds \tag{12}$$

$$= \sum_{m=1}^{\infty} \int_{s \geq t_1} \mathbb{P}[X(\epsilon) = 0] \mathbb{P}[X(s - t_1) = m - 1] g(s) ds \tag{13}$$

$$= \mathbb{P}[X(\epsilon) = 0] \int_{s \geq t_1} g(s) \left[\sum_{m=1}^{\infty} \mathbb{P}[X(s - t_1) = m - 1] \right] ds \tag{14}$$

$$= \exp(-\lambda\epsilon) \int_{s \geq t_1} g(s) ds \tag{15}$$

$$= \exp(-\lambda\epsilon) \tag{16}$$

where from Equation (12) to Equation (13) we use the memoryless property of Poisson processes jump times, and denote $\{X(t), t \geq 0\}$ as a new Poisson process with rate λ, starting from $t = 0$. From Equation (13) to Equation (14) we use $\mathbb{P}[X(\epsilon) = 0] = \exp(-\lambda\epsilon)$ is independent of s, $g(s)$ is independent of m (as a result of it being independent of the sample path after t_1), and $\mathbb{P}[X(s - t_1) = m - 1] \geq 0, \forall m \geq 1, s \geq t_1$, thus we can exchange the summation and integral, and put $g(s)$ outside of the summation. From Equation (14) to Equation (15) we use the fact that the summation is of the probability mass function of a Poisson random variable with rate $\lambda(s - t_1)$ over its range, thus evaluates to 1, and finally the last equation follows from the definition of $g(s)$ as a probability density function. We note that the final equation corresponds to the tail probability of an exponential random variable with rate parameter λ, and thus conclude our claim. □

Proof of Theorem 2. We now prove Theorem 2. The main idea of the proof is to decompose the likelihood into full conditional probabilities, and then express each part using the conditions we laid out and Lemma 2. The difficulty comes from separating terms involving λ from terms independent of λ, which requires the two conditions, and arranging the likelihood function to exhibit the structure of a Poisson distribution likelihood, where we rely on Lemma 2 above.

We prove the result in three cases, based on how many reports are observed within $(s, e]$: 0, 1, and 2 or more, as these scenarios have varying level of complexity.

As reference, the likelihood of a Poisson distribution with $M \geq 0$ occurrences and rate $\lambda(e-s) > 0$ is as follows:

$$p(M | \lambda(e - s)) = \frac{[\lambda(e - s)]^M}{M!} \exp(-\lambda(e - s)).$$

First, for $M = 0$, since we do not observe any jump times within $(s, e]$ in this case, $\mathcal{J}_0$ is an

empty set and we omit it from the equation.

$$\begin{aligned}
&\mathbb{P}[S=s, E=e, X(e)=X(s)=m|\mathcal{H}] && (17)\\
=&\mathbb{P}[S=s|\mathcal{H}]\mathbb{P}[E=e|S=s,\mathcal{H}]\mathbb{P}[T_{m+1}-s>e-s|E=e,S=s,\mathcal{H}] && (18)\\
=&g(s)h_m(e)\exp(-\lambda(e-s)) && (19)\\
=&f(e,s,\mathcal{H},\mathcal{J}_M)p(0|\lambda(e-s)) && (20)
\end{aligned}$$

From Equation (17) to Equation (18) we decompose the likelihood by the full conditional probabilities. The likelihood of the data contains three parts: (i) conditional on the history of reports $\mathcal{H}$, the likelihood of observing a realization of $S=s$; (ii) conditional on the history of reports $\mathcal{H}$ and $S=s$, the likelihood of observing a realization of $E=e$; (iii) conditional on the history of reports $\mathcal{H}$, $S=s$ and $E=e$, the likelihood of *not* observing the next jump time, that is $T_{m+1}>e$.

From Equation (18) to Equation (19) we use **condition 1** to express (i) as $g(s)$, **condition 2** to express (ii) as $h_m(e)$, and Lemma 2 to express (iii) as $\exp(-\lambda(e-s))$. The final equation follows by defining

$$f(e,s,\mathcal{H},\mathcal{J}_M)=g(s)h_m(e),$$

which is independent of λ.

For the case with $M=1$, we cannot omit J_1, but similarly have:

$$\begin{aligned}
&\mathbb{P}[S=s, E=e, X(e)=m+1, X(s)=m, \mathcal{J}_1|\mathcal{H}] && (21)\\
=&\mathbb{P}[S=s, E=e>t_{m+1}, T_{m+1}=t_{m+1}, T_{m+2}>e|\mathcal{H}] && (22)\\
=&\mathbb{P}[S=s|\mathcal{H}]\times\mathbb{P}[E>t_{m+1}|S=s,\mathcal{H}]\times\mathbb{P}[T_{m+1}-s=t_{m+1}-s|S=s,E>t_{m+1},\mathcal{H}]\\
&\times\mathbb{P}[E=e|S=s,\mathcal{H},\{T_{m+1}=t_{m+1}\}]\times\mathbb{P}[T_{m+2}>e|E=e,S=s,\mathcal{H},\{T_{m+1}=t_{m+1}\}] && (23)\\
=&g(s)\times\left[\int_{t_{m+1}}^{T}h_m(t)dt\right]\times\lambda\exp(-\lambda(t_{m+1}-s))\times h_{m+1}(e)\times\exp(-\lambda(e-t_{m+1})) && (24)\\
=&\lambda\exp(-\lambda(e-s))g(s)h_{m+1}(e)\left[\int_{t_{m+1}}^{T}h_m(t)dt\right] && (25)\\
=&f(e,s,\mathcal{H},\mathcal{J}_M)p(1|\lambda(e-s)) && (26)
\end{aligned}$$

From Equation (21) to Equation (22) we express the data in a different but equivalent manner. Note that the fact $\mathcal{J}_1$ contains the first jump time t_{m+1}. It (along with $X(e)=m+1$ equivalently can be stated as two events: first, trivially, we know $T_{m+1}=t_{m+1}$; second, we also know that T_{m+2} arrives later than e, which is why that jump is not observed. From Equation (22) to Equation (23) we similarly decompose the likelihood into full conditional probabilities. The likelihood now contains five parts: (i) conditional on the history of reports $\mathcal{H}$, the likelihood of observing $S=s$; (ii) conditional on the history of reports $\mathcal{H}$ and $S=s$, the likelihood of observing E to be greater than t_{m+1}; (iii) conditional on the history of reports $\mathcal{H}$, $S=s$ and E still not realized, the likelihood of observing $T_{m+1}=t_{m+1}$; (iv) conditional on the history of reports $\mathcal{H}$, $S=s$, $T_{m+1}=t_{m+1}$, the likelihood of observing a realization $E=e$; (v) conditional on the the history of reports $\mathcal{H}$, $S=s$, $E=e$, and $T_{m+1}=t_{m+1}$, the likelihood of *not* observing the next jump time, that is $T_{m+2}>e$.

From Equation (23) to Equation (24) we use **condition 1** to express (i), **condition 2** to express (ii), Lemma 2 to express (iii), **condition 2** again to express (iv), and then the memoryless property of Poisson process jump times to express (v). Equation (25) follows by collecting terms. The final

equation follows by defining

$$f(e, s, \mathcal{H}, \mathcal{J}_M) = \frac{1}{(e-s)} g(s) h_{m+1}(e) \int_{t_{m+1}}^{T} h_m(t) dt,$$

which is independent of λ.

Next, for any $M > 1$, we similarly have:

$$\begin{aligned}
&\mathbb{P}[S = s, E = e, X(e) = m + M, X(s) = m, \mathcal{J}_M | \mathcal{H}] && (27)\\
=&\mathbb{P}[S = s, E = e > t_{m+M} > \cdots > t_{m+1}, T_{m+1} = t_{m+1}, \ldots, T_{m+M} = t_{m+M}, T_{m+M+1} > e | \mathcal{H}] && (28)\\
=&\mathbb{P}[S = s | \mathcal{H}] \\
&\times \prod_{i=1}^{M} \mathbb{P}[E > t_{m+i} | S = s, \mathcal{H}, \{T_{m+1} = t_{m+1}, \ldots, T_{m+i-1} = t_{m+i-1}\}] \\
&\times \prod_{i=1}^{M} \mathbb{P}[T_{m+i} = t_{m+i} | S = s, \mathcal{H}, \{T_{m+1} = t_{m+1}, \ldots, T_{m+i-1} = t_{m+i-1}\}] \\
&\times \mathbb{P}[E = e | S = s, \mathcal{H}, \{T_{m+1} = t_{m+1}, \ldots, T_{m+M} = t_{m+M}\}] \\
&\times \mathbb{P}[T_{m+M+1} > e | E = e, S = s, \mathcal{H}, \{T_{m+1} = t_{m+1}, \ldots, T_{m+M} = t_{m+M}\}] && (29)\\
=&g(s) \\
&\times \prod_{i=1}^{M} \lambda \exp(-\lambda(t_{m+i} - t_{m+i-1})) \\
&\times \prod_{i=1}^{M} \int_{t_{m+i}}^{T} h_{m+i-1}(t) dt \\
&\times h_{m+M}(e) \times \exp(-\lambda(e - t_{m+M})) && (30)\\
=&\lambda^M \exp(-\lambda(e-s)) \times g(s) h_{m+M}(e) \prod_{i=1}^{M} \int_{t_{m+i}}^{T} h_{m+i-1}(t) dt && (31)\\
=&f(e, s, \mathcal{H}, \mathcal{J}_M) p(M | \lambda(e-s)). && (32)
\end{aligned}$$

From Equation (27) to Equation (28) we similarly express the data in an equivalent manner, also noting that the data provides information that $T_{m+M+1} > e$. From Equation (28) to Equation (29) we decompose the likelihood by the full conditional probabilities, which similar to the previous case contains five parts: (i) conditional on the history of reports $\mathcal{H}$, the likelihood of observing $S = s$; (ii) for $i = 1, \ldots, M$, conditional on the history of reports $\mathcal{H}$, $S = s$ and all jump times T_{m+1} to T_{m+i-1},[1] the likelihood of observing E to be greater than t_{m+i}; (iii) for $i = 1, \ldots, M$, conditional on the history of reports $\mathcal{H}$, $S = s$, all jump times T_{m+1} to T_{m+i-1} and E still not realized, the likelihood of observing $T_{m+i} = t_{m+i}$; (iv) conditional on the history of reports $\mathcal{H}$, $S = s$, all jump times $T_{m+1}, \ldots, T_{m+M}$, the likelihood of observing a realization $E = e$; (v) conditional on the the history of reports $\mathcal{H}$, $S = s$, $E = e$, and $T_{m+1} \ldots T_{m+M}$, the likelihood of *not* observing the next jump time, that is $T_{m+M+1} > e$.

From Equation (29) to Equation (30) we use **condition 1** to express (i), **condition 2** to express (ii), Lemma 2 to express (iii), **condition 2** again to express (iv), and then the memoryless property of Poisson process jump times to express (v). Equation (31) follows by collecting terms, and the

[1] Where $i = 1$, this degenerates to be an empty set, the same for part (iii) below.

last step follows by defining

$$f(e, s, \mathcal{H}, \mathcal{J}_M) = \frac{(e-s)^M}{M!} g(s) h_{m+M}(e) \prod_{i=1}^{M} \int_{t_{m+i}}^{T} h_{m+i-1}(t) dt,$$

which is independent of λ.

Combining Equations 20, 26 and 32 concludes our claim.

□

3 Comparing estimation methods via simulation

Before applying our methods to real-world 311 data, we demonstrate the effectiveness of our methods via simulated data—in such simulations, we have full control over the data-generating process and thus can compare the estimation results with the true parameters. In particular, we use the simulator to illustrate (a) Proposition 1, that attempting to recover reporting rates λ_θ from $N_\theta^{\text{observed}}(T)$ is prone to bias; (b), that in Theorem 1, the conditions on S and E are essential, (c), that in a correctly-specified model setting, i.e., following the assumptions in Theorem 1, both the MLE in Equation (3) and the homogeneous Poisson regression recover the ground truth; and (d), that the regression approach is more data-efficient than the MLE, leading to tighter parameter estimates.

Simulator setup We simulate a basic time-homogeneous system, as follows. We set the incident type to be a two-dimensional vector $\theta \in \mathbb{R}^2$. Our simulator needs parameters for three processes: the incident birth process governed by homogeneous Poisson rate Λ_θ, the reporting process governed by homogeneous Poisson rate λ_θ, and a lifetime T_i of each incident, generated in the following way: after each report, we sample two competing exponential clocks, one representing incident death and another the next report; the incident death rate can depend on the number of reports so far, reflecting, for example, that the agency prioritizes inspections for incidents with more reports. If the report happens before death, we increment the number of reports for the incident and repeat; otherwise, the incident has "died" and no more reports are logged.

Formally, for each type of incident, we specify a parameter μ_θ that is independent of λ_θ as the "death rate" of incidents; conditional on there has been m reports of an incident of type θ, we generate two exponentially distributed random variables, $d_\theta^m \sim \text{Exponential}(\mu_\theta \times (\gamma_\theta)^m)$, where

Supplementary Table 1: Simulation results comparing five estimates of reporting rate λ_θ under different incident rates Λ_θ. The true reporting rates are all $\lambda_\theta = 2$. We find that with correctly specified stopping times, the Poisson regression estimators are more precise than the MLE, especially as the incident rate decreases. The naive estimator, or the incorrectly specified observation ending time both introduce bias into the estimation. Estimate standard deviations are in parentheses.

	Incident rate Λ				
Estimates for reporting rate λ	1.0	2.0	3.0	4.0	5.0
Naive	1.201(0.016)	2.394(0.032)	3.589(0.051)	4.788(0.066)	5.999(0.084)
MLE, correct	2.041(0.188)	2.018(0.091)	2.005(0.057)	2.005(0.045)	2.013(0.035)
MLE, incorrect	8.992(4.276)	8.686(1.696)	8.649(1.221)	8.615(0.859)	8.596(0.635)
Poisson regression, correct	2.033(0.118)	2.014(0.058)	2.003(0.036)	2.004(0.029)	2.011(0.022)
Poisson regression, incorrect	8.826(2.356)	8.613(1.008)	8.599(0.755)	8.581(0.539)	8.572(0.402)

γ_θ is a scaling parameter by our choice, and $r_\theta^m \sim \text{Exponential}(\lambda_\theta)$; if $d_\theta^m \leq r_\theta^m$, we consider the incident dead, and let $T_i = \sum_{j=0}^{m-1} r_\theta^j + d_\theta^m$, otherwise increment m and repeat the process. Both the incident and reporting rates are set as a function of the type covariates, as in Equation (5):

$$\Lambda_\theta = \exp\left(\alpha_{\text{incident}} + \beta_{\text{incident}}^T \theta\right) \qquad \lambda_\theta = \exp\left(\alpha_{\text{report}} + \beta_{\text{report}}^T \theta\right)$$

where $\alpha_{incident} \in \mathbb{R}$, $\beta_{incident} \in \mathbb{R}^2$, $\alpha_{report} \in \mathbb{R}$ and $\beta_{report} \in \mathbb{R}^2$ are varied across simulations. For simplicity in comparing the various methods, here we report simulation results for the case in which there are only 5 distinct types, and the reporting rate for each type is 2, i.e., $\alpha_{report} = \log(2)$ and $\beta_{report} = \mathbf{0}$, but where incident rates Λ_θ vary by type.

Next, we need to set μ_θ and γ_θ, which governs the duration that the incident is alive. Given the above parameters, we do so in a manner that matches the distribution of the number of reports received per incident to the real data studied in the next section: that 18.7% of the received reports in the reporting period are duplicates of incidents already reported. For simplicity, we set these uniformly across all types. This calibration results in $\mu_\theta = 0.065$ and $\gamma_\theta = 100$ for all θ.

Finally, we fix the time span of our observation to be $T = 300$ days, which by our parameter settings, is long enough for the system to converge to its long-run stationary distribution (in terms of the number of active incidents). Incidents and reports are generated according to the simulator setup. It is possible that some death times and consequently reports generated may be beyond the 300-day observation period. We discard any such reports.

The output of the simulator consists of all the incidents that occur and are reported at least once during this 300-day period. For all these incidents, the available data for our estimators is the time that each incident was reported the first time, the times and number of the subsequent reports, and the times of the incident death if they occur before the end of the time span. To specify the observation period of each incident, in the setting of Theorem 1, we let S_i be the time of the first report of such incident, and E_i be the incident death time, or the end of the time span if the death time is greater than it. This specification satisfies the assumption in Theorem 1. As a comparison, we add an incorrectly specified version of this observation period by letting E_i be the time of the last report of this incident, in which case it no longer satisfies the conditions.

Simulator results We compare five estimators: a "naive" estimator, that calculates $\frac{N_\theta^{\text{observed}}(T)}{T}$, the ratio between the observed number of reports and the observation period; the MLE as derived in Equation (3) with correctly specified stopping times; the MLE with incorrectly specified observation ending times; a Poisson regression with correctly specified stopping times, and finally, a Poisson regression with incorrectly specified observation ending times. The regression methods were implemented using Scikit-learn [31]. We run each of the methods on the same simulated datasets and iterated 300 times. Supplementary Table 1 summarizes our results: showing, for each distinct type with a differing incident rate, the estimates for λ_θ for each of the methods.

