BlueConduit

PROJECT SUMMARY REPORT FOR

Waterford Township (MI0006910)

Statistical Analysis Report for Absence of Lead

September 23, 2024

Report Prepared by:
BlueConduit
Ryan Treves, Data Scientist

Table of Contents

Executive Summary

This report presents the findings of an analysis of data/records and an inspection of water service line materials within the water distribution system, of the Charter Township of Waterford (Waterford), Michigan, conducted to assess compliance with the Lead and Copper Rule (LCR) and the recent LCR Revisions (LCRR). The objective was to determine the prevalence of lead service lines in the water distribution system. For the purposes of this report, the term "unlabeled" is used for service lines which do not have a material label.

Key Findings

- **Lead-Free Service Lines:** During the inspection of service lines for 388 structures, it was revealed that none of the water service lines were constructed with lead. This includes 388 inspections of the public side of the service line and 388 inspections of the private side. Staff interviews suggest a uniform absence of lead in Waterford's system, with the exception of a single anomalous service line (described in Step 2 on page 7). Together, this evidence strongly supports the claim that there are no additional lead service lines within Waterford's system.

- **Statistical Evidence:** Statistical analysis, detailed in subsequent sections of this report, provide evidence supporting the conclusion that the absence of lead in the set of inspected service lines is representative of the broader system. The statistical methodology employed ensures a high level of confidence in the accuracy and reliability of the findings. On an address-level basis, **we can conclude with 95% confidence that the 18,287 unlabeled service lines on the public side have no more than a 0.77% probability of being lead, and on the private side the 18,912 unlabeled service lines have no more than a 0.77% probability of being lead**.

- **Recommendations:** Based on the conclusive results, we recommend that Waterford indicate that there are **no lead service lines** (with the exception of one anomalous line) in their inventory. This conclusion is supported by the absence of lead in the inspected sample of service lines and is further validated through statistical evidence. The conclusion should be open to updates as new information arrives.

Conclusion

- This report affirms that Waterford Township has successfully met the requirements of the Lead and Copper Rule, as evidenced by the absence of lead service lines in the inspected sample and the statistical analysis supporting this conclusion. The findings provide a strong foundation for the utility to confidently assert the absence of lead in its service line inventory. Thus, the enclosed evidence supports Waterford's use of a category label of "Non-Lead" for the remaining unlabeled service lines in their inventory, using statistical analysis as the basis of material classification.

Introduction

Waterford, in partnership with BlueConduit, developed this report to submit to the Michigan Department of Environment, Great Lakes, and Energy (EGLE), outlining our Statistical Analysis Approach for the inventory development and specifically our no-lead outcomes.

BlueConduit's methodology and approach is closely aligned with the guiding Principles of Data Science for Lead Service Line Inventories and Replacement Programs and in accordance with the EPA Guidance for Developing and Maintaining a Service Line Inventory, where our work is cited several times.

To ensure that the statistical methods' results are accurate and interpreted appropriately, BlueConduit recommends that water suppliers adhere to the following fundamental statistical and research principles when using statistical methods to inform the classification of the service line material at each property in the Waterford system:

1. Develop a data management plan.

2. Understand existing verified service lines materials, historical records, and all other information about service lines and addresses.

3. Verify service line material for a representative set of service lines, which can be focused on some types of service lines or addresses based on risk factors. Evaluate the correspondence between historical records and verified materials.

If lead is discovered, then:
4a. Develop the predictive model, providing the estimated likelihood of lead service line by service line, and demonstrate that the predictive model performs well for the supply region.

If there is zero lead discovered, then:
4b. Provide valid statistical documentation about how much lead may be in the service area (i.e., estimated likelihood that the entire service area has a very small fraction of lead service lines). Provide an estimated, address-level likelihood of lead probability.

5. Ensure transparency by submitting model and analysis results, explaining how results were utilized to inform the prioritization of properties for service line investigation/replacement and/or to develop the inventory and classify material types.

Before we go any further, we want to be clear about what we mean by "lead" and "no lead." We consider lead to be present if any of these is present:

- utility-owned or customer-owned lateral service line is lead, or
- utility-owned or customer-owned lateral service line is galvanized requiring replacement (according to EPA).

When we say "no lead" as it relates to LCRR compliance, we mean that none of the above is present.

During the course of our work, it became statistically clear that Waterford Township had an extremely low chance of finding lead in its service area.

Out of 388 field verified lines on the public side, and 388 field verified lines on the private side, no lead service lines (LSLs) were found. These known materials came from a set of randomly selected verified lines. Waterford Township maintains approximately 25,027 lines across its service area.

We go into detail about our methodology for non-lead determination in Step 4 on page 13.

Regulations on Statistical Methods for LCRR

In its March 2021 guidance, EGLE makes clear that utilities "are not expected to physically verify every service line, but rather a statistically sound subset. To effectively evaluate the accuracy of service line records and/or predict service line materials, a representative, uniformly random number of service lines must be physically verified." This information can be used by predictive tools to estimate the likelihood of a given line being lead or not.

