Guidance for Statistical Methods and Predictive Modeling

Virginia Department of Health, Office of Drinking Water

The primary purpose of the Lead Service Line Inventory is to identify the location of lead service lines (LSLs) and galvanized requiring replacement (GRR) service lines so that they can be replaced when the time comes, to eliminate the possibility of lead exposure and to protect public health. The inventory is useful only if it accurately identifies the locations of LSLs and GRRs. A secondary purpose for the inventory is to document the material, size, and age of service lines owned by the waterworks to facilitate asset management and related planning by the waterworks. The inventory can also contain information useful for developing a Lead and Copper Rule tap sampling plan and identifying appropriate tap sample locations.

VDH ODW has a goal of allowing use of statistical methods and predictive modeling to facilitate identification of service line materials, provided it is sufficient to meet the primary purpose of protecting public health.

Statistical Methods Guidance

Refer to the *Minimum Service Line Material Verification Requirements*, revised 3/2021, published by Michigan EGLE.

The terms *verify* and *verification* used here and in the referenced documents refers to a physical or visual inspection of a service line in the field to determine the service line materials. Physical or visual inspections are considered more accurate than records review, and thus have been termed *verification.* This term is used regardless of whether there are records documenting the service line materials. *Verification* can be used to evaluate the truth or accuracy of records if they exist.

How to demonstrate service lines with unknown materials are Non-Lead:

Statistical Methods as described here can be a standalone strategy for demonstrating service lines with unknown materials are Non-Lead. Follow this protocol, including:

- Pick a random set to verify
- Complete verifications
- If no lead and no GRR are found, the unknowns can be classified as non-lead - other.
- If lead is found, additional study is necessary to identify lead service line locations.

Virginia waterworks may use the protocol described on page 3 of *Minimum Service Line Material Verification Requirements* to characterize the unknown service lines in a distribution system, as follows:

1. **Identify all unknown service lines (utility side and private side)**

   Identify all service lines (utility side and private side) of unknown material. Determine the total number of unknown service lines. Service lines larger than two inches in diameter may be excluded from this verification process since they can be classified as non-lead based on size.

The number and location of unknowns on the utility side and the private side may be different. Many utilities will have more unknowns on the private side. Consider the utility side and the private side as two different pools of unknowns.

2. **Identify how many service lines to physically verify**

   - Waterworks with fewer than 1,500 unknown service lines (utility or customer side) must physically verify at least 20 percent of the total number of unknown lines.
   - Waterworks with 1,500 or more unknown service lines must physically verify enough lines to reach a 95 percent confidence level. See *Minimum Service Line Material Verification Requirements* Appendix A to determine the number of service lines requiring verification.

3. **Randomly select service lines for verification**

   Service Line Material Field Verified or verification means the service line material was determined by observing the service line by field methods.

   From the list of unknown service lines identified in Step 1, randomly select enough service lines to at least meet the number requiring physical verification as determined in Step 2. Selection must be uniformly random and not selected based on any specific criteria which can introduce bias. In other words, each unknown service line must have an equal chance of being chosen for verification. See *Minimum Service Line Material Verification Requirements,* Appendix B for an easy way to generate a uniformly random set of locations for verification. Note: It may be tempting to introduce a "logic" to the site selection process, such as selecting within periods of construction or targeting portions of town. However, doing so can unintentionally bias the data set. Be certain to use a truly random selection method.

4. **Track records and materials during verification**

   Create a spreadsheet (or other tracking mechanism) containing the randomly selected locations identified in Step 3. Include columns or fields for tracking both existing records and results of the physical verification. In this step, enter currently available information about service line materials based on pre-verification records. You may use the ODW Community Service Line Inventory Template, available on ODW's website at https://www.vdh.virginia.gov/drinking-water/lcrr-guidance/.

5. **Conduct physical verification**

   Physically verify the materials for the utility and customer portions of the service lines. Use pre-approved methods listed in the *Lead Service Line Inventory Methods* document. Physical verification includes visual inspection of the service line materials at the meter setting, visual inspection inside the home/building, customer self-identification, closed circuit television inspection (external), magnet and scratch test, and excavation – vacuum and mechanical. The waterworks has the responsibility to select and validate the protocol for the use of each physical verification method. For example, ODW is not mandating a specific number or specific locations for test pits for physical verification, so that it is up to the waterworks to determine the number and locations of test pits to accurately determine the materials of a service line.