These results indeed illustrate Proposition 1, that attempting to recover reporting rates λ_θ from $N_\theta^{\text{observed}}(T)$ is prone to bias: the naive method of counting the number of observed incidents conflates the incident rate with the reporting rate. Second, the stopping times assumption in Theorem 1 is indeed crucial to a valid result, and in a correctly specified model setting, both the MLE in Equation (3) and the homogeneous Poisson regression recover the ground truth reporting rate λ_θ, regardless of the incident rate. However, third, the regression approach is more data-efficient than the MLE, leading to tighter parameter estimates – especially as the incident rate decreases, i.e., as the sample size in terms of the number of incidents decreases. This data efficiency is important in high dimensions, as in our real-world data application in the next section.

4 Supplementary Empirical analysis

4.1 Two potential sources of bias

There are two potential sources of bias in the NYC data, due to ways in which the data differs from the model assumed in Section 5.2.1. Here, we evaluate these concerns and conclude that they likely do not substantially affect our measurements.

Repeat callers about the same incident. One potential worry is that duplicate reports are a mirage: they are primarily generated by the same resident repeatedly calling about an incident until it is addressed. If that is the case, our method does not work: we rely on a Poisson rate assumption for the reporting behavior (that reporting behavior is memory-less and so that one report does not affect the likelihood of another for that incident), which is likely violated if the same person makes multiple reports about the same incident. In theory, such repeats should be minimal: NYC makes available a portal to check the status of past reports, so a reporter does not need to call again to remain up-to-date. In practice, however, our contacts at NYC DPR indicated that repeat calls occur.

To mitigate the effect of such repeat callers on our analysis, we obtain anonymized (hashed) caller information from NYC DPR for each report: if the caller provided it, their name, phone number, and/or email address. We then filter out the duplicate reports for each incident where either the phone number or emails match, or both the first and last names match. Our analyses are run on the resulting filtered dataset. The above approach may not filter out all repeat callers: if callers choose not to leave their information, but call multiple times. Thus, we also run our main analyses on a filtered dataset where we additionally assume that any caller who did not leave their information is a repeat of a previous caller with no information. Our estimates are largely the same on this more conservative dataset, suggesting that repeat callers do not substantially bias our estimates. Results are given in Supplementary Information 4.4.

Censored data: incidents that were not inspected The NYC DPR only marks duplicate reports corresponding to incidents that were inspected: for the service requests not connected to an inspected incident, we do not know which (if any) other reports also refer to the same incident. As our method relies on the rate of duplicate reports, we must discard service requests that were not inspected.

This censoring may limit the generalizability of our findings, from measuring the reporting rates of all incidents to measuring the reporting of incidents that tend to be inspected. This limitation to external validity may be acceptable: if the inspection decisions are correlated with incident importance (likely), then studying the heterogeneous reporting behavior for these incidents is a more important task than is studying that of minor incidents not deemed worth inspecting.

There is a second reason we believe that the censoring is relatively acceptable. In particular, we measure reporting rates as a function of incident type θ, where the type includes characteristics such as report category and incident risk. If our models are correctly specified, and θ is rich enough to capture inspection decisions (there is no confounding), then this censoring does not affect our estimates.[2] While a seemingly strong assumption, we note that the θ we have available is the

[2]We aim to model $\tilde{Pr}(Y|\theta)$, where Y is reporting behavior. However, with just data on inspected incidents, we can only estimate $\tilde{Pr}(Y|\theta, \text{inspected})$. If Y is independent of the inspection decision given θ, then $\tilde{Pr}(Y|\theta, \text{inspected}) = \tilde{Pr}(Y|\theta)$. One potential source of bias is if, even conditional on θ (which includes the *content* of the reports), NYC DPR is making decisions that strongly correlate with the *number* of reports. Then, all of our rate estimates would be biased upwards, as we selectively observe data for incidents with many reports. However, as we primarily care about

same data that the NYC DPR sees about a report through their portal when making an inspection decision; any confounding would have to come from another source.

Nevertheless, to the extent that the above (likely small) bias affects practice, it may be valuable for 311 systems to systematically tag duplicates for all reports and then apply our methods. (Relative to the missing data challenges that we centrally tackle in this work, i.e., incidents not reported and birth and death times, this duplicate censoring is cheaply addressable by city agencies). We note that the Chicago data *does* mark duplicates even for open/uninspected incidents, and so this bias does not appear there.

4.2 NYC data preprocessing

Before training models, we need to construct a dataset in which each row corresponds to an incident, and where we have the number of reports $\tilde{M}_i$ in an observation interval, the duration $\tilde{T}_i$ of that interval, and covariates θ. We separated out an exploratory dataset of 8,000 unique incidents, on which we conducted covariate selection as detailed below.[3]

We filter out the reports corresponding to the uninspected service requests (as we do not have duplicate information for these) and then use the provided incident label to group all service requests for the same incident. Then, we remove repeat caller reports, comparing each caller to previous (ordered by time) callers for the same incident. Next, we must construct a valid observation interval for each incident.

Constructing an observation interval $(S_i, E_i]$ As outlined in Theorem 1, we must be careful in how we choose an observation interval $(S_i, E_i]$ in which we count reports – we need that the interval is inside the incident lifetime, i.e., we must end the interval before the incident is addressed, $E_i \leq t_i + T_i$. Both endpoints of the interval must also satisfy the conditions outlined in Theorem 1. As discussed above, the best choice to start the observation period S_i is the time of the first report, but choosing the observation end is a design choice.

We make the following choice. Let t_i^{INSP} be the inspection time of incident i, and, t_i^{WO} be the time that a work order is placed for incident i, if applicable. Then, E_i of each incident i is:

$$E_i = \min\left\{100 \text{ days} + S_i,\ t_i^{\text{INSP}},\ t_i^{\text{WO}}\right\}. \tag{33}$$

The maximum duration of 100 is a design choice, for which we perform robustness checks (with 30 and 200 days); a maximum mitigates – for incidents not inspected for a long time – the inclusion of a time period in which an issue might have been resolved before an inspection, which would bias our estimates downward. Equation (33) requires that inspection and work order times are stopping times; that they do not depend on the future trivially holds, and it is likely that they do not depend on the reporting rate except through the type θ and the sample path number of reports received

heterogeneous reporting rates across types, such a bias matters to the extent that it heterogeneously affects different types of incidents or geographic locations. Furthermore, according to NYC DPR, the primary drivers of inspection decisions are the report characteristics, which are included in θ. Nevertheless, an important direction for future work is directly addressing this censoring challenge.

[3]While the exploratory data was used to filter variables for ultimate analysis, and to develop and fine-tune our models, we note that we did not hold out a separate test set at the outset of the project; it was not clear how to cluster assign reports to test and train before developing our empirical strategy, and there were initial (ultimately resolved) data errors on how reports were tagged to unique incidents. Thus, our overall approach was selected and developed using the NYC DPR data on which we ultimately report results, but not using the Chicago dataset. The exploratory dataset was ultimately composed of 4463 unique incidents, after filtering.

up to that time.[4] (It is not a problem that incidents with more reports are inspected sooner).

Supplementary Figure 2a shows the histogram of the number of reports per incident during the observation interval; Supplementary Figure 2b shows the distribution of durations; and Supplementary Table 2 shows how the average duration differs by Borough and report category. The heterogeneity in duration length (due to the speed of being inspected or worked on) demonstrates the value of Theorem 1, which allows us to maximally utilize the data without introducing bias. For example, compared to 'Prune' incidents, 'Hazard' incidents tend to have more reports on average and shorter duration: residents have a higher reporting rate for hazardous incidents, and these incidents tend to be addressed more quickly. Suppose we had to use a fixed duration D, instead of an inspection/work order dependent time. If D is large (e.g., $D \approx 15$ days), then we bias our estimates downward, as we're including time periods after an incident has already been addressed – and the bias heterogeneously affects incident types, since some incident types are typically addressed more quickly than are other types. On the other hand, a much shorter duration would substantially limit the data. Finally, Supplementary Figure 4 shows the average number of days after the first report that the ℓth duplicate report was submitted for an incident, for incidents with at least $k \geq \ell$ reports. The plots are largely linear (i.e., the average delay between the first and second report is the same as that between the third and fourth report), consistent with reporting rates being approximately homogeneous Poisson within the interval.

Covariate selection and processing Next, we select the covariates that compose type θ. The data given to us by NYC DPR includes a set of *report* covariates (e.g., report *Category*),[5] *inspection* results (e.g., condition of the tree at inspection time), and *tree* characteristics (e.g., the diameter of the tree at breast height, tree species). We augment this data with *socioeconomic characteristics* as follows. Most of the reports in our data contain latitude-longitude coordinates for each inspection (and thus incident), using which we identify which of the over 2000 census tracts in New York City the incident is in, through an FCC API.[6] We then join this information with 2020 Census data obtained from the IPUMS NHGIS [23], which include socioeconomic characteristics such as race/ethnicity, education, income, and population density for each census tract.

Next, we perform covariate selection using the exploratory dataset. We remove report and inspection variables that are highly collinear, have low variance, or with a high number of missing values. Conversations with NYC DPR also played a role in the selection. Finally, we log transform several variables and standardize all data. Supplementary Table 3 contains the covariates we use.

Starting with the dataset discussed above, we filter out the incidents for which any of the covariates are missing and those with short-logged reporting periods (Duration less than 0.1 days). We are left with a dataset of 81,638 incidents on which we conduct our main analyses; for each incident i, we have the duration of observance $e_i - s_i$, the number of total reports $\tilde{M}_i$, and all the various geographic and demographic covariates associated with it.

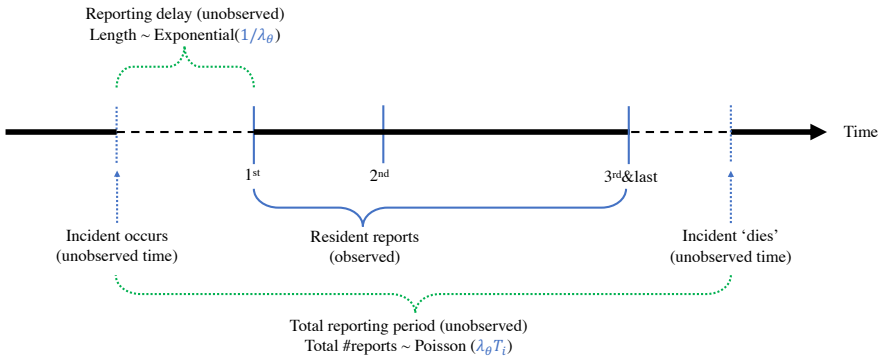
[4] We observe largely the same dashboard data that the inspector does when making inspection decisions, and so there is minimal unobserved confounding; the exception is that we do not process report free-form text, though in our conversations with NYC DPR these hold secondary importance after the structured fields.

[5] Occaisonally, different reports about the same incident disagree on the report covariates. We select the first report characteristics in those cases.

[6] FCC Area API, https://geo.fcc.gov/api/census/

4.3 Additional information

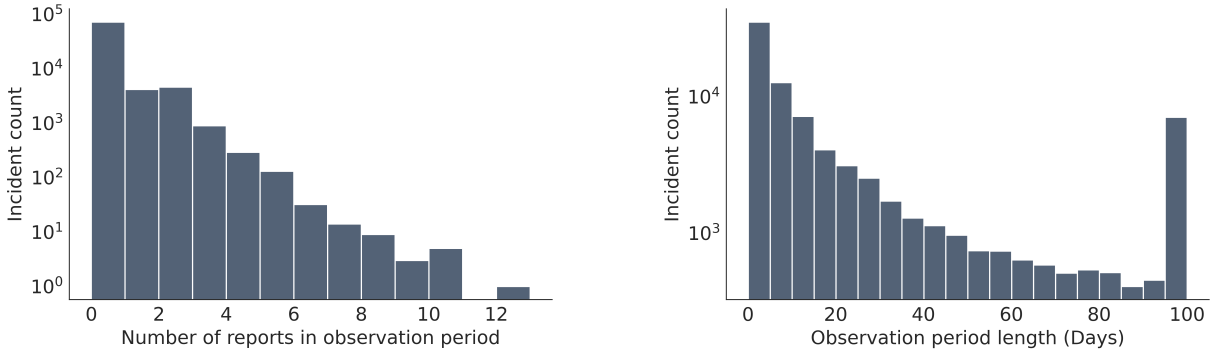
In this section, we provide some additional information about our model, dataset and results. Supplementary Figure 1 illustrate our data-generating model. Supplementary Table 2 contains summary statistics of the NYC dataset on which we conduct our main analyses in the main text. Supplementary Figure 2a shows the histogram of the number of reports per incident during the observation interval; Supplementary Figure 2b shows the distribution of durations. Supplementary Table 3 provides a description of the covariates selected; Supplementary Figure 3b shows the relationship between the number of unique incidents observed versus the census tract fixed effect; Supplementary Figure 5 shows the posterior distribution of the number of reports as estimated by the basic Poisson regression model, and the zero-inflated Poisson regression model, with reference to the observed distribution. Note that the Zero-inflated model has a better fit to the observed distribution since most of the density of this distribution is on reports 1, 2, and 3. We note that our framework also allows for zero inflation coefficients to be Category-dependent. Using Category-dependent zero inflation coefficients further improves the fit, but the gain is not substantial. Supplementary Table 4 and Supplementary Table 5 lists the full information of the coefficients for census tract socioeconomic covariates as estimated alone in a regression alongside the incident-specific covariates and the borough fixed effects. Supplementary Table 7 lists the information of the coefficients for a subset of census tract socioeconomic covariates as estimated together in a regression alongside the incident-specific covariates.



Supplementary Figure 1: Model of resident reports of an incident i of type θ. Observed and unobserved events are marked with solid and dotted lines, respectively. In this example, there are a total of 3 reports made about this incident. When the reports are generated according to a homogeneous Poisson process with rate λ_θ, the average reporting delay is $\frac{1}{\lambda_\theta}$. Incident 'death' refers to the incident being resolved or otherwise marked such that no future reports are submitted and logged. The goal is to estimate the reporting rate λ_θ (and consequently, the unobserved reporting delay) for each incident of type θ.

Supplementary Table 2: Summary statistics: service requests and inspections statistics are directly from the NYC DPR reports data; unique incidents statistics are derived from the inspected service requests, where the average duration is calculated using the stopping times definition. The *Others* category includes four other categories: Rescue/Preservation, Remove Stump, Pest/Disease, Planting Space, that together account for less than 0.4% of the service requests; we exclude them in the analysis. Categories are reported by the person raising the service request: for example, Prune reflects a request to prune the leaves of an over-grown tree, and Hazard reflects a request for NYC DPR to attend to a potentially hazardous condition concerning trees.

	Service requests	**Inspections**		**Incidents (from inspected SRs)**		
		Inspected SRs	Fraction inspected	Unique incidents	Avg. reports per incident	Median Days to Inspection
Total number	223,416	140,057	0.63	98,994	1.41	5.79
By Borough						
Queens	90,930	55,904	0.61	42,724	1.31	5.50
Brooklyn	67,852	45,666	0.67	27,368	1.67	8.68
Staten Island	27,263	15,601	0.57	11,755	1.33	4.07
Bronx	22,629	14,928	0.66	10,059	1.48	6.76
Manhattan	14,702	7,938	0.54	7,072	1.12	2.25
By Category						
Hazard	87,864	59,667	0.68	40,167	1.49	1.93
Prune	48,649	26,706	0.55	20,589	1.30	10.29
Remove Tree	44,177	29,307	0.66	22,275	1.32	7.73
Root/Sewer/Sidewalk	30,856	17,392	0.56	14,694	1.18	17.86
Illegal Tree Damage	11,061	6,533	0.59	5,525	1.18	22.87
Other	809	452	0.56	429	1.05	5.48



(a) Number of reports per incident.

(b) Length of observation period.

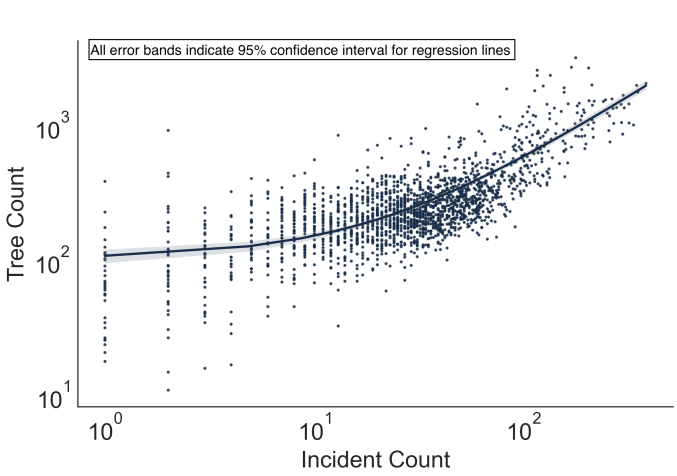
Supplementary Figure 2: Distribution of number of reports and length of observation for each unique incident in the aggregated dataset. For most incidents, there are no reports after the first report (at least not in the observation period). There is a peak at 100 days for the observation period, due to our configuration in Equation (33), where we truncate longer periods to 100 days.

Supplementary Table 3: Description of covariates in the aggregated dataset.