EGLE also provides guidance on the minimum verification requirements. The predictive tools are focused on unlabeled service lines so lines that are "labeled" – whether LSLs or not LSLs – are excluded. The only types of unlabeled service lines that may be excluded are those four or more inches in diameter.

Utilities must physically verify a minimum number of unlabeled service lines based on a statistical calculation. The minimum number ranges from 20% for those utilities with fewer than 1,500 unlabeled service lines with smaller percentages as the number of unlabeled service lines increases to a maximum of 386. The physical verification is to include 18 inches of the public side at the curb stop, 18 inches of the private side at the curb stop, and the portion of the service line coming into the structure.

To be considered "known," and, therefore, excluded from the pool of sites from which a random sample will be drawn, a service line must either be:

Physically verified after 1988 based on specific criteria and the records are demonstrated to be reliable; or ordinances or other controls were in place at the time the service line was installed specifying the materials that must be used in service line construction. In addition, the utilities must not have observed deviations from those conditions during operations and maintenance.

The results of the physical verification of the random selection of unlabeled service lines are to be compared to existing records to determine if the records are reliable and may prompt additional investigation. The results are used to develop a system-wide inventory that forms the basis of the CDSMI and can be used to notify residents whether they have a LSL.

Advantages of Using Statistical Methods for Inventory

With the number of service lines of unknown material in communities across the United States, water utilities do not have the financial, human, or technical resources to physically inspect each service line segment. The use of statistical analysis and predictive modeling built on the fundamental best practices in statistics allows water utilities to be more precise in their compliance efforts, efficient in resource allocation, and transparent with their customers. This approach also offers transparency into decision-making.

1. **Identifying Unlabeled Service Lines.** Statistical methods can “fill in the blanks” for service lines (SL) with unlabeled materials in SL inventories.

2. **Prioritize areas of High Concentrations of Lead.** Statistical methods have shown to quickly, justifiably, and equitably locate lead service lines in order to prioritize their removal. Prioritizing LSL replacements means that we need to know both where the lead is AND where the lead is not.

3. **Proactive Decision Making.** Some state specific regulatory agencies recent guidance on the use of statistical methods for service line inventory recognizes that physically verifying every LSL is expensive and time consuming. Additionally, time and money used to physically verify service lines that are very likely made of non-lead materials could be used to replace lead service lines or meet other pressing utility needs.

4. **Public Communication**. LCRR requires water systems to make their service line inventories publicly available and to annually notify residents with lead service lines or service lines of unknown material. The use of statistical methods for inventories allows water systems to confidently and transparently communicate about the relative risk to their customers.

5. **Adaptable to new information.** Water utilities can adapt and update their service line inventory with new information from field work or previously unavailable sources. As this information is integrated into statistical analysis, it can update the information for each service line to reflect how the new information changes the understanding for the entire system.

6. **Transparency.** Making available information about the performance of the statistical methods allows for increased transparency into the decision-making and can also shed light into ways to improve the results.

BlueConduit Methodology and Approach

Step 1: Data Collection & Preliminary Inventory Development

The very first step to building an accurate and comprehensive inventory starts with data collection, records and system review. The first guiding principle in running any statistical analysis is to ensure that data is organized and consistent. This means ensuring all of the information collected related to a point of service is associated with that point of service (i.e., a structure’s specific water SL).

BlueConduit begins the process by analyzing data that can be classified as service lines of "known" materials. This involves reviewing verified service line material records, building codes, community ordinances about banned service line materials, and investigating other sources that provide certainty about pipe materials in the system. This initial step provides a baseline for the inventory and helps set the strategy for reducing "unlabeled service lines” for locations where pipe material is not known with high degrees of certainty.

BlueConduit requested information from Waterford to begin the statistical analysis process. Waterford provided the following additional data sources to support the statistical analysis:

- Recently Verified Service Line Material Records (both Public and Private-side)
- Historical Water Service Line Material Records (both Public and Private-side)

- Taxable Parcel Records (year built, land size, value, zoning, etc.)

Summary of Waterford Inventory Work to Date

Waterford Township has gone to great lengths to collect and organize their system data. This dataset was provided to BlueConduit to begin the statistical analysis process.

The Waterford Township Department of Public Works (DPW) received bids for its Complete Distributions System Materials Inventory (CDSMI) on April 14, 2022. The project included locating and taking inventory of water distribution service lines at various locations throughout the Township. The low bid contractor, Underground Infrastructure Services, LLC, (UIS) hydro-excavated water services on the public and private side of the curb stop/box at 388 randomly generated locations. The DPW used Michigan EGLE random location selection tool as described in: https://www.michigan.gov/documents/egle/egle-dwehd-min-service-line-material-verification-req_720143_7.pdf to select water services to investigate. UIS exposed water services 18-inches from the curb stop on both the private side and the public side of the curb stop. Service material was identified, documented, and photographed. Data was input into an ArcGIS Online Dashboard that was accessible to contractor and DPW staff members throughout the project. DPW staff also investigated the portion on the service line entering the structure, photographed the line, and cataloged the material.

