Service Line Inventory Statistical Methods and Predictive Modeling Guidance for North Carolina

In accordance with the Lead and Copper Rule Revisions, all community and non-transient, non-community water systems are required to complete an initial inventory of the material of all service lines in their distribution system, regardless of ownership status and intended use. Initial inventories are due October 16, 2024.

A statistical method uses known locations to estimate the probability of a service line being lead and to prioritize areas of investigation and replacement.[1] Predictive modeling, or machine learning, is a specific type of statistical method that develops rules or algorithms for assessing lead likelihood by looking for patterns in a dataset.

In this guidance, the North Carolina Public Water Supply (PWS) Section will outline the basic requirements for statistical methods and predictive models used for lead service line inventories in North Carolina. Water systems will be required to submit a Statistical Method or Predictive Model Proposal to the PWS Section and gain approval prior to using the model. All proposals must be emailed to pwss.lcr@deq.nc.gov.

1.0 Historical records review to identify all unknown service lines

A historical records review must be completed prior to the use of other service line identification methods, including statistical methods. Historical records include construction and plumbing codes, water system records, inspection records, or other records with relevant information. Identify all service lines (water system owned and customer owned) of unknown material to determine the number of unknown service lines. An unknown service line is a service line where the material is not known to be lead, galvanized requiring replacement, or non-lead, such as where there is no documented evidence supporting material classification (40 CFR §141.2).[2]

Any service line installed after March 1987, the lead ban date for North Carolina, can be classified as non-lead and any service line with a diameter greater than 2 inches can be classified as non-lead.

2.0 Identify how many service lines must be physically verified

The number of random verifications required to predict the material of the unknown service line depends on the number of unknown service lines.

- Water systems with fewer than 1,500 unknown service lines (water system or customer owned) must physically verify at least 20 percent of the total number of unknown service lines.
- Water systems with 1,500 or more unknown service lines must physically verify enough service lines to reach a 95 percent confidence level. See *Minimum Number of Unknown Service Lines Requiring Physical Verification* (Appendix A) to determine the number of service lines requiring verification.

3.0 Randomly select service lines for physical verification

From the list of unknown service lines identified in Section 1.0, randomly select enough service lines to at least meet the number requiring physical verification as determined in Section 2.0. The subset must be random and representative of the unknown service lines across the entire distribution system. In other words, each unknown service line connected to the distribution system should have an equal chance of being selected for random verification.

Create a spreadsheet or other tracking tool containing the randomly selected service lines. Include columns or fields for tracking both existing records and results of the physical verification. Enter currently available information about service line materials based on pre-verification records.

Water systems must demonstrate that the set selected is, in fact, representative of the service lines with unknown materials across the entire distribution system. This requires comparing the characteristics of the initial subset to be inspected to the characteristics of the pool of all unknowns from which that subset was selected. This can be done using one or more of the following plots and tables.

(1) A graph or table comparing the percents of each service line coming from each decade of year built for the verification subset versus all unknowns. That is, if 22% of all unknown service lines are from the 1920s, then approximately 22% of service lines in the initial verification subset should be from the 1920s.

(2) A map or table comparing the verification subset versus all unknowns by geographic region. For instance, if 10% of unknown service lines come from Zone A of the city, then approximately 10% of the initial subset's service lines should come from Zone A.

(3) A graph or table comparing the number of service lines with a partial identification by historical records in the verification subset versus all unknowns. For instance, if 30% of unknowns have a historical record that indicates a system-owned material classification, but the customer side does not have any historical record, then about 30% of the initial subset of service lines to be verified should have that same pattern of historical records.

(4) A graph or table showing that the initial verification subset has the same percentage of any given Environmental Justice factor that the full set of unknowns in the system has. Environmental Justice factors can include, but are not limited to, percent of population living below the poverty line, census blocks with at least 30% of households living below the poverty line, percent non-white.