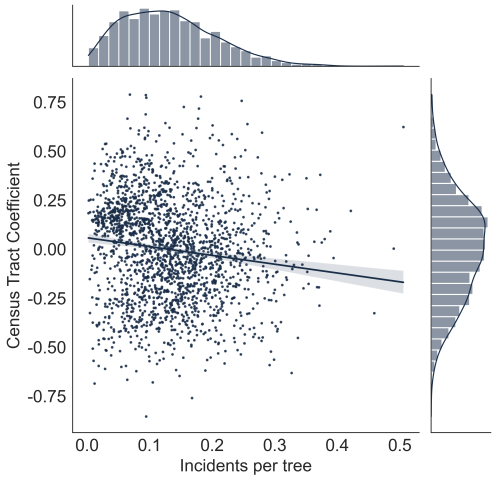
Covariate	Description
Incident Global ID	An identifier unique to each incident.
Duration	The observation duration as defined in Equation (33)
Number Reports	Number of reports on this incident in the observation duration, after filtering out repeat callers
INSPCondition	The inspection outcome regarding the condition of the tree, indicates whether the tree is dead, in good to excellent conditions, or in fair conditions.
INSP_RiskAssessment	The inspection outcome regarding how dangerous the reported incident is, as determined by inspector. Ranges from 0 to 12, with 11 and 12 being category "A" incident and prioritized for work orders.
Tree Diameter at Breast Height (TDBH)	Main characteristic of the tree describing how large the tree trunk is. Measured in inches.
Borough	Indicating which borough in NYC this incident is located.
Category	The incident category as reported.
Median Age	Median age in the census tract.
Fraction Hispanic	Fraction of residents that identify as Hispanic in the census tract.
Fraction white	Fraction of residents that identify as white in the census tract.
Fraction Black	Fraction of residents that identify as Black in the census tract.
Fraction no high school degree	Fraction of residents that have not graduated from high school in the census tract.
Fraction college degree	Fraction of residents that have graduated from college in the census tract.
Fraction poverty	Fraction of residents that are identified to be in poverty in the census tract.
Fraction renter	Fraction of residents that rent their current residence in the census tract.
Fraction family	Fraction of family household in the census tract.
Median household value	Median value of household in the census tract.
Income per capita	Income per capita of residents in the census tract.
Density	Population density in the census tract.

Supplementary Table 4: Census Tract socioeconomic coefficients in NYC, estimated alone in a regression alongside the incident-specific covariates. Full table corresponding to Supplementary Table 3.

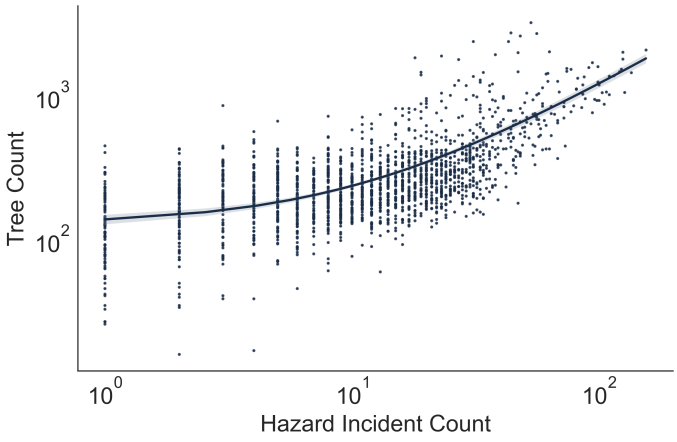
	Mean	StdDev	2.5%	97.5%
Median age	-0.033	0.008	-0.051	-0.018
Fraction Hispanic	0.030	0.009	0.011	0.046
Fraction white	0.057	0.008	0.040	0.072
Fraction Black	-0.039	0.009	-0.059	-0.024
Fraction no high school degree	-0.031	0.008	-0.049	-0.016
Fraction college degree	0.043	0.009	0.024	0.059
Fraction poverty	-0.010	0.009	-0.027	0.007
Fraction renter	0.054	0.009	0.035	0.070
Fraction family	-0.081	0.009	-0.099	-0.065
Log(Median house value)	0.065	0.009	0.045	0.081
Log(Income per capita)	0.048	0.009	0.031	0.068
Log(Density)	0.077	0.010	0.056	0.099



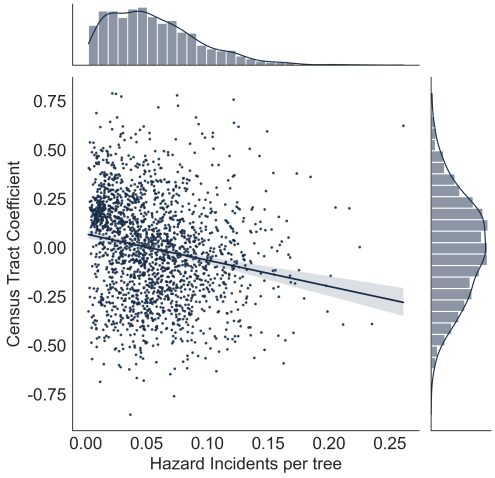
(a) Reported incidents vs number of trees



(b) Incidents/Tree vs Census Tract Coefficient

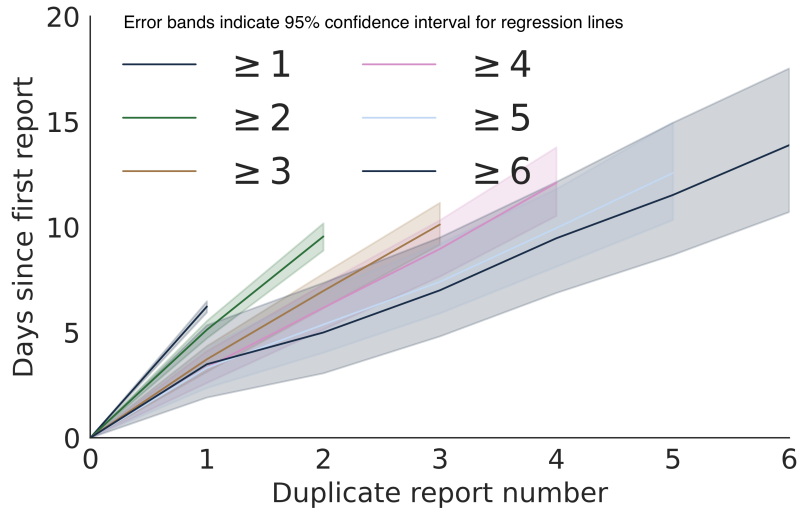


(c) Reported **Hazard** incidents vs number of trees



(d) **Hazard** Incidents/Tree vs Census Tract Coefficient

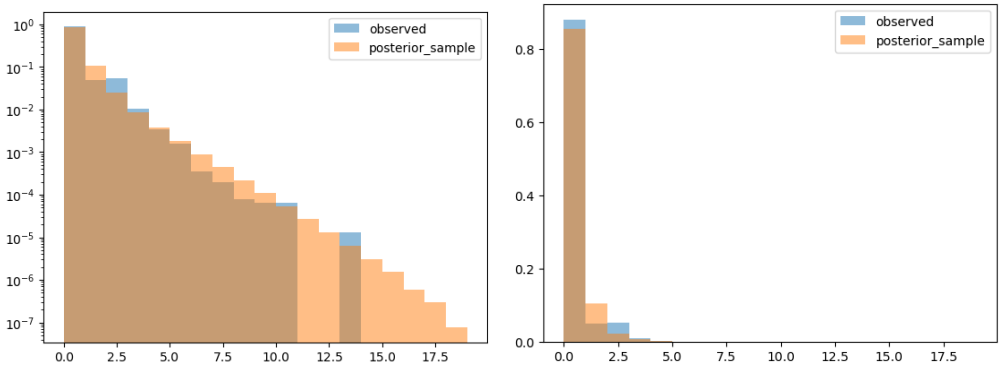
Supplementary Figure 3: (a) For each census tract, the number of trees according to the 2015 New York Street Tree Census versus the number of reported incidents in that census tract. As expected, more trees result in more reported incidents. (b) The relationship between reported incidents per tree and the census tract fixed effect. (c) and (d) reproduce these plots while restricting to only **Hazard** incidents. The value, "incidents per tree" is an attempt to normalize the number of incidents we observe with the number we expect to observe – and thus the ratio is a measure of *reporting* rates as is done in prior work. We observe a slightly negative relationship between this measure and the one we develop, the census tract coefficient in the Poisson regression. While it is possible that one can construct better proxies for how many incidents we expect to observe than the raw counts of trees, the relationship suggests that our method's results can differ substantially from those of prior work. We prefer our measure, as it automatically controls for 'legitimate' incident-level characteristics (such as risk) that may correlate with geography but are not captured with the number of trees—without needing to construct an estimate for the number of expected incidents for each such type. In practice, there can be many types of incidents; see, e.g., Supplementary Table 23. The Street Tree data is available here: https://data.cityofnewyork.us/Environment/2015-Street-Tree-Census-Tree-Data/pi5s-9p35.



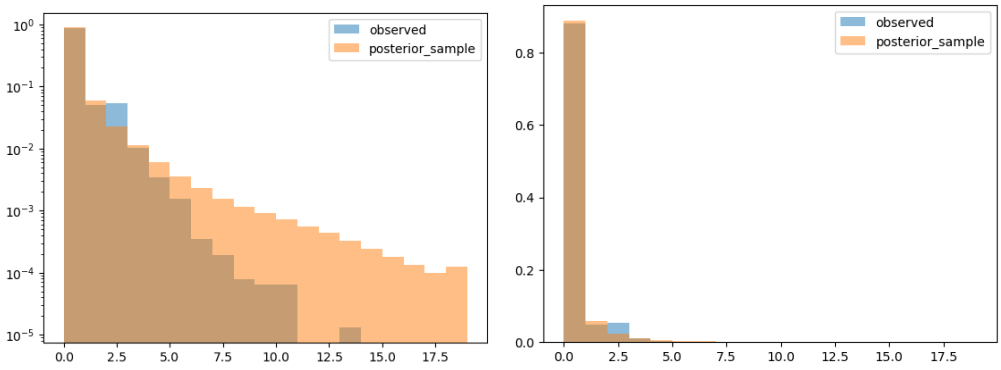
Supplementary Figure 4: Conditional on having at least k duplicate reports, on average how many days after the $0th$ report was the ℓth report, where $\ell \leq k$? The linear nature of each plot is consistent with a homogeneous Poisson process within the given period. That incidents with more reports also receive reports *faster* is consistent with those incidents being more severe in nature and having a higher reporting rate λ_θ.

Supplementary Table 5: Census Tract socioeconomic coefficients in NYC, estimated alone in a regression alongside the incident-specific covariates and the **borough fixed effects**.

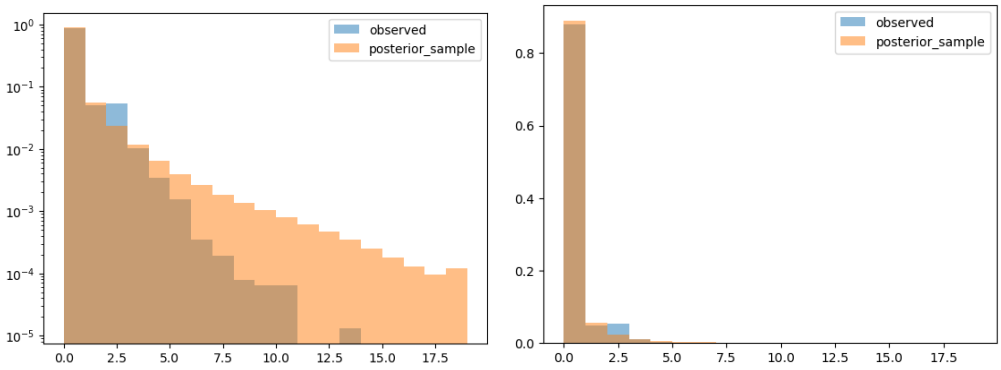
	Mean	StdDev	2.5%	97.5%
Median age	-0.009	0.008	-0.025	0.005
Fraction Hispanic	0.053	0.009	0.035	0.069
Fraction white	0.068	0.008	0.052	0.083
Fraction Black	-0.052	0.009	-0.070	-0.037
Fraction no high school degree	-0.039	0.008	-0.056	-0.024
Fraction college degree	0.035	0.008	0.019	0.049
Fraction poverty	-0.020	0.008	-0.036	-0.006
Fraction renter	0.022	0.008	0.004	0.038
Fraction family	-0.075	0.009	-0.092	-0.060
Log(Median house value)	0.023	0.008	0.006	0.038
Log(Income per capita)	0.056	0.008	0.038	0.072
Log(Density)	0.032	0.009	0.015	0.048



(a) Standard Poisson regression



(b) Zero inflated Poisson regression



(c) Zero inflated Poisson regression, with Category-dependent zero inflation coefficient

Supplementary Figure 5: Comparison among posterior distributions sampled from the Standard Poisson regression model and two Zero-inflated variants. Left-hand plots are in log space for the y axis, and right-hand plots are in regular probability space, to highlight the different parts of the distribution.

Supplementary Table 6: Census Tract socioeconomic coefficients in NYC, estimated together in a regression alongside the incident-specific covariates. Covariates with high levels of collinearity are dropped to maintain interpretability.

	Mean	StdDev	2.5%	97.5%
Median age	-0.014	0.009	-0.034	-0.003
Fraction white	0.058	0.009	0.037	0.076
Fraction college degree	-0.047	0.013	-0.073	-0.023
Fraction renter	0.042	0.010	0.021	0.060
Log(Income per capita)	0.086	0.014	0.057	0.110

Supplementary Table 7: Census Tract socioeconomic coefficients in NYC, estimated together in a regression alongside the incident-specific covariates. Compared with Supplementary Table 6, we additionally include for (log) population density. With the exception of the fraction of residents with college degrees which becomes insignificantly associated with the reporting rate, most socioeconomic variables are still significantly associated with the reporting rate in the same direction.

	Mean	StdDev	2.5%	97.5%
Median age	-0.021	0.009	-0.038	-0.003
Fraction white	0.048	0.010	0.028	0.066
Fraction college degree	-0.011	0.013	-0.037	0.013
Fraction renter	0.038	0.012	0.015	0.061
Log(Income per capita)	0.074	0.015	0.041	0.103
Log(Density)	0.087	0.012	0.063	0.108

4.4 Robustness checks – Supplementary Table 1 with other specifications

In this section, we present evidence for the robustness of our results in Supplementary Table 1 by reproducing its results with different configurations in Supplementary Table 8 through Supplementary Table 11. Additionally, we show results when the risk assessment scores are binned according to their priority level in Supplementary Table 12.

4.5 Calculation of mean reporting delay

In this section, we give the example calculation for the mean reporting delay of a Hazard, tree in Poor condition, risk assessment score 12 incident in Manhattan. The mean delay is calculated as:

$$1/\exp(\underbrace{-3.229}_{\text{Intercept, tree in Poor condition}} + \underbrace{1.418}_{\text{Hazard}} + \underbrace{0.438}_{\text{Manhattan}} + \underbrace{\frac{12-6.4915}{2.1788}\times 0.240}_{\text{Standardized risk assessment score}}) \approx 2.2$$

There are a few points worth mentioning with this calculation. First, we take the exponential of the sum of these coefficients, in accordance with the specification of the Poisson regression we fit in Equation (4); we further take the reciprocal, since the mean of an Exponential random variable is the reciprocal of its rate. Second, coefficient estimate for the dimension of 'tree in Poor condition' is integrated into the estimate for the intercept term; for tree in other conditions, an additional appropriate coefficient estimate needs to be added to the exponent. Third, tree size and risk assessment scores are standardized in the train set, thus in the calculation, we need to do the same standardization process as when we obtained the train set. Since we are concerned with trees of average size, the tree size variable is standardized to 0 here and omitted.

4.6 Additional information on socioeconomic and spatial reporting inequities

In this section, we provide some additional information on socioeconomic and spatial reporting inequities omitted from the Results section.

Calculation of cumulative association of socioeconomic variables. We first give an example of how the cumulative association scores in Figure 1 are calculated. Note that all coefficients given in Supplementary Table 6 are on standardized covariates. (Standardization was done on all covariates on the data frame used for the Poisson regression, and average refers to the corresponding average on that dataset). Let us suppose one hypothetical census tract has the following socioeconomic profile:

- Median age: 1 standard deviation above average;
- Fraction of white residents: 1 standard deviation below average;
- Fraction of residents with a college degree: 0.5 standard deviations above average;
- Fraction of renters: 0.5 standard deviations below average;
- Log income per capita: 0.5 standard deviations above average.

Then, according to Supplementary Table 6, the cumulative association can be calculated as:

$$1\times(-0.014)+(-1)\times 0.058+0.5\times(-0.047)+(-0.5)\times 0.042+0.5\times 0.086=-0.0735,$$

where each additive part in the equation corresponds to each of the socioeconomic variables above.

Supplementary Table 8: Regression coefficients for Standard Poisson regression with incident-level covariates and Borough fixed effects for Max Duration 100 days, Default repeat caller removal.

	Mean	StdDev	2.5%	97.5%	R_hat
Intercept	-4.577	0.023	-4.629	-4.539	1.0
INSPCondition[T.Dead]	-0.334	0.029	-0.396	-0.285	1.0
INSPCondition[T.Excellent_Good]	-0.428	0.023	-0.472	-0.382	1.0
INSPCondition[T.Fair]	-0.296	0.022	-0.344	-0.254	1.0
INSP_RiskAssessment	0.286	0.010	0.265	0.305	1.0
Log(Tree Diameter at Breast Height)	-0.020	0.009	-0.039	-0.003	1.0
Borough[Bronx]	0.103	0.022	0.061	0.144	1.0
Borough[Brooklyn]	-0.148	0.016	-0.185	-0.117	1.0
Borough[Manhattan]	-0.081	0.038	-0.163	-0.016	1.0
Borough[Queens]	-0.112	0.015	-0.142	-0.080	1.0
Borough[Staten Island]	0.237	0.026	0.183	0.284	1.0
Category[Hazard]	1.448	0.015	1.416	1.473	1.0
Category[Illegal Tree Damage]	0.009	0.028	-0.045	0.059	1.0
Category[Prune]	-0.083	0.024	-0.136	-0.043	1.0
Category[Remove Tree]	0.001	0.021	-0.042	0.043	1.0
Category[Root/Sewer/Sidewalk]	-1.375	0.029	-1.432	-1.325	1.0

Supplementary Table 9: Regression coefficients for Zero-inflated Poisson regression with incident-level covariates and Borough fixed effects for Max Duration 30 days, Default repeat caller removal.