Step 2: Data Evaluation and Validation

There are many potential sources of data about SL information, and the types and accuracy of SL data will vary between water systems. Existing data about SL materials comes from different sources (e.g., water main repairs, water meter replacement programs, construction records) and the accuracy and reliability of these records varies by record type and location. Replacements may have been made over time without proper record keeping or records simply may be incomplete or incorrect. It is therefore crucial to establish how correct a water system's historical records are. **Out of caution, BlueConduit does not treat historical data as 100% truth when performing statistical analysis.**

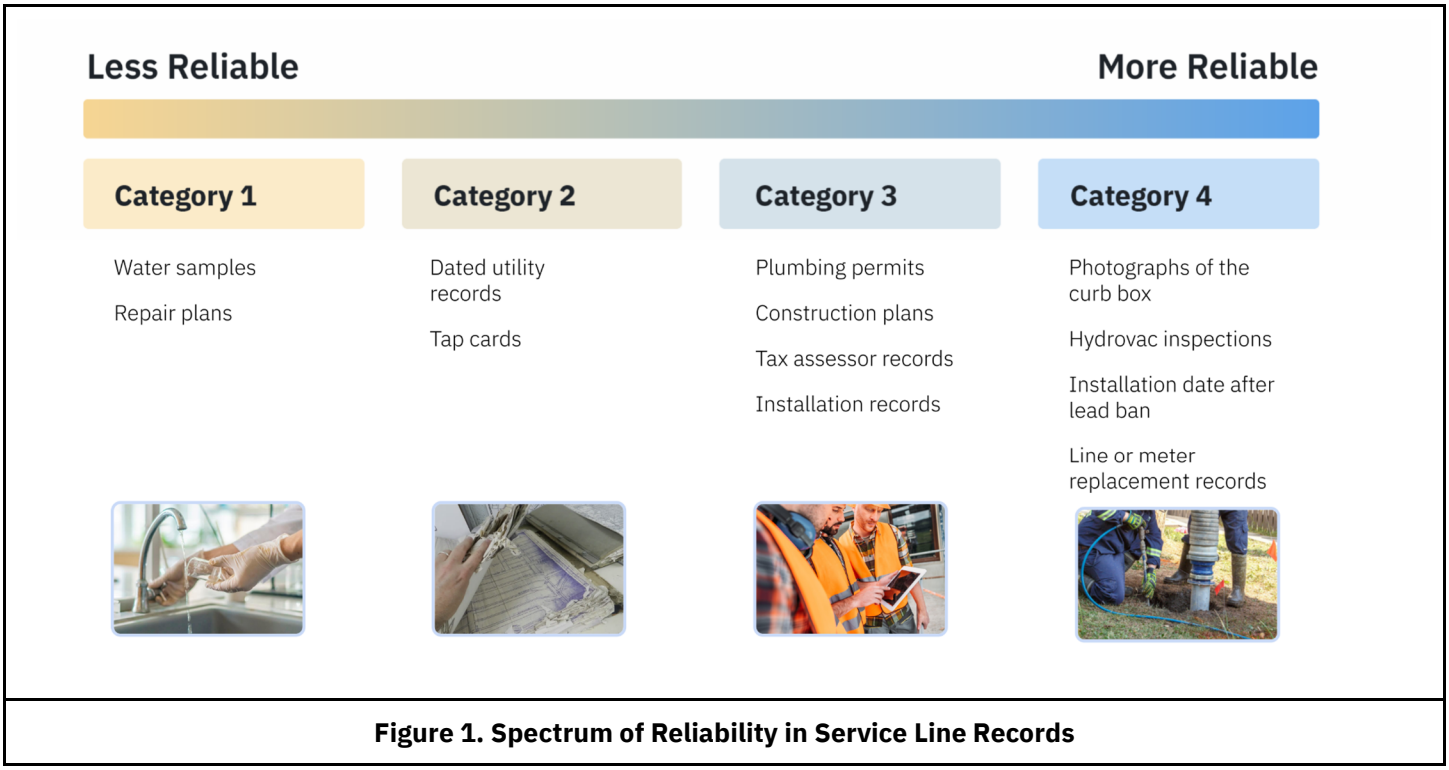


Figure 1. Spectrum of Reliability in Service Line Records

Many water systems do not know which types of records are correct and which are not. For BlueConduit, it is important to establish an understanding of how accurate those records are, while also noting that some types of records are going to be more accurate than others (Figure 1). The process of learning just how accurate (or otherwise) a system's records are, is a powerfully informative piece of this data-driven approach.