The characteristic percentages of the verification subset versus all unknowns does not have to be an exact match, but the subset should be made to be as representative as possible. You must submit documentation in the form of graphs and/or tables demonstrating that the sites selected are appropriately representative of the unknowns in the distribution system.

4.0 Conduct physical verification

Randomly selected service lines must be verified through at least one point of physical identification. Water systems can use any of the accepted physical identification methods, such as excavation, customer-submitted visual inspection, meter box inspection, or other forms of visual inspection approved by the PWS Section. Water systems should use the verification method that is most accurate and efficient for them. More than one point of physical identification should be used if there is any reason to believe that one point verification will not provide accurate results.[3]

In the spreadsheet or tracking tool created to contain the randomly selected service lines, enter the service line material observed for the utility and customer sides of the service line during the physical verification. Record the actual material observed, such as lead, copper, plastic, galvanized steel, or other.

5.0 Evaluate results of physical verification

The representative, uniformly random sample of service line materials should be, with a high probability, representative of service line materials from the portion(s) of the distribution system from which the random service lines were selected. Use this information to assess the reliability of existing records and make sound assessments/decisions to complete the system-wide inventory.

To assess the reliability of existing records, compare any existing historical records of the randomly selected service lines with the new physically verified materials from those sites. Evaluate what percent of the time the physical verification matched the historical records. If the historical records are found to be unreliable, additional physical verification may be necessary.

6.0 Predict service line materials at other locations

The results of the physical verifications, in combination with other available information, can be used to predict service line materials at other sites.

If a water system can demonstrate that there are no lead service lines and no galvanized requiring replacement service lines in the random sample set, then a statement can be released regarding the probability of all remaining unknown service lines being non-lead. The statement should be formatted as "With 95% confidence we can conclude we have fewer than 1% of service lines containing lead." These service lines can be classified as "Non-Lead" in the initial lead service line inventory.

For water systems that have found lead or galvanized requiring replacement service lines, predictive modeling methods can be used to predict individual service line materials. For the purposes of predictive modeling, a lead gooseneck, pigtail, or connector, or a galvanized requiring replacement service line may count as a lead service line in the model development.

6.1 Classification

A lead likelihood factor can be assigned to unknown service lines based on the results of the method (i.e., likely lead, unlikely lead). The PWS Section may reevaluate the classification criteria for predictive modeling based on future guidance.

6.1.1 Non-Lead Service Line

A service line should be classified as “Non-Lead” if the likelihood of lead produced by the model is less than or equal to 10%.

6.1.2 Unlikely Lead Service Line

A service line should be classified as “Unknown – Unlikely Lead” if the likelihood of lead produced by the model is greater than 10% but less than or equal to 30%. These service lines should be classified after all other unknown service lines have been determined.

6.1.3 Likely Lead Service line

A service line should be classified as “Unknown - Likely Lead” if the likelihood is between 60% and 80%. These service lines should be verified before all other unknown service lines.

6.1.4 Lead Service Line

A service line should be classified as “Lead” if the likelihood of lead is at least 80%.

Any service lines that fall outside of the above ranges shall remain “Unknown – Service Line Material Unknown.”

7.0 Statistical method proposal

7.1 Proposal

In advance of employing statistical methods for the purpose of assigning service line materials for the lead service line inventory, the water system must provide a proposal to the PWS Section containing the following information:

1. The scope of the investigation. For example, which service connections are addressed by the method. If the scope is unusual, such as addressing more than one service area or distribution system, further explanation is necessary.
2. The anticipated number of unknowns (water system owned and customer owned), anticipated number of field verifications (water system owned and customer owned) and intended field verification method(s).
3. Explain how the project approach ensures representative or unbiased data.
4. Explain how the approach treats historical records, including how the team will assess the reliability of historical records.
5. Explain how Environmental Justice populations are represented in the verification dataset.
6. Explain how galvanized requiring replacement service lines and lead connectors will be accounted for.