	Mean	StdDev	2.5%	97.5%	R_hat
Intercept	-2.679	0.031	-2.738	-2.617	1.0
Zero Inflation fraction	0.728	0.003	0.722	0.735	1.0
INSPCondition[T.Dead]	-0.226	0.042	-0.312	-0.144	1.0
INSPCondition[T.Excellent_Good]	-0.368	0.029	-0.422	-0.311	1.0
INSPCondition[T.Fair]	-0.178	0.027	-0.230	-0.125	1.0
INSP_RiskAssessment	0.204	0.012	0.178	0.227	1.0
Log(Tree Diameter at Breast Height)	-0.009	0.009	-0.026	0.008	1.0
Borough[Bronx]	-0.041	0.027	-0.099	0.012	1.0
Borough[Brooklyn]	-0.188	0.019	-0.226	-0.155	1.0
Borough[Manhattan]	0.297	0.051	0.191	0.388	1.0
Borough[Queens]	-0.240	0.020	-0.279	-0.203	1.0
Borough[Staten Island]	0.172	0.035	0.097	0.238	1.0
Category[Hazard]	1.336	0.019	1.295	1.371	1.0
Category[Illegal Tree Damage]	0.257	0.039	0.173	0.323	1.0
Category[Prune]	-0.113	0.034	-0.186	-0.050	1.0
Category[Remove Tree]	-0.030	0.025	-0.084	0.016	1.0
Category[Root/Sewer/Sidewalk]	-1.449	0.040	-1.532	-1.378	1.0

Supplementary Table 10: Regression coefficients for Zero-inflated Poisson regression with incident-level covariates and Borough fixed effects for Max Duration 200 days, Default repeat caller removal.

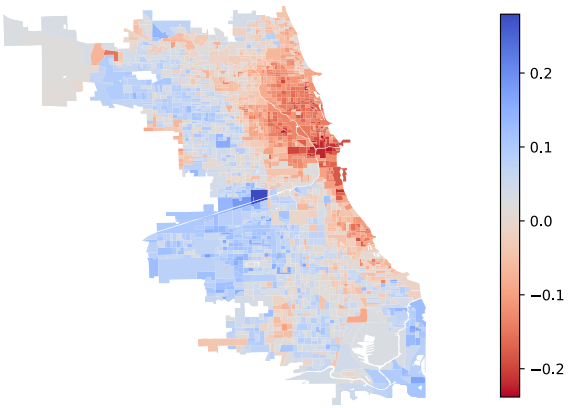
	Mean	StdDev	2.5%	97.5%	R_hat
Intercept	-3.495	0.029	-3.548	-3.440	1.0
Zero Inflation fraction	0.626	0.004	0.619	0.634	1.0
INSPCondition[T.Dead]	-0.462	0.034	-0.536	-0.401	1.0
INSPCondition[T.Excellent_Good]	-0.205	0.027	-0.254	-0.153	1.0
INSPCondition[T.Fair]	-0.123	0.026	-0.179	-0.077	1.0
INSP_RiskAssessment	0.238	0.011	0.216	0.258	1.0
Log(Tree Diameter at Breast Height)	-0.042	0.008	-0.059	-0.025	1.0
Borough[Bronx]	-0.139	0.027	-0.191	-0.093	1.0
Borough[Brooklyn]	-0.393	0.019	-0.433	-0.357	1.0
Borough[Manhattan]	0.512	0.050	0.405	0.596	1.0
Borough[Queens]	-0.244	0.019	-0.278	-0.208	1.0
Borough[Staten Island]	0.264	0.032	0.196	0.326	1.0
Category[Hazard]	1.482	0.018	1.445	1.515	1.0
Category[Illegal Tree Damage]	0.178	0.034	0.105	0.243	1.0
Category[Prune]	-0.087	0.027	-0.141	-0.034	1.0
Category[Remove Tree]	0.086	0.021	0.040	0.123	1.0
Category[Root/Sewer/Sidewalk]	-1.659	0.035	-1.730	-1.598	1.0

Supplementary Table 11: Regression coefficients for Zero-inflated Poisson regression with incident-level covariates and Borough fixed effects for Max Duration 100 days, Remove all repeat callers and missing caller information.

	Mean	StdDev	2.5%	97.5%	R_hat
Intercept	-3.267	0.029	-3.326	-3.214	1.0
Zero Inflation fraction	0.652	0.004	0.642	0.659	1.0
INSPCondtion[T.Dead]	-0.334	0.034	-0.401	-0.270	1.0
INSPCondtion[T.Excellent_Good]	-0.308	0.027	-0.359	-0.256	1.0
INSPCondtion[T.Fair]	-0.186	0.027	-0.241	-0.133	1.0
INSP_RiskAssessment	0.232	0.012	0.209	0.256	1.0
Log(Tree Diameter at Breast Height)	-0.028	0.008	-0.045	-0.013	1.0
Borough[Bronx]	-0.059	0.029	-0.116	-0.006	1.0
Borough[Brooklyn]	-0.378	0.021	-0.425	-0.339	1.0
Borough[Manhattan]	0.431	0.057	0.311	0.539	1.0
Borough[Queens]	-0.242	0.021	-0.286	-0.202	1.0
Borough[Staten Island]	0.248	0.031	0.185	0.307	1.0
Category[Hazard]	1.413	0.017	1.378	1.447	1.0
Category[Illegal Tree Damage]	0.237	0.035	0.157	0.302	1.0
Category[Prune]	-0.106	0.028	-0.164	-0.056	1.0
Category[Remove Tree]	0.025	0.022	-0.021	0.062	1.0
Category[Root/Sewer/Sidewalk]	-1.569	0.036	-1.643	-1.499	1.0

Supplementary Table 12: Regression coefficients for Zero-inflated Poisson regression with incident-level covariates and Borough fixed effects for Max Duration 100 days, Default repeat caller removal. Risk assessment scores binned according to levels of prioritization.

	Mean	StdDev	2.5%	97.5%	R_hat
Intercept	-3.177	0.205	-3.528	-2.872	1.0
Zero Inflation fraction	0.656	0.004	0.650	0.662	1.0
INSPCondition[T.Dead]	-0.339	0.033	-0.390	-0.285	1.0
INSPCondition[T.Excellent_Good]	-0.410	0.024	-0.451	-0.373	1.0
INSPCondition[T.Fair]	-0.188	0.025	-0.227	-0.147	1.0
Risk assessment category A	0.871	0.206	0.555	1.224	1.0
Risk assessment category B	0.594	0.183	0.290	0.916	1.0
Risk assessment category C	-0.297	0.189	-0.586	0.008	1.0
Risk assessment category D	-0.146	0.172	-0.400	0.158	1.0
Risk assessment category E	-1.206	0.834	-2.689	-0.018	1.0
Risk assessment category Unknown	0.184	0.175	-0.071	0.490	1.0
Log(Tree Diameter at Breast Height)	-0.005	0.010	-0.021	0.010	1.0
Borough[Bronx]	0.086	0.114	-0.122	0.268	1.0
Borough[Brooklyn]	-0.350	0.127	-0.553	-0.155	1.0
Borough[Manhattan]	0.765	0.181	0.445	1.048	1.0
Borough[Queens]	-0.600	0.085	-0.737	-0.461	1.0
Borough[Staten Island]	0.099	0.192	-0.219	0.416	1.0
Category[Hazard]	1.473	0.018	1.442	1.503	1.0
Category[Illegal Tree Damage]	0.220	0.035	0.159	0.276	1.0
Category[Prune]	-0.057	0.029	-0.105	-0.012	1.0
Category[Remove Tree]	0.052	0.022	0.016	0.088	1.0
Category[Root/Sewer/Sidewalk]	-1.687	0.040	-1.749	-1.624	1.0



Supplementary Figure 6: Coefficients for each census block group in Chicago, representing the combined association of socioeconomic variables (log income per capita, fraction of white residents, fraction of renter, median age, and fraction of residents with college degrees) on reporting rates.

Supplementary Figure 6 is a reproduction of Figure 1 using data from Chicago. Substantial spatial differences in reporting across different areas of the city are visible. Supplementary Figure 7 present the results when our model includes indicator variables for the over 2000 census tracts in New York City and the over 2000 census block groups in Chicago. These estimates capture more fine-grained spatial differences, beyond which may occur due to socioeconomic differences.

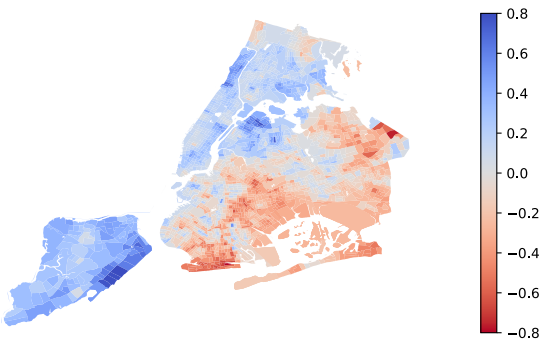
4.7 Replicating NYC results with public data

In the main text, we present results from NYC that were obtained from partially private data. In this subsection, we show that with all public data, we can nontheless reproduce all the results. In this subsection, we will refer to the data used there as the "full dataset" and the data used here as the "public dataset"

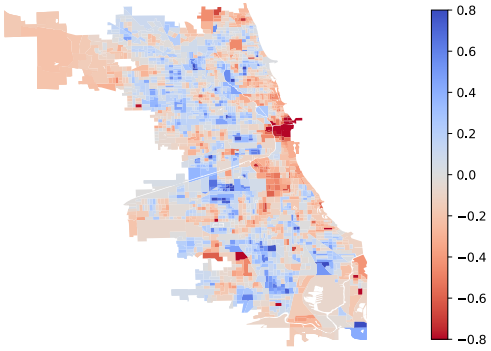
In the Methods section, we outlined that internal data were used to identify inspections and work orders associated with the service requests, and identify (anonymized) information about the caller. With the public dataset, we can still perform the first task, by joining the public service request data[7] with public inspection data[8], public work order data[9], and public risk assessment data[10]. During this process, we note that all of the covariates which we used in the previous analyses are retained. The second task, however, cannot be performed since public datasets do not contain any information about the caller. This means that we are unable to identify repeat callers and remove them in our analyses. This, however, does not change the final results qualitatively.

Summary of preprocessing The aforementioned public datasets are joined using unique ID's on service requests and inspections. These datasets are all actively maintained and updated, and the version used here contains 569,340 unique service requests made by the public within the period between 2/28/2015 and 9/01/2022. We then created an aggregated dataset as described

[7] https://data.cityofnewyork.us/Environment/Forestry-Service-Requests/mu46-p9is
[8] https://data.cityofnewyork.us/Environment/Forestry-Inspections/4pt5-3vv4
[9] https://data.cityofnewyork.us/Environment/Forestry-Inspections/4pt5-3vv4
[10] https://data.cityofnewyork.us/Environment/Forestry-Risk-Assessments/259a-b6s7

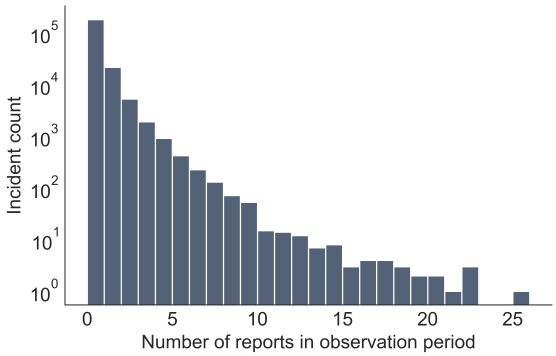


(a) Census tract coefficients in New York City

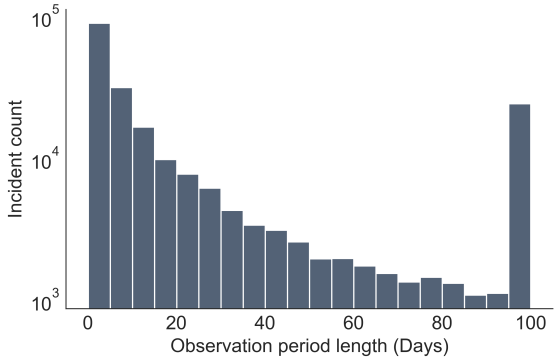


(b) Census block group coefficients in Chicago

Supplementary Figure 7: Coefficients on spatial variables in NYC and Chicago. These spatial coefficients are estimated using the ICAR spatial zero-inflated Poisson regression. More positive coefficients indicate higher reporting rates.



(a) Number of reports per incident.



(b) Length of observation period.

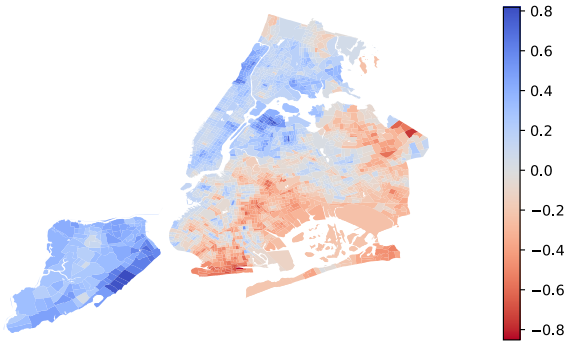
Supplementary Figure 8: Distribution of number of reports and length of observation for each unique incident in the aggregated dataset using public data.

in the Methods section on preprocessing. See Supplementary Table 13 for summary statistics, Supplementary Figure 8a and Supplementary Figure 8b for distribution of number of reports per incident and length of observation inverval.

Summary of results We replicate all the analyses presented in the main text on the public data set. Supplementary Table 14 summarizes results with the **Base** variables, while Supplementary Table 15 and Supplementary Table 16 serve as robustness checks. Supplementary Figure 9 shows the coefficients on **Spatial** variables. Supplementary Table 17, Supplementary Table 18, and Supplementary Table 20 show the coefficients on census tract **Socioeconomic** covariates.

Supplementary Table 13: Summary statistics from the NYC public dataset.

	Service requests	**Inspections**		**Incidents (from inspected reports)**		
		Inspected SRs	Fraction inspected	Unique incidents	Avg. reports per incident	Median Days to Inspection
Total number	569,340	363,676	0.64	246,744	1.47	6.53
By Borough						
Queens	235,363	152,985	0.65	107,524	1.42	5.99
Brooklyn	176,030	118,782	0.67	73,067	1.63	8.19
Staten Island	69,436	35,415	0.51	25,710	1.38	5.24
Bronx	54,839	38,336	0.70	26,110	1.47	7.89
Manhattan	33,659	18,156	0.54	14,346	1.27	2.82
By Category						
Hazard	236,593	179,477	0.76	118,135	1.52	2.80
Prune	124,725	42,929	0.34	34,305	1.25	17.97
Remove Tree	105,960	83,179	0.79	63,422	1.31	7.55
Root/Sewer/Sidewalk	72,900	41,837	0.57	33,450	1.25	38.07
Illegal Tree Damage	27,479	15,554	0.57	12,447	1.25	17.38
Other	1,683	700	0.42	645	1.09	5.70



Supplementary Figure 9: Coefficients on spatial variables in NYC, estimated with public data

Supplementary Table 14: Regression coefficients for Base variables in NYC, for Max Duration 100 days, estimated with public data

	Mean	StdDev	2.5%	97.5%	R_hat
Intercept	-3.449	0.028	-3.507	-3.401	1.0
Zero Inflation fraction	0.539	0.005	0.529	0.547	1.0
INSPCondition[T.Dead]	-0.163	0.035	-0.237	-0.103	1.0
INSPCondition[T.Excellent_Good]	-0.698	0.028	-0.755	-0.646	1.0
INSPCondition[T.Fair]	-0.436	0.025	-0.492	-0.387	1.0
INSP_RiskAssessment	0.262	0.013	0.235	0.287	1.0
Log(Tree Diameter at Breast Height)	0.079	0.011	0.053	0.097	1.0
Borough[Bronx]	0.029	0.026	-0.022	0.076	1.0
Borough[Brooklyn]	-0.178	0.017	-0.214	-0.147	1.0
Borough[Manhattan]	0.635	0.039	0.558	0.701	1.0
Borough[Queens]	-0.516	0.017	-0.549	-0.485	1.0
Borough[Staten Island]	0.030	0.029	-0.030	0.084	1.0
Category[Hazard]	1.583	0.019	1.547	1.621	1.0
Category[Illegal Tree Damage]	0.039	0.038	-0.033	0.111	1.0
Category[Prune]	-0.108	0.028	-0.166	-0.056	1.0
Category[Remove Tree]	-0.173	0.023	-0.216	-0.131	1.0
Category[Root/Sewer/Sidewalk]	-1.341	0.035	-1.413	-1.278	1.0

Supplementary Table 15: Regression coefficients for Base variables in NYC, for Max Duration 30 days, estimated with public data