As a best-practice, to be considered a reliable-known material in the preliminary analysis and therefore excluded from further field investigation, a service line should meet criteria 1 or 2 below, supported by data from a reliable source:

1. The service line was recently physically verified **-OR-**
2. Ordinances or controls were in place and all of the following apply:
    - Ordinances or other controls were in place at the time the service line was installed specifying materials used in service line construction **-AND-**
    - The water supply has not observed deviations from these ordinance(s) or control(s) during operations and maintenance.

Any service line that does not meet one of the two criteria above may have been included in the list of locations from which a set of sites were uniformly randomly selected for verification in Step 3.

In addition to field verifications, Waterford used other methods to identify non-lead service lines in their system. These designations tend to come from categories 1-3 in Figure 1. This includes service lines where the associated structure was built after the lead ban, an install record is after the lead ban, or a record of the diameter of the line excludes it from the unknown categories.

Overall, Waterford has designated 6,363 public side service lines and 5,738 private side service lines as non-lead from categories 1-3. On the public side, 6,181 have an unknown material but were designated as "Non-Lead - Other" based

on the aforementioned categories, and in total these materials consist of 96.44% Non-Lead - Other (6,184), 3.15% Non-Lead - Plastic (202), and 0.41% Non-Lead - Copper (26). On the private side, 100% of the 5,738 service line materials are designated Non-Lead - Other based on the year that the associated structure was built.

BlueConduit does not use these locations as verified material to base the risk assessment of lead on, as in Step 3, but they do reduce the unlabeled population. The unlabeled population of service lines is what we aim to estimate the risk of finding lead in.

BlueConduit worked with Waterford to understand how their preliminary inventory was developed, and what data and records were collected and utilized to classify known materials. In collaboration with Waterford, we evaluated the accuracy and reliability of those records.

Outlier Service Line with Lead Material

Waterford's records outside of field verifications indicate that one (1) service line contains lead on the private side. Multiple staff members with significant experience report that this service line is an anomaly, not reflective of any broader pattern of lead service lines in the system. Waterford is actively working to gain access to this service line and replace it with non-lead material. Access to this residential property has been denied to the DPW by the homeowner since 2018. The property has recently sold, and efforts to replace the private side of the service are continuing. Nevertheless, the presence of this single lead service line does not undermine the statistical finding (see Step 4: Statistical Analysis page 13).

Step 3: Representative Field Investigation

The goal is to best estimate how much lead there may be in the water system and quantify the likelihood that there is very little lead. The accepted best practice in statistics to be able to make these kinds of estimates is gathering verified service line material data at a random set of structures where the service line material is unknown. Statistically, only such a representative set of verified service points will truly reflect the whole system. This representative randomized sample is critical for understanding the entire system's service line materials.

Even when the water system expects to find zero lead service lines out of their currently unlabeled service line materials, the best approach is to inspect the materials of service lines that are representative of all of the remaining unlabeled service lines across the whole system.

In order to determine the appropriate number of field verifications necessary, the following information was considered: (1) total number of service lines in the system, (2) total number of unlabeled service lines in the system, and (3) desired confidence level (e.g., 95%).

Service line selection is designed to be representative to mitigate data biases. By doing this, we ensure that every service line that has not yet been verified has an equal chance of being verified. That means the data will represent all of the service lines with unlabeled materials across all neighborhoods, socioeconomic factors, age of structure, and any historical records about service lines. This is critical to ensure that the water system does not make conclusions about all locations based on only the existing data, as that may not have previously been equally representative of all parts of the system.

Field Verification Method

While BlueConduit does not require the water system to use a specific verification method in the field, we do recommend potholing/hydrovac at the curb box as the most accurate and efficient method to verify material on the public and private sides of the service line. We recognize that every system's infrastructure is different and recommend utilizing the verification method(s) most efficient and effective for Waterford, so long as the chosen methods are aligned with any specific field verification requirements set forth by state regulation. Waterford chose to hydrovac the curb box/stops. This verification method was the most accurate available while being minimally invasive to the customer.

Description of Representative Field Inspections

The 388 public side field inspections consist of 81.19% Non-Lead - Plastic (315), 17.78% Non-Lead - Copper (69), and 1.03% Non-Lead - Other (4). The 388 private side field inspections consist of 74.23% Non-Lead - Plastic (288), 24.48% Non-Lead - Copper (95), 1.03% Non-Lead - Other (4), and 0.26% Galvanized (1). The locations of verified service lines are numerous and cover the entire service area (Figure 2).

Ultimately, a total of 388 public service lines and 388 private service lines were verified. After compiling Waterford Township's inspections, we could see that after sampling widely and densely, the specific non-lead materials on both the public (Figure 2) and private (Figure 3) side were spatially distributed.