7. Explain how the process will be iterative. For example, how the inventory will be updated as new information becomes available.
8. Explain the expected findings from this approach and how the team intends to assign service line materials for the unknowns.

If lead or galvanized requiring replacement service lines are known to be present in the distribution system and a predictive model will be used, the water system must provide the following additional information as part of their proposal:

9. Explain how training data and testing data will be used to calibrate and test the model. Explain how many times the model will be re-run and calibrated after the initial run.
10. Explain which types of materials will be assigned (i.e., lead service lines, lead connectors, galvanized requiring replacement, etc.)
11. Results supporting the performance of a classification model, such as a receiver operating characteristic curve (summarized by the area under the curve; showing how well the model can distinguish between the two outcomes, e.g., lead vs not lead, in the test dataset) or a calibration curve (showing how well the model's probabilities match the frequencies with which each outcome occurs in test dataset).

All proposals must be emailed to pwss.lcr@deq.nc.gov. The PWS Section will review the proposal, and if satisfactory, approve it.

8.0 Statistical method report

Prior to submitting the initial service line inventory using a statistical method, including a predictive model, for the purpose of assigning service line materials, the water system must provide a report describing the following information:

1. The scope of the investigation. If a predictive model was used, explain how the model addresses all service connections with unknown materials. If the scope of the method is unusual, for example, addressing more than one service area or distribution system, additional explanation and justification must be provided.
2. The number of unknowns (water system and customer owned), number of field verifications (water system and customer owned), and the field investigation method(s).
3. Explanation of how the project approach ensured representative or unbiased data.
4. Explanation of how the project approach used historical records, including the method used to assess reliability of historical records, and the findings.
5. Explanation of how the project addressed galvanized requiring replacement.
6. Explanation of the method to assign service line materials for unknowns.
7. A full listing of service lines and materials, noting the materials assigned using statistical methods. If a model was used, provide a detailed listing of the model output, including the predicted likelihood of lead for each service line (water system and customer owned) and the resulting material assignments.

If a predictive model was used, provide responses to items 8 – 13 below. If not, proceed to item 14.

8. Explanation of how the project approach ensured representative or unbiased data, including the "training data" and "testing data".
9. Explanation of how the "training data" and the "testing data" were used to calibrate and test the model. Provide a report of the last calibration showing comparison to the "testing data". Provide an assessment of the accuracy of the model.
10. Explanation of how the modeling process incorporated new information.
11. Explanation of the sequence of collecting additional information and running the model after the initial run, including how many times the model was calibrated and run.
12. Explanation of how success was measured.
13. Explanation of how model results were used to develop the inventory, including how the modeling team mapped the service line "ranking" into a material classification for the lead service line inventory reporting.

14. If a predictive model was not used, provide a statement on the likelihood of lead in the set of remaining unknowns that the water system will release.

15. Explanation of the next steps, including additional or ongoing work on the lead service line inventory.

The water system must provide the report to the PWS Section for review prior to final submission of the inventory spreadsheet. All reports must be emailed to pwss.lcr@deq.nc.gov. The PWS Section will review the results of the statistical analysis and retains the authority to reject service line material determinations. The PWS Section will notify the water system of the review results.

If at any point it is determined that a service line was misclassified as non-lead when it was actually lead or galvanized requiring replacement using a statistical method, the water system must re-evaluate their method and send a written notification to the PWS Section with a proposed solution to validate their method.

Appendix A: Minimum Number of Unknown Service Lines Requiring Physical Verification[3]

Number of Unknown Service Lines Remaining	Number of Service Lines Requiring Physical Verification
Fewer than 1,500	20% of Unknown Service 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

Note: If the number of unknown service lines falls between two values, either interpolate or round up to the higher number.

References

[1]Guidance for Developing and Maintaining a Service Line Inventory
[2]40 CFR 141.2
[3]MI EGLE Minimum Service Line Material Verification Requirements