	Mean	StdDev	2.5%	97.5%	R_hat
Intercept	-2.923	0.029	-2.987	-2.873	1.0
Zero Inflation fraction	0.618	0.004	0.610	0.626	1.0
INSPCondition[T.Dead]	-0.210	0.037	-0.282	-0.140	1.0
INSPCondition[T.Excellent_Good]	-0.691	0.032	-0.751	-0.627	1.0
INSPCondition[T.Fair]	-0.429	0.029	-0.482	-0.369	1.0
INSP_RiskAssessment	0.241	0.013	0.213	0.264	1.0
Log(Tree Diameter at Breast Height)	0.083	0.011	0.060	0.103	1.0
Borough[Bronx]	-0.099	0.026	-0.153	-0.051	1.0
Borough[Brooklyn]	-0.053	0.018	-0.088	-0.018	1.0
Borough[Manhattan]	0.536	0.038	0.455	0.604	1.0
Borough[Queens]	-0.374	0.018	-0.413	-0.339	1.0
Borough[Staten Island]	-0.010	0.029	-0.069	0.041	1.0
Category[Hazard]	1.458	0.019	1.415	1.492	1.0
Category[Illegal Tree Damage]	0.252	0.035	0.175	0.311	1.0
Category[Prune]	-0.191	0.033	-0.256	-0.126	1.0
Category[Remove Tree]	-0.241	0.026	-0.298	-0.193	1.0
Category[Root/Sewer/Sidewalk]	-1.277	0.040	-1.364	-1.211	1.0

Supplementary Table 16: Regression coefficients for Base variables in NYC, for Max Duration 200 days, estimated with public data

	Mean	StdDev	2.5%	97.5%	R_hat
Intercept	-3.670	0.027	-3.720	-3.614	1.0
Zero Inflation fraction	0.508	0.005	0.499	0.517	1.0
INSPCondition[T.Dead]	-0.116	0.037	-0.192	-0.049	1.0
INSPCondition[T.Excellent_Good]	-0.631	0.028	-0.690	-0.579	1.0
INSPCondition[T.Fair]	-0.389	0.027	-0.447	-0.340	1.0
INSP_RiskAssessment	0.256	0.012	0.233	0.279	1.0
Log(Tree Diameter at Breast Height)	0.059	0.011	0.039	0.079	1.0
Borough[Bronx]	0.130	0.024	0.079	0.173	1.0
Borough[Brooklyn]	-0.177	0.018	-0.216	-0.147	1.0
Borough[Manhattan]	0.585	0.039	0.495	0.657	1.0
Borough[Queens]	-0.525	0.018	-0.562	-0.492	1.0
Borough[Staten Island]	-0.013	0.028	-0.067	0.039	1.0
Category[Hazard]	1.595	0.017	1.562	1.624	1.0
Category[Illegal Tree Damage]	0.183	0.033	0.111	0.242	1.0
Category[Prune]	-0.209	0.023	-0.260	-0.167	1.0
Category[Remove Tree]	-0.148	0.025	-0.209	-0.107	1.0
Category[Root/Sewer/Sidewalk]	-1.422	0.032	-1.482	-1.360	1.0

Supplementary Table 17: Census Tract Socioeconomic coefficients, estimated alone in a regression alongside the incident-specific covariates, with all public data

	Mean	StdDev	2.5%	97.5%
Median age	-0.061	0.008	-0.078	-0.045
Fraction Hispanic	0.042	0.009	0.022	0.059
Fraction white	0.109	0.008	0.092	0.124
Fraction Black	-0.065	0.008	-0.083	-0.050
Fraction no high school degree	-0.007	0.009	-0.026	0.011
Fraction college degree	0.056	0.009	0.037	0.072
Fraction poverty	0.030	0.008	0.014	0.046
Fraction renter	0.126	0.010	0.104	0.143
Fraction family	-0.111	0.008	-0.129	-0.097
Log(Median house value)	0.085	0.009	0.066	0.101
Log(Income per capita)	0.050	0.010	0.030	0.067
Log(Density)	0.145	0.010	0.122	0.164

Supplementary Table 18: Census Tract Socioeconomic coefficients, estimated alone in a regression alongside the incident-specific covariates and the **borough fixed effects**, with all public data

	Mean	StdDev	2.5%	97.5%
Median age	-0.032	0.008	-0.050	-0.018
Fraction Hispanic	0.043	0.009	0.023	0.059
Fraction white	0.069	0.009	0.050	0.087
Fraction Black	-0.075	0.008	-0.092	-0.060
Fraction no high school degree	0.005	0.009	-0.014	0.023
Fraction college degree	0.027	0.009	0.009	0.043
Fraction poverty	-0.014	0.009	-0.033	0.002
Fraction renter	0.078	0.010	0.057	0.098
Fraction family	-0.045	0.010	-0.064	-0.026
Log(Median house value)	0.061	0.010	0.039	0.082
Log(Income per capita)	0.015	0.009	-0.005	0.032
Log(Density)	0.104	0.011	0.082	0.123

Supplementary Table 19: Census Tract Socioeconomic coefficients, estimated together in a regression alongside the incident-specific covariates, with all public data

	Mean	StdDev	2.5%	97.5%
Median age	-0.024	0.010	-0.046	-0.005
Fraction white	0.100	0.010	0.081	0.118
Fraction college degree	-0.018	0.014	-0.048	0.007
Fraction renter	0.142	0.010	0.123	0.160
Log(Income per capita)	0.073	0.014	0.047	0.103

Supplementary Table 20: Census Tract Socioeconomic coefficients, estimated together in a regression alongside the incident-specific covariates, with all public data. Compared with Supplementary Table 20, we additionally include (log) population density, which does not substantially affect the results.

	Mean	StdDev	2.5%	97.5%
Median age	-0.010	0.010	-0.030	0.009
Fraction white	0.110	0.010	0.091	0.129
Fraction college degree	-0.022	0.013	-0.050	0.004
Fraction renter	0.081	0.012	0.057	0.104
Log(Income per capita)	0.069	0.015	0.039	0.095
Log(Density)	0.116	0.011	0.092	0.137

4.8 Supplementary information on method validation

In this section, we first provide more information on the validation of model results using storm-inferred ground truth (in Section 5.2.3), in Supplementary Information 4.8.1. We further present two additional validation techniques in Supplementary Information 4.8.2 and Supplementary Information 4.8.3. Finally, we validate our estimated cumulative association of socioeconomic variables with reporting rate against a measure of civic engagement – participation in voting – in Supplementary Information 4.8.4.

4.8.1 Supplementary information on validation using reports after hurricanes

In Section 5.2.3, we validate our results using data immediately after Tropical Storm Isaias affected New York City, until 12 PM on 8/14/2020. Here we present results from similar analyses using data on Isaias with different end-time filters and from Tropical Storm Ida (Supplementary Figure 10). Note that we must filter for incidents only immediately after the storm, because new incidents, e.g., 2 months after the event may not have been caused by the storm, as opposed to events immediately after the storm.

For Tropical Storm Ida, we filter for service requests between 12 PM on 9/01/2021 and 12 PM on various days, as specified in Supplementary Figure 11, and define the true reporting delay as the time between the first report of an incident and 12 PM on 9/01/2021. We then estimate the reporting delays, similarly using coefficients on census tract and incident level variables. The individual incident level Pearson correlation between true reporting delays and model-estimated reporting delays is significant, while the means of them within each of the 30 bins defined on the model-estimated reporting delays exhibit even stronger levels of correlation, as shown in Supplementary Figure 11.

For other storm events that affected New York City in recent years, no significant increase in the amount of tree-related service requests was observed as in Supplementary Figure 12, and so we limit our analyses to these two hurricanes.

4.8.2 Supplementary information on validation using delays between first and second reports

Next, we use another form of validation: can our model predict (out-of-sample) true delays between the *first* and *second* report between incidents? If reporting rates are homogeneous (not changing over time for a given incident), then our model should be able to accurately recover these rates, and the model would similarly be accurate for estimating the (unknown) delay before the first report. We analyze service requests submitted to DPR from 9/1/2020 to 8/31/2022 (after the model training data period). For each incident that received 2 or more service requests, we calculate the time between the first and second reports and estimate their reporting delay using the spatial model.

Supplementary Information 4.8.2 shows the relationship between model-estimated delays and observed delays between the first and second reports when model-estimated delays are binned. There is a strong correlation, and predictions are approximately on the $y = x$ line.

4.8.3 Supplementary information on using a training set starting from the second report

In Section 5.2.3 and in Supplementary Information 4.8.1, the predictions were generated using the results of the spatial model we presented in the main text. The training set used to obtain such

results was constructed by defining the time of the first report as the start of our observation period, as detailed in Supplementary Information 4.2.

However, our theoretical analysis also holds if we instead use as the Poisson interval the time starting after the *second* report – if we split a Poisson process on the time of the first jump, and start counting subsequent jumps at that time, the resulting counting process would still be a Poisson process with the same rate. Thus, we can estimate the Poisson process rate using our methods, starting with the second report time.[11] However, crucially, we as researchers know for the incident the *time between the first and second report*, which was not used to train the model (but which the model is trying to predict). Intuitively, this procedure evaluates our methods in a setting in which the ground truth is known.

Formalizing this idea, we further validate our empirical approach as follows. We first construct a training set from the same raw public data introduced in Supplementary Information 4.7, with the time of the second report as the start of our observation period, and the end of our observation period defined analogous to Equation (33), but with 20, 30, and 50 days as the maximum duration.[12] We then use this alternative training set to estimate the coefficients for the set of **Base** covariates, and then use these coefficients to estimate a reporting delay for each incident. The reason we only included the **Base** covariates is that the training set constructed in this way only retains 16,460 incidents, which makes it hard to estimate the effect of high-dimensional covariates, such as the census tract spatial variables. Finally, we compare these estimates to the observed time between the first and second reports for each incident.

On the individual incident level, our estimates and the true delays between the first and second reports are significantly correlated. Supplementary Figure 14 showcase a comparison of the means when all incidents are binned according to their model-estimated delays. We note that though the choice of the maximum duration of observation interval affects the specific model predictions, all three choices produced reasonable estimates that are close to the true delays on the group level. (We use a smaller maximum duration as these incidents are on average higher risk, and we only start counting after the second report).

4.8.4 Supplementary validation of census tract socioeconomic coefficients estimates

The disparities in reporting behavior along socioeconomic variables highlight *individual-level* behavioral heterogeneity in resident crowdsourcing, and civic engagement at large. For example, O'Brien et al. [28] use the reporting delay as part of their measure for the 'civic response rate', that relates to other forms of activities as well. The level of civic engagement can be measured by many other means, chief among which is participation in political voting. In this section, we validate our coefficient estimates on socioeconomic variables using this idea.

Voter-level public data on participation in voting is generally scarce in New York; however, the NYC Campaign Finance Board published a dataset containing participation scores of more

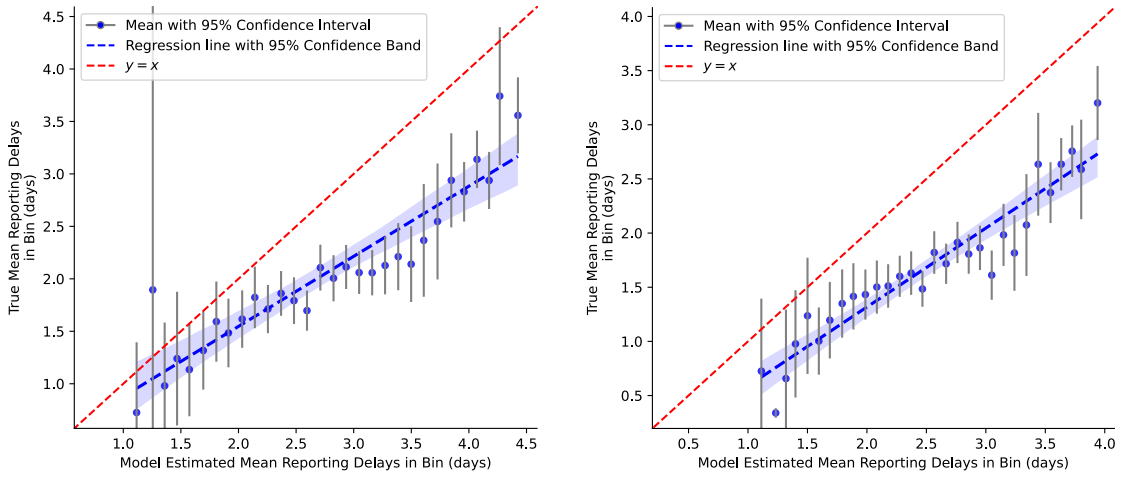
[11] Formally, consider a Poisson process with rate λ, denoted as $\{X(t), t \geq 0\}$, with the first jump time denoted as T_1. Based on the memoryless property of Poisson process jump times, the process $\{X(t) - 1, t \geq T_1\}$ is still a Poisson process with the same rate parameter (see e.g., [33]). Thus in our context, for an incident with type θ, if we treat the first report as the start of a Poisson process, treat the second report as equivalent to the first report, and count duplicates starting at the third report, this reporting process is still a Poisson process with rate λ_θ, but now the start of this process is observed.

[12] We note that in Theorem 1, we require the end of the observation period to be before incident death time. The maximum duration is thus a design choice to alleviate the effect of the incident being resolved before an inspection or a work order. Using the time of the second report as the start of the observation interval requires us to set the maximum duration at a smaller value compared to using the time of the first report, in order to have the equivalent effect. Furthermore, incidents with two or more reports are likely incidents with higher than average urgency, and thus more likely to be resolved than outside sources, which further justifies the choice of shorter maximum duration.

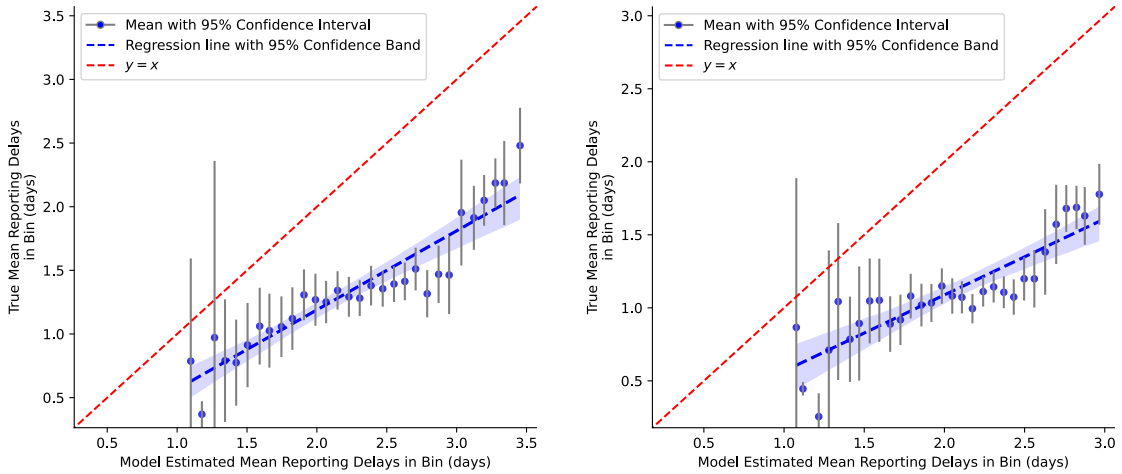
than 4 million voters calculated based on their voting history from 2008 up to 2018, along with the census tract in which they resided in 2010.[13] We calculate cumulative association scores [14]) for each 2010 census tract, using the corresponding census data. We compare these scores with the average voter participation score in each tract.Our model-estimated mean engagement score is significantly correlated with the mean voter participation score as shown in Supplementary Figure 15. This shows that our estimated disparities in 311 system usage relate to other forms of civic participation and representation.

[13] https://data.cityofnewyork.us/City-Government/Voter-Analysis-2008-2018/psx2-aqx3

[14] Since we are focusing on individual-level behavior, for that figure and this analysis we do not use the density variable, as that does not reflect individual-level characteristics, and only use the remaining five: Median age, Fraction white, Fraction college degree, Fraction renter, log of per capita income

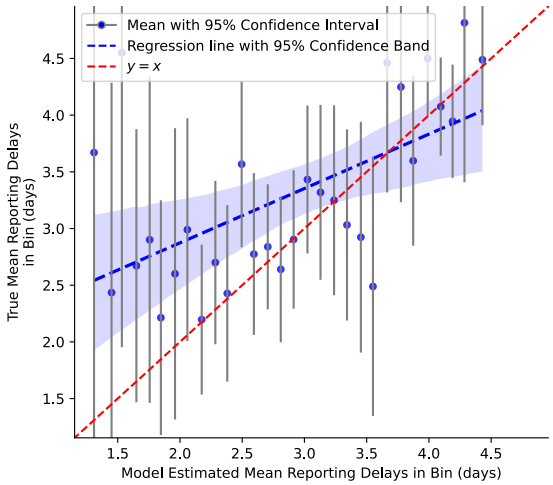


(a) Filter for reports before 12 PM on 8/13/2020.(b) Filter for reports before 12 PM on 8/12/2020.

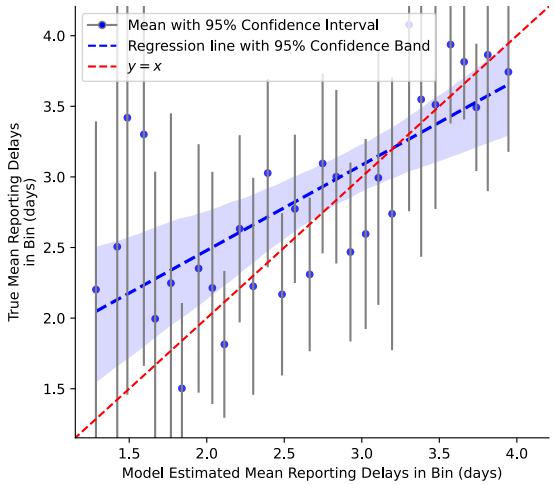


(c) Filter for reports before 12 PM on 8/11/2020.(d) Filter for reports before 12 PM on 8/10/2020.