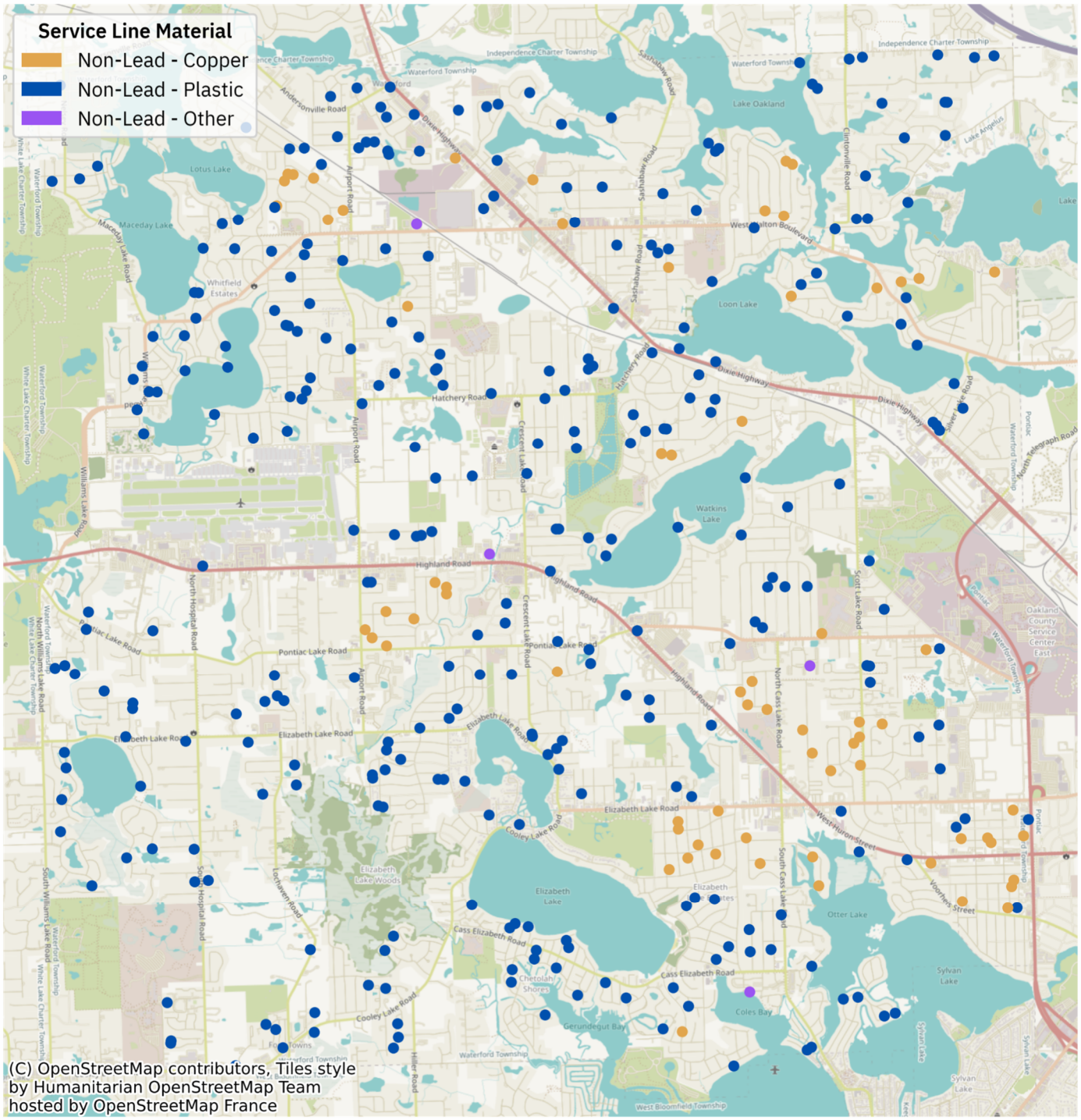


Figure 2. Locations of Verified Public Service Line Material.