6. **Record physical verification results**

   In the spreadsheet or tracking tool created in Step 4, enter the service line material observed for the utility and customer sides of service line during the physical verification conducted in Step 5. Record the actual material observed, such as lead, copper, plastic, galvanized steel, or other.

7. **Evaluate Results of physical verification**

   This 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 system from which the random set was selected. You can use this information to assess the reliability of your existing records and make sound assessments/decisions as you complete a system-wide inventory.

   Profile of unknowns: The profile of the random sample set will be representative of the entire set of unknowns. For example, the random sample set will give the percentages of lead, plastic, galvanized, and copper. However, this method does not by itself identify the specific locations of lead and galvanized requiring replacement service lines.

   - Profile of unknowns with no lead service lines and no galvanized requiring replacement: If a waterworks can demonstrate that there are no lead service lines and no galvanized requiring replacement in the random sample set, this means that the set of unknowns consists entirely of non-lead service lines. The waterworks may classify these as "Non-Lead - Other" in their initial LSL inventory.
   - Predict Service Line Materials at Other Locations: The results of this verification, in combination with other available information, can be used to predict service line materials at other sites. Predictive modeling methods can be used to predict individual service line materials.
   - Assess Reliability of Existing Records: Compare the historical records of your randomly selected sites with physically verified materials for those sites to evaluate the reliability of your water supply's historical records. For example, what percent of the time did the physical verification match the historical records? If historical records are found to be unreliable, additional physical verification may be necessary.

<u>Addressing Galvanized Requiring Replacement Service Lines</u>

"Galvanized Requiring Replacement" means where a galvanized service line is or was at any time downstream of a lead service line or is currently downstream of a "Lead Status Unknown" service line. If the water system is unable to demonstrate that the galvanized service line was never downstream of a lead service line, it must presume there was an upstream lead service line. (40 CFR 141.84 a 4 ii)

When a waterworks identifies a galvanized service line, most frequently on the customer side, it must decide how to classify the resulting combination of utility and customer service line materials. Here are some possible scenarios:

System-Owned Portion Service Line Material Observed	If Non-Lead in Column F, Was Material Ever Previously Lead?	Customer-Owned Portion Service Line Material Observed	Customer-Owned Portion Service Line Material Classification	Basis
Galvanized	No	Galvanized	Galvanized	• Same material (galvanized) on System Service Line. • No history and no records of Lead Service Lines.
Copper	No	Galvanized	Galvanized	• Copper replaced an unknown previous material. • No history and no records of Lead Service Lines
Copper	Unknown	Galvanized	Galvanized Requiring Replacement	• Copper replaced an unknown previous material. • History of Lead Service Lines present. • Unknown if prior system SL was lead.
Unknown	Not applicable	Galvanized	Galvanized Requiring Replacement	• Classified by definition as GRR.
Copper	Yes	Galvanized	Galvanized Requiring Replacement	• Previous lead system SL • History of Lead Service Lines present.
Galvanized	Unknown or Yes	Galvanized	Galvanized Requiring Replacement	• Previous lead system SL • History of Lead Service Lines present.
Lead	Yes	Galvanized	Galvanized Requiring Replacement	• Previous lead system SL • History of Lead Service Lines present.

Waterworks that answer "If Non-Lead in Column F, Was Material Ever Previously Lead?" with "No" will need to explain the basis for this assertion. For those that do not have records specific to a service line but assert there is no history of lead service lines, provide the basis for the statement, including the available data to back up the assertion.

The most conservative approach for addressing galvanized service lines is to treat all galvanized service lines as if the system service line was lead in the past and classify the galvanized service lines as galvanized requiring replacement.

Statistical Methods Proposal Required

In advance of employing statistical methods for the purpose of assigning service line materials, the waterworks owner must provide a work plan to VDH ODW describing the following items:

1. The scope of the investigation. For example, the investigation will address all service connections with unknown materials on the utility and/or customer service lines. If the scope of the population is unusual, for example, addressing more than one service area or distribution system, additional explanation and justification is necessary.