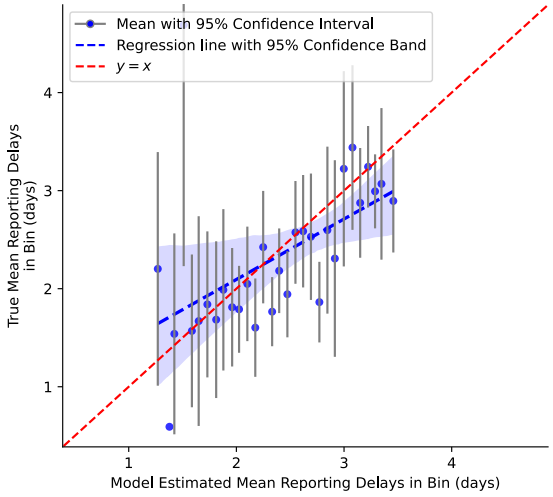
Supplementary Figure 10: Replication of Figure 2 using data with different filtering for the end date. (a) Individual incident level Pearson $r = 0.196$ (Beta$(3956.5, 3956.5) = 0.598, P < 0.001$), bin level Pearson $r = 0.925$ (Beta$(14, 14) = 0.962, P < 0.001$). (b) Individual incident level Pearson $r = 0.192$ (Beta$(3779, 3779) = 0.596, P < 0.001$), bin level Pearson $r = 0.945$ (Beta$(14, 14) = 0.973, P < 0.001$). (c) Individual incident level Pearson $r = 0.176$ (Beta$(3526.5, 3526.5) = 0.590, P < 0.001$), bin level Pearson $r = 0.930$ (Beta$(14, 14) = 0.965, P < 0.001$).(d) Individual incident level Pearson $r = 0.159$ (Beta$(3321.5, 3321.5) = 0.579, P < 0.001$), bin level Pearson $r = 0.874$ (Beta$(14, 14) = 0.937, P < 0.001$).



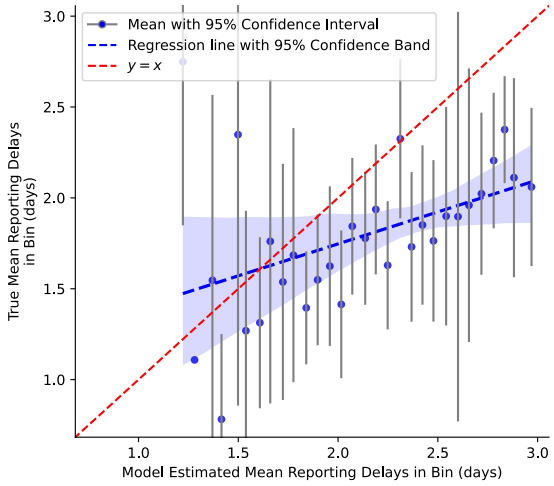
(a) Filter for reports before 12 PM on 9/10/2021.



(b) Filter for reports before 12 PM on 9/9/2021.

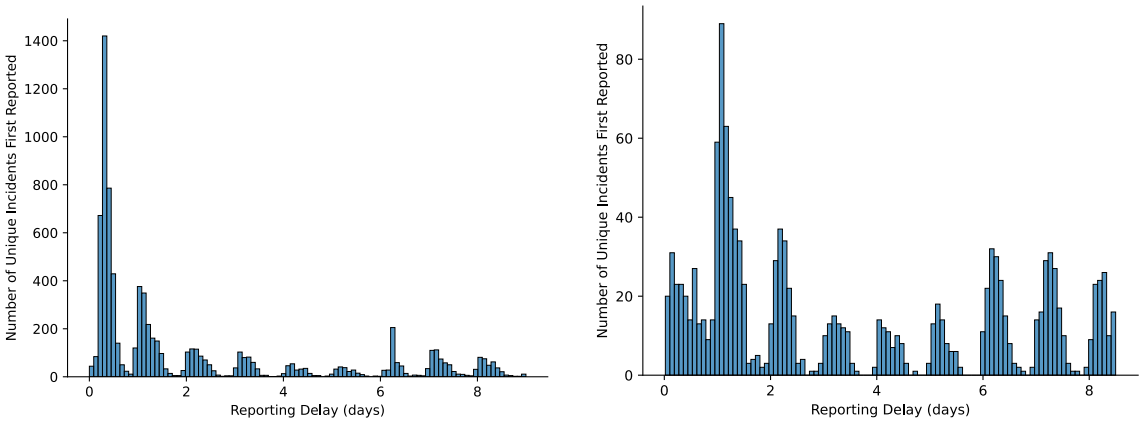


(c) Filter for reports before 12 PM on 9/8/2021.



(d) Filter for reports before 12 PM on 9/7/2021.

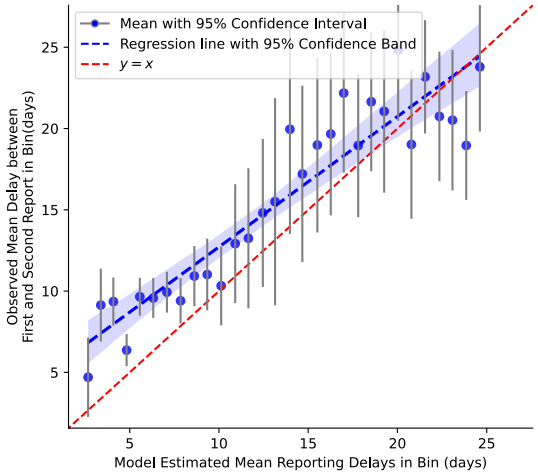
Supplementary Figure 11: Replication of Figure 2 using data after Tropical Storm Ida, with different end date specifications. (a) Individual incident level Pearson $r = 0.222$ (Beta$(679.5, 679.5) = 0.598, P < 0.001$), bin level Pearson $r = 0.576$ (Beta$(14, 14) = 0.788, P < 0.001$). (b) Individual incident level Pearson $r = 0.226$ (Beta$(626, 626) = 0.613, P < 0.001$), bin level Pearson $r = 0.689$ (Beta$(14, 14) = 0.845, P < 0.001$). (c) Individual incident level Pearson $r = 0.228$ (Beta$(552, 552) = 0.614, P < 0.001$), bin level Pearson $r = 0.518$ (Beta$(14, 14) = 0.759, P = 0.003$). (d) Individual incident level Pearson $r = 0.159$ (Beta$(477.5, 477.5) = 0.580, P < 0.001$), bin level Pearson $r = 0.446$ (Beta$(14, 14) = 0.723, P = 0.013$).



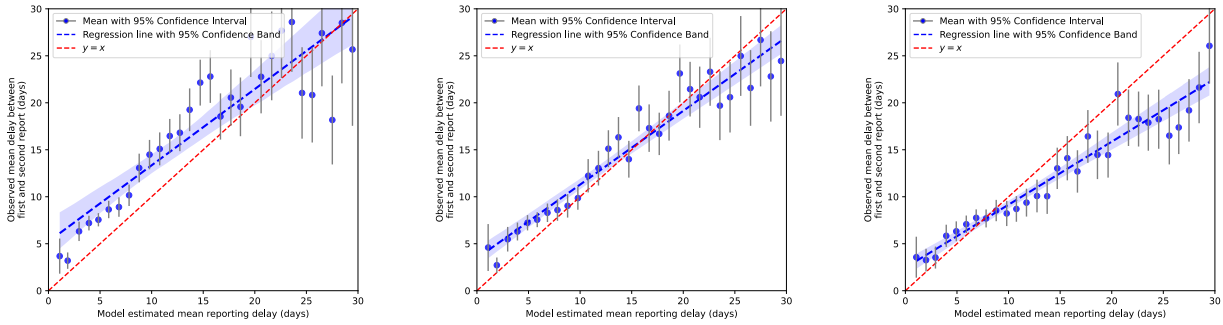
(a) Histogram of reporting delay after Isaias.

(b) Histogram of reporting delay after Ida.

Supplementary Figure 12: Histogram of reporting delays of unique incidents after Isaias and Ida. After each hurricane, we immediately observe a surge in reported incidents.

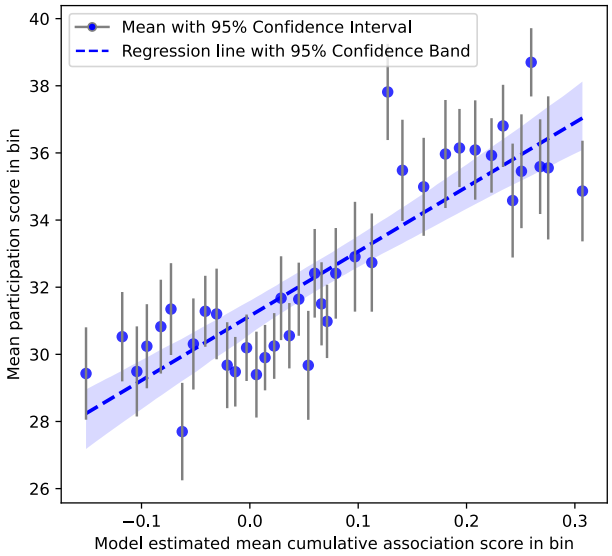


Supplementary Figure 13: Comparison of observed delay between first and second reports and model estimated reporting delays, based on data between 9/1/2020 and 8/31/2022. Model-estimated delays are obtained using the spatial model. All incidents are categorized into 30 bins based on their estimated reporting delays, and we calculate the means of true and estimated delays within each bin. Individual level Pearson $r = 0.21$ ($\text{Beta}(3802, 3802) = 0.604, P < 0.001$), bin level Pearson $r = 0.931$ ($\text{Beta}(14, 14) = 0.965, P < 0.001$)



(a) Maximum duration 20 days. (b) Maximum duration 30 days. (c) Maximum duration 50 days.

Supplementary Figure 14: Comparison of observed delays between first and second reports and model-estimated reporting delays, which are estimated using a model trained with data constructed by defining the start of the observation interval as the time of the second report, and the end as in Equation (33), but with maximum duration set as 20 days, 30 days, and 50 days, respectively. All incidents are then categorized into 30 bins based on their model-estimated reporting delays, and we calculate the means of true and estimated delays within each bin. (a) Individual incident level Pearson $r = 0.295$ (Beta$(6781.5, 6781.5) = 0.648, P < 0.001$), bin level Pearson $r = 0.905$ (Beta$(14, 14) = 0.952, P < 0.001$). (b) Individual incident level Pearson $r = 0.292$ (Beta$(6781.5, 6781.5) = 0.646, P < 0.001$), bin level Pearson $r = 0.967$ (Beta$(14, 14) = 0.983, P < 0.001$). (c) Individual incident level Pearson $r = 0.261$ (Beta$(6781.5, 6781.5) = 0.631, P < 0.001$), bin level Pearson $r = 0.962$ (Beta$(14, 14) = 0.981, P < 0.001$).



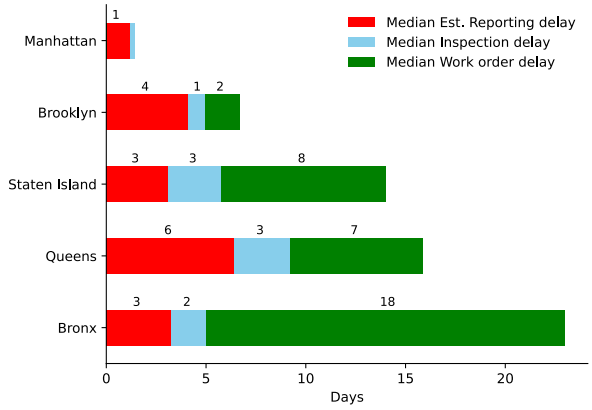
Supplementary Figure 15: Relationship between voter participation rates in a census tract, and our model estimated cumulative association of socioeconomic variables with reporting rate. At the bin level, Pearson $r = 0.861$ (Beta$(19, 19) = 0.930, P < 0.001$)

4.9 Additional information for comparative delay analysis

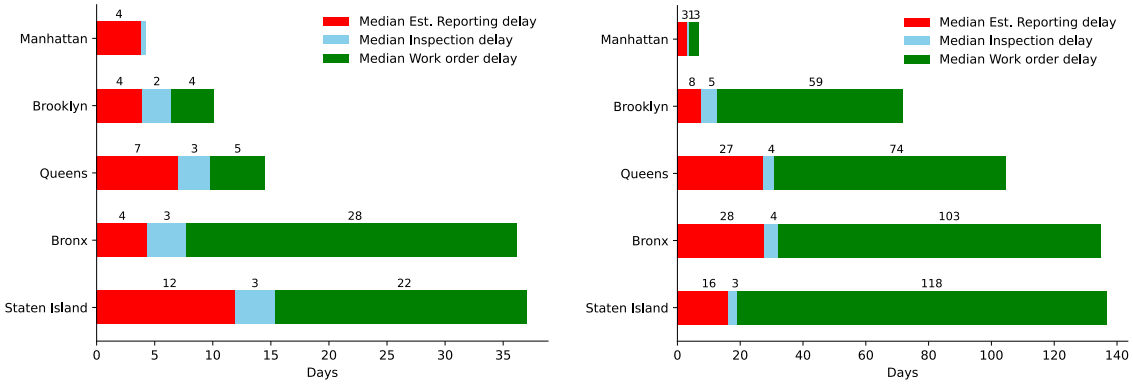
In this section, we present results with alternative analysis choices for applications of reporting delay quantification: imputing missing values (with infinite delays), other risk levels, analyzing by request category instead of risk prioritization level, and analyzing all incidents instead of the subset that was inspected or worked on; results are qualitatively similar.

For each unique incident, inspection delays are measured as the time between the first inspection (if it was inspected) and the first service request for that incident; work order delays are measured as the time between actually finishing the work order (if it was completed) and first inspection; reporting delays are *estimated* – for each incident, we estimate the mean of an Exponential random variable with reporting rate our model estimates for an incident of the given characteristics. In this plot, we only consider incidents that are inspected, and ignore missing values; in each Borough over 89% of such high-risk inspected incidents eventually have a completed work order.

Similar to analyses in the main text, we present results with other risk prioritization groups.



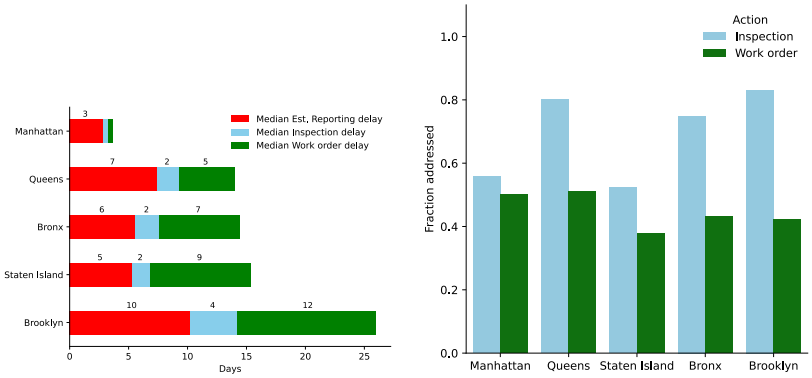
Supplementary Figure 16: Same as Figure 3a (highest risk prioritization A) except with incomplete work orders imputed as having infinite delays



(a) Risk prioritization B

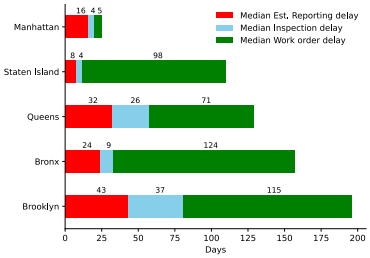
(b) Risk prioritization C

Supplementary Figure 17: Same as Figure 3a (highest risk prioritization, not imputed missing values) except for other risk prioritization levels. Risk prioritization group D is omitted due to a low percentage of incidents that eventually got worked on.

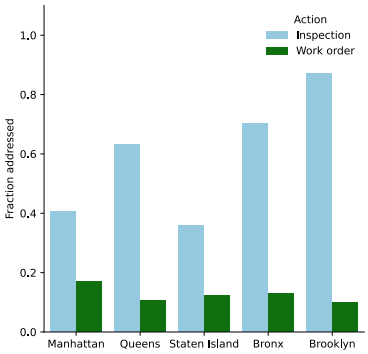


(a) *Hazard* delays (b) *Hazard* fraction addressed

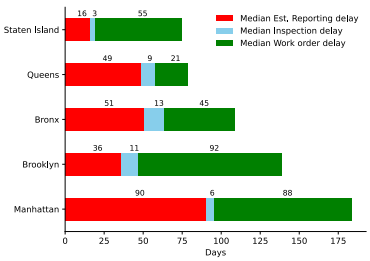
Supplementary Figure 18: We repeat the analysis except, instead of separating out by *Risk prioritization level*, we separately analyze each *report-time Service Request Category* (here, we show the *Hazard* category. Since we have this feature for *all service requests* (not just those inspected, as in the case where we use the risk prioritization covariate), we can (1) estimate reporting delays for *all* incidents after training a model using just report-time incidents (but on the inspected set, due to labeling of duplicates; and (2) we can also report the fraction of incidents of this category that were inspected and worked on, respectively, as we do in (b). We get qualitatively similar results as in the main text: the ordering of Borough end-to-end delays (as well as their individual parts) is about the same. The same Boroughs with higher work order delays just counting the incidents that were actually worked on (Staten Island, Bronx, Brooklyn) also have lower fractions of incidents that were addressed. Figure (a) does not impute missing values, since (b) reports the fraction inspected and worked on, respectively.