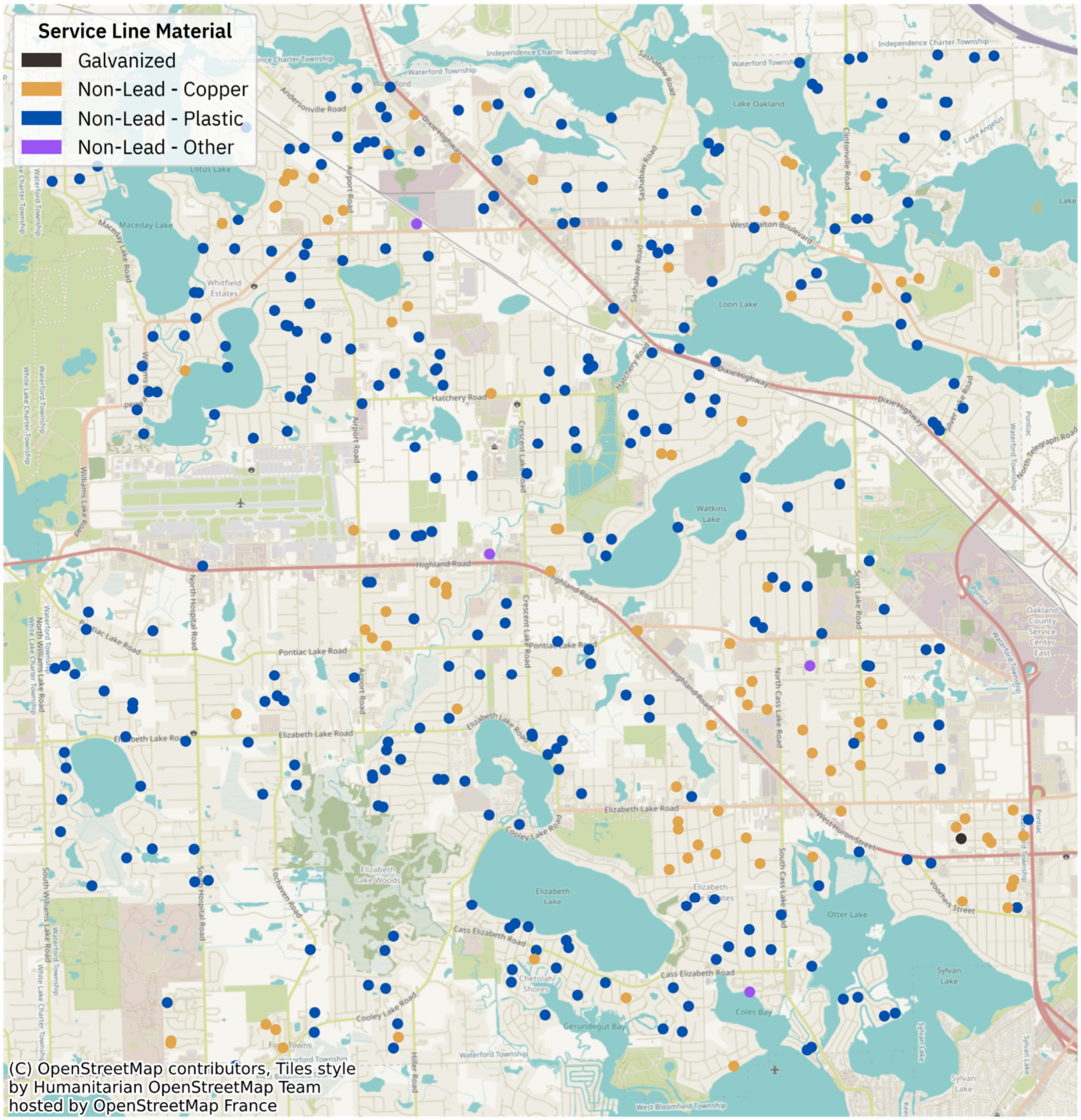


Figure 3. Locations of Verified Private Service Line Material.

The distribution of materials found in all 388 field verifications is described in Table 1 on the following page.

Table 1: Material counts found through field investigations.

Service Line Material	Public Side Count	Private Side Count
Galvanized	0	1
Non-Lead - Plastic	315	288
Non-Lead - Copper	69	95
Non-Lead - Other	4	4
Total	*388*	*388*

In addition to being numerous and spatially representative of the water system, the verifications are representative of the underlying population. For instance, all levels of structure sale price have been sampled with these verifications (Appendix C).

In summary, Waterford Township's numerous inspections amount to a strong representation of the underlying population. The inspections have been geographically distributed and have sampled across the population's spectrum of types of structures. These verifications will be the basis of the statistical analysis in the following step.

Step 4: Statistical Analysis

Method for Showing Evidence of the Absence of Lead

Many water systems report having never encountered a lead pipe in all of their time maintaining their water infrastructure and have no historical records of lead line installations.

However, the absence of evidence of lead at a subset of service lines alone cannot "prove" an absence of lead altogether across the whole service area — but it's an important part of the picture. If the preliminary inventory data **and** representative field investigations reveal *zero* lead - (no lead service lines, no galvanized service lines requiring replacement, or no lead goosenecks) - the remaining unlabeled service lines will have a very low probability of being lead.

The only way to absolutely prove with complete certainty there are no lead service lines in a water system is to dig up and visually inspect every single length of every single service line. Water systems and regulators alike recognize that 100% physical verification is not feasible from a time or cost perspective. This is where BlueConduit's statistical analysis approach provides significant value and is the most efficient, scientific alternative to putting eyes on every single pipe. This approach has also proven to be more effective than using historical records alone.

We can rely on a body of research in statistics spanning decades to utilize the most productive way to think about situations where information is scarcer than we would like for typical analysis methods. For instance, when patients in a

particular clinical trial of a new drug have not experienced a given side effect, medical researchers do not conclude that the side effect is impossible, but that they quantify how unlikely it is. Our approach does the same.

If we have representative data that shows zero lead, our data scientists take two inputs:

1. the number of service lines in the representative sample that you inspected where you found no lead **and**
2. the desired level of confidence - and create the highest possible percentage of lead services lines you could reasonably expect to find in the water system.

To illustrate this concept, we present a hypothetical utility with 30,000 service lines, 10,000 of which are verified, and they haven't seen any lead in their system. They want to know if they can claim there aren't lead service lines or goosenecks in the 20,000 unlabeled service lines. The utility verifies randomly selected 200 inspections (that are representative of the service area) and finds no lead, so they can conclude they are 95% certain that fewer than 1.5% of their unlabeled service lines are lead. They can also conclude that they are 99% certain that less than 2.3% of those unlabeled service lines are lead.

And as they continue to gain more information, say, another 50 representative service lines are found to be not lead, then they can update those numbers to be 95% certain that fewer than 240 (1.2% of unlabeled service lines) are lead. As the number of representative service lines found to not have any lead increases, then these metrics evolve.

When a representative set of unlabeled service lines are inspected (on both portions of the service line) and when not even one of them turns out to be lead, then we can characterize our uncertainty about that water system's remaining unlabeled service lines with the above variable calculations. The resulting calculations can guide community and utility action even when it is very likely that there is a very low number of lead service lines in the system.

Results and Conclusion Showing Evidence of the Absence of Lead

Based on representative field investigations revealing 0 indications of lead, as well as historical records containing only a single lead service line, we can conclude with 95% confidence that of the system's 25,027 service lines, there are fewer than 142 lead service lines that may be present on the public side of the system and fewer than 147 lead service lines that may be present on the private side of the system. These figures represent 0.57% of all public service lines and 0.59% of private service lines in the whole system. Furthermore, even including the single anomalous lead service line in the statistical analysis, we can apply the appropriate statistical method (see Clopper-Pearson confidence interval) and still conclude with 95% confidence that the probability of any single unknown service line being lead is less than 0.014 (1.4%).

Thus, we provide evidence that could support Waterford's use of a category label of "Non-Lead" for the remaining unlabeled service lines in their inventory, using statistical analysis as the basis of material classification, recognizing the following:

BlueConduit recommends that Waterford recognize that additional physical verifications may be requested after review of the report and findings, including possible excavation of additional service lines.

Even when there is a low likelihood of a property having lead (e.g., 99% chance that service line is non-lead), there remains a small chance that the property will have lead (e.g., 1% chance of lead).

BlueConduit guides Waterford to expect the lead service line inventory to remain a living and evolving dataset as future field inspections and potential replacements are completed. In accordance with LCRR, if after a service line classified as non-lead based on the results of statistical analysis is eventually found to be lead in the field, the classification of the material should be adjusted accordingly in the inventory and proper notification and remediation steps taken.

Appendix

A: Michigan EGLE Guidance: Minimum Number of Service Lines Requiring Physical Verification

B: References

C: Comparison of Sale Price for Samples and Total Population

Appendix A

Source Document: Michigan EGLE - MINIMUM SERVICE LINE MATERIAL VERIFICATION REQUIREMENTS
Minimum Number of Service Lines Requiring Physical Verification (p.6)

Number of "Unknown" Service Lines*	Number to Physically Verify
Fewer than 1,500	20% of "unknown" lines
1,500	306
1,600	310
1,700	314
1,800	317
1,900	320
2,000	322
2,200	327
2,400	331
2,600	335
2,800	338
3,000	341
3,500	346
4,000	351
4,500	354
5,000	357
6,000	361
7,000	364
8,000	367
9,000	368
10,000	370
15,000	375
20,000	377
30,000	379
40,000	381
60,000	382
90,000	383
225,000 or more	384

*For the purposes of the physical verification process, this represents the number of service lines that do NOT meet the criteria for "known" service lines described in Step 2 of this document. If the number of "unknowns" falls between two values on the chart, BlueConduit will either interpolate or round up to the higher number when creating the final recommended inspection list.

Appendix B

State-Guidance References

The method outlined in this plan is grounded in the academic literature for how to use data science to inventory and locate lead service lines. It draws upon recognized best practices from regulators and policy makers on how to adapt that research for the purpose of service line inventory requirements required by LCRR.