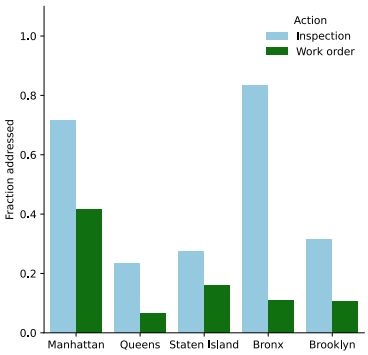
(a) *Illegal tree damage* delays



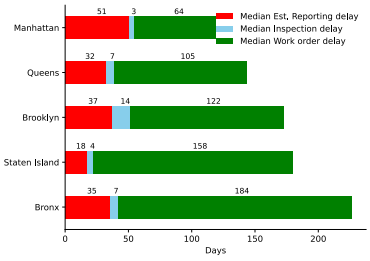
(b) *Illegal tree damage* fraction addressed



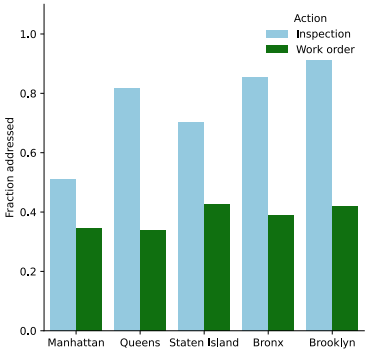
(c) *Prune* delays



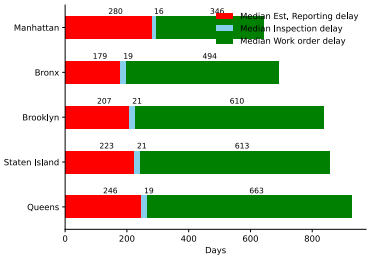
(d) *Prune* fraction addressed



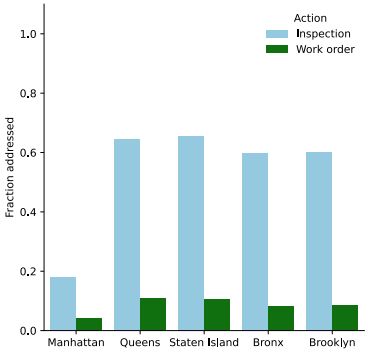
(e) *Remove Tree* delays



(f) *Remove Tree* fraction addressed

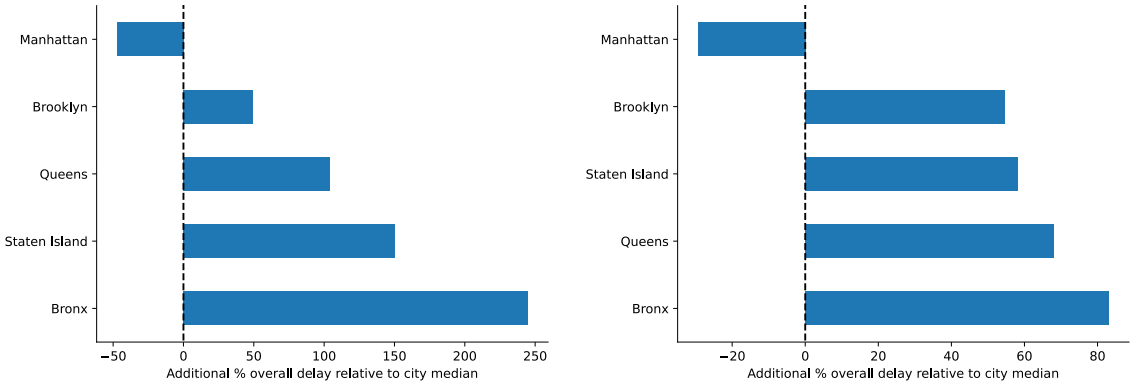


(g) *Root/Sewer/Sidewalk* delays



(h) *Root/Sewer/Sidewalk* fraction addressed

Supplementary Figure 19: Same as Supplementary Figure 18 with other *Service Reqest Categories*.



(a) Risk prioritization B

(b) Risk prioritization C

Supplementary Figure 20: Same as Figure 3b except for other risk prioritization levels. Risk prioritization group D is omitted due to a low percentage of incidents that eventually got worked on.

4.10 Preprocessing for Chicago dataset

In this section, we detail the preprocessing done on the Chicago dataset before model training.

For each request, we also have location and incident-level covariates. We further observe whether and when this service request is marked as 'completed', 'open' or 'canceled', and whether this is a duplicate request; if the request is indeed a duplicate, CDOT and CDWM mark which other requests refer to the same incident. At most three timestamps are associated with each service request: for 'open' service requests, we observe the created time and last modified time; for 'completed' and 'canceled' service requests, we further observe the closed time. One crucial difference in the Chicago dataset, compared with New York City, is that duplicate reports on "open" incidents are also marked. This requires different treatments in constructing observation intervals, as we detail below.

Constructing an observation interval $(S_i, E_i]$ For Chicago reports, we make the following choice. Let $\tilde{t}_i$ be the first time an incident is reported, t_i^{CLOSED} be the time an incident is marked closed, and $\bar{t}$ be the time that we retrieved the dataset (and that it was last updated), 7/4/2022 at 21:40 EDT. Then, E_i of each incident i is:

$$E_i = \min\left\{100 \text{ days} + \tilde{t}_i,\ t_i^{\text{CLOSED}},\ \bar{t}\right\}. \tag{34}$$

The choice of adding $\bar{t}$ ensures that for "open" incidents, the observation interval is not overly long, which could bias our estimates for the reporting rate downward.

Though the Chicago dataset contains reports that were generated as early as 7/1/2018, it was not until 2/27/2019 that reports started to have a closed time associated with them[15]. As a consequence, we filter out any reports made prior to 3/1/2019 and filter out "E-scooter" and "Vehicle Parked in Bike Lane Complaint" incidents, for which no duplicates are marked. After this, we are left with 949,352 reports, which represent 698,365 unique incidents. Supplementary Table 21 lists some summary statistics of this dataset at this stage. We further filter out any incidents that have a negative or extremely short (< 0.01 days) observation interval. These most likely represent human errors in logging the time. This leaves us with a total of 575,882 unique incidents.

Covariate selection and processing Next, we select the covariates that compose type θ. Similar to the NYC dataset, most service requests in Chicago come with latitude-longitude coordinates, using which we identified which of the over 2,000 census block groups in Chicago this incident is in, through the FCC API. The reason we used census block groups in Chicago instead of census tracts, which were used in NYC, is that the Chicago dataset contains far more incidents to allow finer-grained analyses; further, there are approximately as many census block groups in Chicago as there are census tracts in NYC. We then join this information with the 2020 Census Data from IPUMS NHGIS. Finally, we log transform several variables, filter out the incidents for which any of the covariates are missing, and standardize all data. During this process, we filter out 10,452 incidents (1.80% of the total number of incidents), due to either missing covariates or unable to match them to census tracts, and are left with 565,430 unique incidents for our further analysis, which represent 794,132 unique service requests. Supplementary Figure 21a shows the histogram

[15] In private correspondence with the team responsible for the maintenance of this dataset, they pointed out that the current version of the 311 system went live on 12/18/2018, and that the loss of data is likely due to both migration of data from an older version of the system and not correctly logging data at the very beginning of the current version. However, they also confirmed that beyond 2/28/2019, the integrity of data should not be a concern.

Supplementary Table 21: Summary statistics from the Chicago dataset, after filtering out service requests prior to 2019/03/01 and categories with no duplicates. There are a total of 28 categories and we selected the top 15 in terms of number of reports. We note that the variation in median days to completion is large in this dataset, and some categories have extremely short completion time. The subsequent filtering of incidents with short duration addresses the concern for mislabeling incidents as completed.

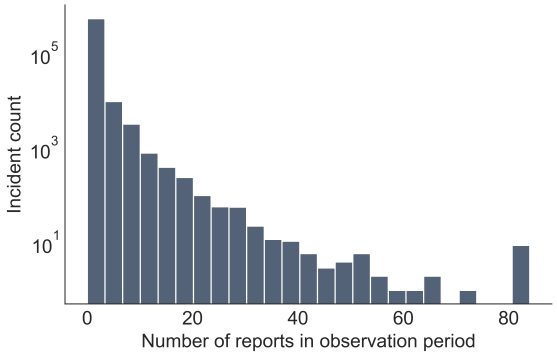
	Service requests	**Completions**		**Incidents (from all reports)**		
		Completed SRs	Percentage completed	Unique incidents	Avg. reports per incident	Median Days to Completion
Total number	949,352	853,989	0.9	698,365	1.36	6.676
By Owner Department						
CDOT	789,063	720,925	0.91	559,777	1.41	6.783
CDWM	160,289	133,064	0.83	138,969	1.15	5.704
By Top 15 Categories						
Street Light Out Complaint	225,408	209,441	0.93	122,302	1.84	9.786
Pothole in Street Complaint	200,132	189,879	0.95	138,832	1.44	10.450
Sign Repair Request - All Other Signs	96,019	88,775	0.92	91,295	1.05	0.002
Traffic Signal Out Complaint	75,828	69,443	0.92	58,705	1.29	0.283
Alley Light Out Complaint	53,010	49,329	0.93	31,831	1.67	65.671
Sewer Cleaning Inspection Request	39,120	34,390	0.88	36,056	1.08	24.965
Alley Pothole Complaint	36,333	33,981	0.94	25,528	1.42	28.973
Water On Street Complaint	29,494	25,605	0.87	26,048	1.13	5.517
Sidewalk Inspection Request	25,515	9,821	0.38	21,724	1.17	323.228
Open Fire Hydrant Complaint	25,105	21,103	0.84	15,548	1.61	0.313
Sewer Cave-In Inspection Request	22,308	18,301	0.82	21,300	1.05	58.093
Snow – Uncleared Sidewalk Complaint	19,532	18,519	0.95	17,338	1.13	6.810
Water in Basement Complaint	16,952	15,563	0.92	16,292	1.04	0.821
Sign Repair Request - Stop Sign	16,850	16,746	0.99	16,023	1.05	0.128
Street Light Pole Damage Complaint	16,171	14,393	0.89	15,194	1.06	1.204

of the number of reports per incident during the observation interval; Supplementary Figure 21b shows the distribution of durations. Supplementary Table 22 lists the covariates we use.

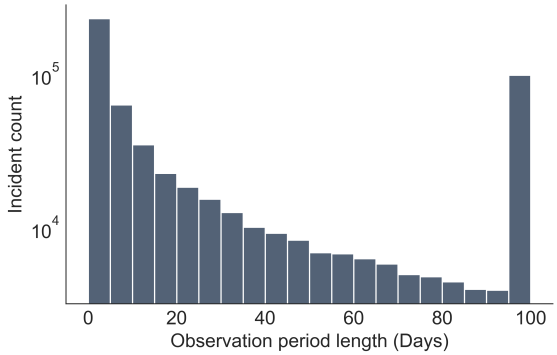
Sampling the dataset for tractability Due to computational constraints, when training the Stan models on Chicago data, we randomly sample 100,000 incidents for each model run.

4.11 Results from Chicago dataset

In this section, we provide results from applying our methods to the Chicago dataset. Supplementary Figure 22 confirms the posterior samples from the zero-inflated model with Base variables reasonably match the observed distribution; Supplementary Table 23 lists the coefficients for the set of Base variables; Supplementary Figure 7b illustrates the coefficients on census tract Spatial covariates; Supplementary Table 26 and Supplementary Table 28 show the coefficients on census tract Socioeconomic covariates.



(a) Number of reports per incident.

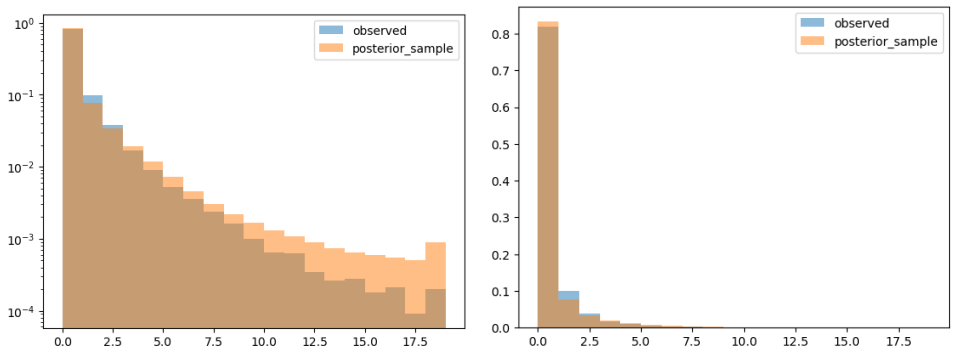


(b) Length of observation period.

Supplementary Figure 21: Distribution of number of reports and length of observation for each unique incident in the Chicago aggregated dataset. For most incidents, there are no reports after the first report (at least not in the observation period). There is one incident not included in the histogram of number of reports that attracted 162 reports, which happened to be a malfunctioning traffic light at a busy junction. There is a peak at 100 days for the observation period, due to our configuration in Equation (34).

Supplementary Table 22: Description of covariates in the Chicago aggregated dataset

Covariate	Description
Incident Global ID	An identifier unique to each incident.
Duration	The observation duration as defined in Equation (34)
Owner Department	Which department (CDOT or DWM) the service request is directed to.
Service Request Type	The incident type as reported.
Created Month	The month that the first report of each incident came in.
Census Block Group	Which census block group the incident occurred in as reported.
Median Age	Median age in the census block group.
Fraction Hispanic	Fraction of residents that identify as Hispanic in the census block group.
Fraction white	Fraction of residents that identify as white in the census block group.
Fraction Black	Fraction of residents that identify as Black in the census block group.
Fraction no high school degree	Fraction of residents that have not graduated from high school in the census block group.
Fraction college degree	Fraction of residents that have graduated from college in the census block group.
Fraction poverty	Fraction of residents that are identified to be in poverty in the census block group.
Fraction renter	Fraction of residents that rent their current residence in the census block group.
Fraction family	Fraction of family household in the census block group.
Median household value	Median value of household in the census block group.
Income per capita	Income per capita of residents in the census block group.
Density	Population density in the census block group.



Supplementary Figure 22: Comparison between posterior distributions sampled from the Zero-inflated Poisson regression model and the observed distribution in the data. The left-hand side plot is in log scale for the y axis, while the right-hand side plot is in the natural scale. The posterior samples reasonably match the observed distribution. Note that due to sampling of the full dataset, the observed distribution is not entirely the same as in Supplementary Figure 21a.

Supplementary Table 23: Regression coefficients for Base variables in Chicago, for Max Duration 100 days

	Mean	StdDev	R_hat
Intercept	-4.053	0.061	1.0
Zero Inflation fraction	0.539	0.003	1.0
Category[Sewer Cave-In Inspection Request]	-1.645	0.094	1.0
Category[Water On Street Complaint]	0.523	0.078	1.0
Category[Water in Basement Complaint]	0.627	0.115	1.0
Category[Alley Sewer Inspection Request]	-0.908	0.127	1.0
Category[Sewer Cleaning Inspection Request]	-1.018	0.074	1.0
Category[Pavement Cave-In Inspection Request]	-2.996	1.122	1.0
Category[Protected Bike Lane - Debris Removal]	-0.308	0.285	1.0
Category[Sidewalk Inspection Request]	-2.246	0.106	1.0
Category[Sign Repair Request - Stop Sign]	2.764	0.111	1.0
Category[Sign Repair Request - One Way Sign]	2.147	0.212	1.0
Category[Sign Repair Request - Do Not Enter Sign]	1.080	0.474	1.0
Category[Sign Repair Request - All Other Signs]	-1.413	0.077	1.0
Category[Bicycle Request/Complaint]	-0.878	0.184	1.0
Category[Alley Pothole Complaint]	0.265	0.067	1.0
Category[Pothole in Street Complaint]	0.831	0.062	1.0
Category[Alley Light Out Complaint]	0.383	0.064	1.0
Category[Traffic Signal Out Complaint]	2.569	0.064	1.0
Category[Viaduct Light Out Complaint]	-0.491	0.163	1.0
Category[Street Light Out Complaint]	1.767	0.061	1.0
Category[Street Light Pole Damage Complaint]	0.680	0.117	1.0
Category[Street Light On During Day Complaint]	-0.267	0.106	1.0
Category[Street Light Pole Door Missing Complaint]	-2.499	0.280	1.0
Category[Snow Removal - Protected Bike Lane or ...	-2.457	0.813	1.0
Category[Snow – Uncleared Sidewalk Complaint]	0.286	0.086	1.0
Category[No Water Complaint]	2.433	0.124	1.0
Category[Low Water Pressure Complaint]	-0.961	0.168	1.0
Category[Open Fire Hydrant Complaint]	4.221	0.065	1.0
Category[Water Quality Concern]	-2.491	0.397	1.0

Supplementary Table 24: Regression coefficients for Base variables in Chicago, for Max Duration 30 days