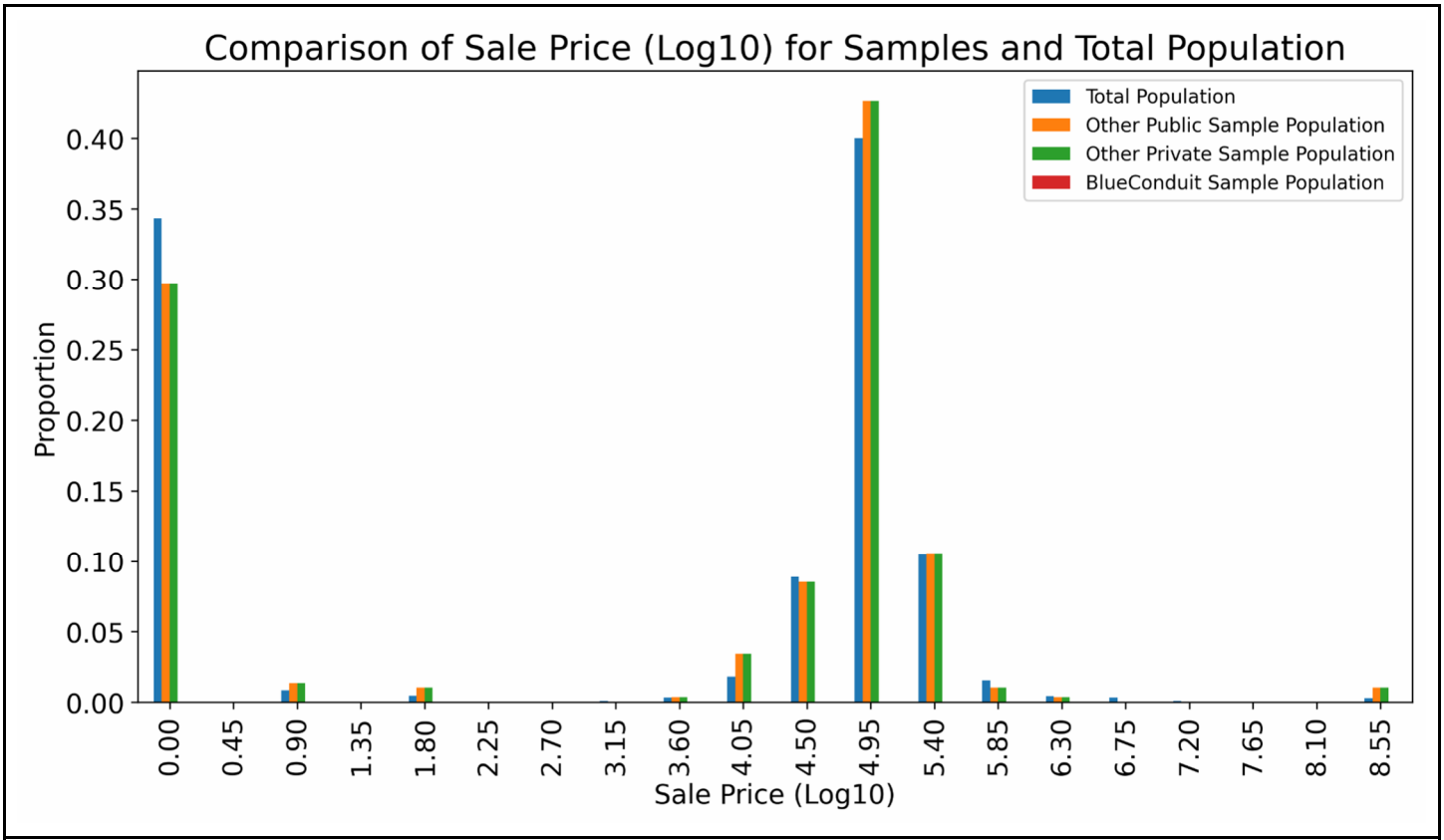
- ActiveRemediation: The Search for Lead Pipes in Flint, Michigan
- Getting the Lead Out: Data Science and Water Service Lines in Flint
- Principles of Data Science for Service Line Inventory and Replacement Programs, Association of State Drinking Water Administrators
- Michigan EGLE guidance on the use of statistical methods for developing service line inventory
    - Complete Distribution System Materials Inventory (CDSMI) Overview
    - Minimum Service Line Materials Verification Requirements
    - Complete Distribution System Materials Inventory (CDSMI), Evaluating Minimum Service Line Materials Verification Data
- Oregon Health Authority
    - Statistical Guidance for Evaluating Unknown Service Lines
- Virginia Department of Health, Office of Drinking Water
    - Guidance for Statistical Methods and Predictive Modeling
- Maryland Department of the Environment
    - MDE LCRR Service Line Inventory Guidance

BlueConduit Blog Article

When a Representative Sample of Service Lines Reveals Zero Lead, What Can You Say?
When a Representative Sample of Service Lines Reveals Zero Lead, What Can You Say? | BlueConduit

By Alice Berners-Lee, Ph.D., Jared Webb and Eric Schwartz, Ph.D.

Appendix C



Comparison of Sale Price (Log10) for Samples and Total Population. Proportion of service lines associated with structures with certain sale prices (the common log of the sale value), as a function of whether and how it was inspected.