	Mean	StdDev	R_hat
Intercept	-3.308	0.048	1.0
Zero Inflation fraction	0.620	0.003	1.0
Category[Sewer Cave-In Inspection Request]	-1.356	0.089	1.0
Category[Water On Street Complaint]	0.592	0.069	1.0
Category[Water in Basement Complaint]	0.782	0.109	1.0
Category[Alley Sewer Inspection Request]	-0.373	0.116	1.0
Category[Sewer Cleaning Inspection Request]	-0.733	0.065	1.0
Category[Pavement Cave-In Inspection Request]	-1.494	0.664	1.0
Category[Protected Bike Lane - Debris Removal]	-0.219	0.344	1.0
Category[Sidewalk Inspection Request]	-2.156	0.120	1.0
Category[Sign Repair Request - Stop Sign]	2.514	0.094	1.0
Category[Sign Repair Request - One Way Sign]	1.666	0.222	1.0
Category[Sign Repair Request - Do Not Enter Sign]	0.194	0.500	1.0
Category[Sign Repair Request - All Other Signs]	-1.189	0.068	1.0
Category[Bicycle Request/Complaint]	-0.317	0.166	1.0
Category[Alley Pothole Complaint]	0.086	0.060	1.0
Category[Pothole in Street Complaint]	0.542	0.050	1.0
Category[Alley Light Out Complaint]	0.070	0.054	1.0
Category[Traffic Signal Out Complaint]	2.733	0.051	1.0
Category[Viaduct Light Out Complaint]	-0.639	0.183	1.0
Category[Street Light Out Complaint]	1.490	0.049	1.0
Category[Street Light Pole Damage Complaint]	0.190	0.118	1.0
Category[Street Light On During Day Complaint]	-0.968	0.124	1.0
Category[Street Light Pole Door Missing Complaint]	-1.309	0.222	1.0
Category[Snow Removal - Protected Bike Lane or ...	-1.584	0.562	1.0
Category[Snow – Uncleared Sidewalk Complaint]	0.037	0.076	1.0
Category[No Water Complaint]	1.849	0.110	1.0
Category[Low Water Pressure Complaint]	-1.258	0.184	1.0
Category[Open Fire Hydrant Complaint]	3.686	0.055	1.0
Category[Water Quality Concern]	-2.835	0.597	1.0

Supplementary Table 25: Regression coefficients for Base variables in Chicago, for Max Duration 200 days

	Mean	StdDev	R_hat
Intercept	-4.448	0.082	1.0
Zero Inflation fraction	0.496	0.003	1.0
Category[Sewer Cave-In Inspection Request]	-1.537	0.107	1.0
Category[Water On Street Complaint]	0.448	0.094	1.0
Category[Water in Basement Complaint]	1.050	0.113	1.0
Category[Alley Sewer Inspection Request]	-0.823	0.140	1.0
Category[Sewer Cleaning Inspection Request]	-1.081	0.092	1.0
Category[Pavement Cave-In Inspection Request]	-2.813	1.231	1.0
Category[Protected Bike Lane - Debris Removal]	0.070	0.340	1.0
Category[Sidewalk Inspection Request]	-2.287	0.111	1.0
Category[Sign Repair Request - Stop Sign]	2.458	0.113	1.0
Category[Sign Repair Request - One Way Sign]	2.830	0.202	1.0
Category[Sign Repair Request - Do Not Enter Sign]	-1.655	0.933	1.0
Category[Sign Repair Request - All Other Signs]	-1.418	0.094	1.0
Category[Bicycle Request/Complaint]	-1.372	0.189	1.0
Category[Alley Pothole Complaint]	0.378	0.085	1.0
Category[Pothole in Street Complaint]	0.976	0.082	1.0
Category[Alley Light Out Complaint]	0.515	0.084	1.0
Category[Traffic Signal Out Complaint]	2.661	0.083	1.0
Category[Viaduct Light Out Complaint]	-0.298	0.152	1.0
Category[Street Light Out Complaint]	2.026	0.082	1.0
Category[Street Light Pole Damage Complaint]	0.987	0.124	1.0
Category[Street Light On During Day Complaint]	-0.668	0.118	1.0
Category[Street Light Pole Door Missing Complaint]	-1.728	0.188	1.0
Category[Snow Removal - Protected Bike Lane or ...	-2.747	1.161	1.0
Category[Snow – Uncleared Sidewalk Complaint]	0.877	0.099	1.0
Category[No Water Complaint]	2.342	0.131	1.0
Category[Low Water Pressure Complaint]	-0.877	0.186	1.0
Category[Open Fire Hydrant Complaint]	4.165	0.085	1.0
Category[Water Quality Concern]	-2.479	0.426	1.0

Supplementary Table 26: Census Block Group Socio-economic coefficients in Chicago, estimated alone in a regression alongside the incident-specific covariates.

	Mean	StdDev	2.5%	97.5%
Median age	-0.017	0.005	-0.028	-0.007
Fraction Hispanic	0.063	0.005	0.053	0.073
Fraction white	0.000	0.006	-0.011	0.011
Fraction Black	-0.022	0.005	-0.033	-0.012
Fraction no high school degree	0.051	0.005	0.040	0.060
Fraction college degree	-0.051	0.006	-0.063	-0.041
Fraction poverty	0.011	0.005	-0.001	0.020
Fraction renter	-0.008	0.006	-0.020	0.003
Fraction family	0.029	0.006	0.017	0.039
Log(Median house value)	-0.056	0.007	-0.069	-0.044
Log(Income per capita)	-0.051	0.005	-0.063	-0.041
Log(Density)	0.070	0.005	0.059	0.081

Supplementary Table 27: Census Block Group Socio-economic coefficients in Chicago, estimated together in a regression alongside the incident-level covariates.

	Mean	StdDev	2.5%	97.5%
Median age	-0.009	0.003	-0.015	-0.004
Fraction white	0.017	0.003	0.010	0.023
Fraction college degree	-0.054	0.006	-0.066	-0.042
Fraction renter	-0.006	0.003	-0.011	-0.001
Log(Income per capita)	-0.003	0.006	-0.015	0.009

Supplementary Table 28: Census Block Group Socio-economic coefficients in Chicago, estimated together in a regression alongside the incident-level covariates. Compared to Supplementary Table 27, we further control for log population density, which only affects the direction of association of median age.

	Mean	StdDev	2.5%	97.5%
Median age	0.008	0.006	-0.004	0.019
Fraction white	0.021	0.007	0.007	0.035
Fraction college degree	-0.037	0.011	-0.059	-0.015
Fraction renter	-0.037	0.006	-0.051	-0.025
Log(Income per capita)	-0.056	0.012	-0.081	-0.035
Log(Density)	0.091	0.006	0.078	0.101

4.12 Supplementary Algorithm 1

```
functions {// Using reduce sum for within-chain parallel processing.
// function for calculating the log-likelihood
real partial_sum_lpmf(int[] y_slice,
                      int start, int end,
                      matrix X_total,
                      vector logduration,
                      vector beta_total, real theta_zeroinflation) {
    int Nloc = end - start;
    real localtarget = 0;
    // two cases for calculation reflecting zero-inflation
    for (n in 1:Nloc+1) {
      int ind = start + n - 1;
        if (y_slice[n] == 0) {
          localtarget +=
          log_sum_exp(bernoulli_lpmf(1 | theta_zeroinflation),
              bernoulli_lpmf(0 | theta_zeroinflation)
            + poisson_log_glm_lpmf(y_slice[n] | X_total[ind:ind, :],
                  logduration[ind:ind], beta_total)
                  );
        } else {
          localtarget +=
          bernoulli_lpmf(0 | theta_zeroinflation)
              + poisson_log_glm_lpmf(y_slice[n] | X_total[ind:ind, :],
                ogduration[ind:ind], beta_total
                    );
        }
    }
    return localtarget;
  }
}

data {// Define variables in data
  int<lower=0> N_incidents;// number of observations
  int<lower=0> covariate_matrix_width; // covariate matrix width
  // design matrix for other covariates
  matrix[N_incidents, covariate_matrix_width] X;
  vector<lower=1,upper=1>[N_incidents] ones; // vector of ones
  int<lower=0> N_category; // number of categories
  int<lower=0> N_tract; // number of tracts
  int<lower=0> N_edges; // tract adjacency matrix number of edges
  // node1[i] adjacent to node2[i]
  int<lower=1, upper=N_tract> node1[N_edges];
  // and node1[i] < node2[i]
  int<lower=1, upper=N_tract> node2[N_edges];
  // design matrix for category
```

```
  matrix[N_incidents, N_category] X_category;
  matrix[N_incidents, N_tract] X_tract;  // design matrix for tract
  vector<lower=0>[N_incidents] duration;// alive time for incident
  // count outcome - duplicates for the incident
  int<lower=0> y[N_incidents];
}

transformed data {// Transform for succinctness and better performance
  vector[N_incidents] logduration; // log of duration
  logduration = log(duration);
  // total design matrix
  matrix[N_incidents, 1 + N_tract +
           covariate_matrix_width + N_category] X_total;
  X_total = append_col(append_col(append_col(ones,
              X_tract), X_category), X);
}

parameters {// Define parameters to estimate
  vector[N_tract - 1] beta_tract_raw; // coefficents for tract
  vector[N_category-1] beta_category_raw; // coefficients for category
  // coefficents for other covariates in the model
  vector[covariate_matrix_width] beta;
  real intercept; // intercept term in the model
  // zero inflation parameter
  real<lower=0, upper=1> theta_zeroinflation;
}

transformed parameters  {// Transform parameters for better performance
  vector[N_tract] beta_tract; // coefficents for tract
  vector[N_category] beta_category; // coefficients for category
  // full coefficient vector
  vector[1 + N_tract + covariate_matrix_width + N_category] beta_total;
  // zero centering coefficients
  beta_category = append_row(beta_category_raw,
                    -sum(beta_category_raw));
  // zero centering coefficients
  beta_tract = append_row(beta_tract_raw,
                 -sum(beta_tract_raw));
  // combine coefficient vectors to get full coefficient vector
  beta_total = append_row(append_row(append_row(intercept,
    beta_tract), beta_category), beta);
}

model {// Prior part of Bayesian inference
  beta_tract ~ normal(0, 1);
  beta_category ~ normal(0, 2*inv(sqrt(1 - inv(N_category))));
  beta ~ normal(0, 1);
  intercept ~ normal(0, 5);
```

```
  // Use reduce sum to calculate the target likelihood
  int grainsize = 1;
  target += reduce_sum(
    partial_sum_lpmf, y, grainsize, X_total,
    logduration, beta_total, theta_zeroinflation
    );
  // add adjacency priors to beta_tract
  target += -5 * dot_self(beta_tract[node1] - beta_tract[node2]);
}
```

References

[1] Nil-Jana Akpinar, Maria De-Arteaga, and Alexandra Chouldechova. The effect of differential victim crime reporting on predictive policing systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 838–849, 2021.

[2] Daren C Brabham. *Crowdsourcing in the public sector*. Georgetown University Press, 2015.

[3] Ryan W Buell, Ethan Porter, and Michael I Norton. Surfacing the submerged state: Operational transparency increases trust in and engagement with government. *Manufacturing & Service Operations Management*, 23(4):781–802, 2021.

[4] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

[5] Sara Cavallo, Joann Lynch, and Peter Scull. The digital divide in citizen-initiated government contacts: A gis approach. *Journal of Urban Technology*, 21(4):77–93, 2014.

[6] Benjamin Y Clark, Jeffrey L Brudney, and Sung-Gheel Jang. Coproduction of government services and the new information technology: Investigating the distributional biases. *Public Administration Review*, 73(5):687–701, 2013.

[7] Benjamin Y Clark, Jeffrey L Brudney, Sung-Gheel Jang, and Bradford Davy. Do advanced information technologies produce equitable government responses in coproduction: An examination of 311 systems in 15 us cities. *The American review of public administration*, 50(3): 315–327, 2020.

[8] James R Elliott and Jeremy Pais. Race, class, and hurricane katrina: Social differences in human responses to disaster. *Social science research*, 35(2):295–321, 2006.

[9] Nikhil Garg and Zhi Liu. nikhgarg/spatial_underreporting_crowdsourcing: Accepted, November 2023. URL https://doi.org/10.5281/zenodo.10086832.

[10] Kathryn P Hacker, Andrew J Greenlee, Alison L Hill, Daniel Schneider, and Michael Z Levy. Spatiotemporal trends in bed bug metrics: New york city. *PloS one*, 17(5):e0268798, 2022.

[11] Mark O Hill. Local frequency as a key to interpreting species occurrence data when recording effort is not known. *Methods in Ecology and Evolution*, 3(1):195–205, 2012.

[12] Bryan D Jones, Saadia R Greenberg, Clifford Kaufman, and Joseph Drew. Bureaucratic response to citizen-initiated contacts: Environmental enforcement in detroit. *American Political Science Review*, 71(1):148–165, 1977.

[13] M Kéry and J Andrew Royle. Hierarchical bayes estimation of species richness and occupancy in spatially replicated surveys. *Journal of Applied Ecology*, 45(2):589–598, 2008.

[14] Konstantin Klemmer, Daniel B Neill, and Stephen A Jarvis. Understanding spatial patterns in rape reporting delays. *Royal Society open science*, 8(2):201795, 2021.

[15] Constantine Kontokosta, Boyeong Hong, and Kristi Korsberg. Equity in 311 reporting: Understanding socio-spatial differentials in the propensity to complain. *arXiv preprint arXiv:1710.02452*, 2017.

[16] Constantine E Kontokosta and Boyeong Hong. Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions. *Sustainable Cities and Society*, 64:102503, 2021.

[17] Gabrielle Kruks-Wisner. Seeking the local state: gender, caste, and the pursuit of public services in post-tsunami india. *World Development*, 39(7):1143–1154, 2011.

[18] Diane Lambert. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14, 1992.

[19] Myeong Lee, Jieshu Wang, Shawn Janzen, Susan Winter, and John Harlow. Crowdsourcing behavior in reporting civic issues: The case of boston's 311 systems. In *Academy of Management Proceedings*. Academy of Management Briarcliff Manor, NY 10510, 2021.

[20] Zhi Liu and Nikhil Garg. *Quantifying Spatial Under-reporting Disparities in Resident Crowdsourcing*. Code Ocean, 2023. doi: 10.24433/CO.0984693.V1. URL https://codeocean.com/capsule/2697258/tree/v1.

[21] Zhi Liu and Nikhil Garg. Reporting rate estimation method, November 2023. URL https://doi.org/10.5281/zenodo.10086346.

[22] Kristian Lum and William Isaac. To predict and serve? *Significance*, 13(5):14–19, 2016.

[23] Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. Ipums national historical geographic information system: Version 17.0 [dataset]. minneapolis, mn: Ipums, 2022.

[24] Sara McLafferty, Daniel Schneider, and Kathryn Abelt. Placing volunteered geographic health information: Socio-spatial bias in 311 bed bug report data for new york city. *Health & Place*, 62:102282, 2020.

[25] Scott L Minkoff. Nyc 311: A tract-level analysis of citizen–government contacting in new york city. *Urban Affairs Review*, 52(2):211–246, 2016.

[26] Mitzi Morris, Katherine Wheeler-Martin, Dan Simpson, Stephen J Mooney, Andrew Gelman, and Charles DiMaggio. Bayesian hierarchical spatial models: Implementing the besag york mollié model in stan. *Spatial and spatio-temporal epidemiology*, 31:100301, 2019.

[27] Daniel T. O'Brien. *The Urban Commons: How Data and Technology Can Rebuild Our Communities*. Harvard University Press, December 2018. ISBN 978-0-674-97529-3. Google-Books-ID: JttwDwAAQBAJ.

[28] Daniel Tumminelli O'Brien, Robert J. Sampson, and Christopher Winship. Ecometrics in the Age of Big Data: Measuring and Assessing "Broken Windows" Using Large-scale Administrative Records. *Sociological Methodology*, 45(1):101–147, August 2015. ISSN 0081-1750. doi: 10.1177/0081175015576601. URL https://doi.org/10.1177/0081175015576601. Publisher: SAGE Publications Inc.

[29] Daniel Tumminelli O'Brien, Dietmar Offenhuber, Jessica Baldwin-Philippi, Melissa Sands, and Eric Gordon. Uncharted Territoriality in Coproduction: The Motivations for 311 Reporting. *Journal of Public Administration Research and Theory*, 27(2):320–335, April 2017. ISSN 1053-1858. doi: 10.1093/jopart/muw046. URL https://doi.org/10.1093/jopart/muw046.

[30] Burak Pak, Alvin Chua, and Andrew Vande Moere. Fixmystreet brussels: socio-demographic inequality in crowdsourced civic participation. *Journal of Urban Technology*, 24(2):65–87, 2017.

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[32] Ian W Renner and David I Warton. Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, 69(1):274–281, 2013.

[33] Sidney I Resnick. *Adventures in stochastic processes*. Springer Science & Business Media, 1992.

[34] Richard W Schwester, Tony Carrizales, and Marc Holzer. An examination of the municipal 311 system. *International Journal of Organization Theory & Behavior*, 2009.

[35] Peter Thijssen and Wouter Van Dooren. Who you are/where you live: do neighbourhood characteristics explain co-production? *International Review of Administrative Sciences*, 82(1): 88–109, 2016.

[36] John Clayton Thomas. Citizen-initiated contacts with government agencies: A test of three theories. *American Journal of Political Science*, pages 504–522, 1982.

[37] Corey Kewei Xu and Tian Tang. Closing the gap or widening the divide: The impacts of technology-enabled coproduction on equity in public service delivery. *Public Administration Review*, 80(6):962–975, 2020.

[38] Qianli Yuan. Co-production of public service and information technology: A literature review. *Proceedings of the 20th Annual International Conference on Digital Government Research*, 2019.