2. State the anticipated number of unknowns (utility side and customer side), anticipated number of field verifications (utility side and customer side) and the 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 reliability of historical records.
5. Explain how the approach addresses galvanized requiring replacement service lines.
6. Explain the expected findings from this approach and how the team intends to assign service line materials for the unknowns.

VDH ODW will review the proposal, and if satisfactory, approve it.

Statistical Methods Report
Prior to submitting the initial service line inventory using statistical methods for the purpose of assigning service line materials, the waterworks owner must provide a report describing the following items:

1. The scope of the investigation.
2. The number of unknowns (utility side and customer side), number of field verifications (utility side and customer side) and the field verification method(s).
3. Explain how the project approach ensured representative or unbiased data.
4. Explain how the project approach used historical records, including how the team assessed reliability of historical records, and the findings.
5. Explain how the project addressed galvanized requiring replacement.
6. Explain how the team assigned service line materials for the unknowns.
7. Provide a full listing of the service lines and materials, noting the materials assigned using statistical methods.

Predictive Modeling Guidance

For specific guidance on predictive modeling as applied to service line inventories, see *Guidance for utilizing predictive modeling to identify lead service lines for inventory development.* New Jersey Department of Environmental Protection. Trenton, NJ. 9 pages. This is available at: https://dep.nj.gov/wp-content/uploads/dsr/lsl-predictive-modeling-guidance.pdf.

Predictive Modeling Proposal
In advance of employing predictive modeling for the purpose of assigning service line materials for the service line inventory, the waterworks owner must provide a work plan describing the following items:

1. The scope of the predictive model. For example, the predictive model will address all service connections with unknown materials on the utility and/or customer service lines. If the scope of the model is unusual, for example, addressing more than one service area or distribution system, additional explanation and justification is necessary.
2. State the anticipated number of unknowns (utility side and customer side), anticipated number of field verifications (utility side and customer side) and the intended field verification method(s).

3. Explain how the project approach ensures representative or unbiased data, including the "training data" and "testing data".
4. Explain how "training data" and the "testing data" will be used to calibrate and test the model.
5. Explain how the modeling approach treats historical records, including how the modeling team assesses reliability of historical records.
6. State the records or information supplied by the utility used by the model.
7. State the records or information beyond what was supplied by the utility used by the model.
8. Explain how the modeling process incorporates new information.
9. Explain if the modeling team plans to collect additional information and re-run the model after the initial run, including how many times the model will be calibrated and run.
10. Explain how success is measured.
11. Explain how the modeling team intends to map the service line "ranking" into a material classification for the LSL Inventory.
12. Explain how the thresholds break between designations of "Lead", "Unknown" and "Non-Lead" will be set.
13. Explain how the model addresses galvanized requiring replacement.
14. Explain what happens at the end of the contract with the consultant.

Predictive Modeling Report

Prior to submitting the initial service line inventory using predictive modeling methods for the purpose of assigning service line materials, the waterworks owner must provide a report. Waterworks must thoroughly explain and defend decisions, including the thresholds used to determine the material of the service line. The report must address:

1. The scope of the predictive model. For example, the predictive model addresses all service connections with unknown materials on the utility and/or customer service lines. If the scope of the model is unusual, for example, addressing more than one service area or distribution system, additional explanation and justification is necessary.
2. State the number of unknowns (utility side and customer side), number of field verifications (utility side and customer side) and the field verification method(s).
3. Explain how the project approach ensured representative or unbiased data, including the "training data" and "testing data".
4. Explain how "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.
5. Explain how the modeling approach treated historical records, including how the modeling team assessed reliability of historical records.
6. State the records or information supplied by the utility used by the model.
7. State the records or information beyond that supplied by the utility used by the model.
8. Explain how the modeling process incorporated new information.
9. Explain the sequence of collecting additional information and running the model after the initial run, including how many times the model was calibrated and run.

10. Explain how success was measured.
11. Explain how model results were used to develop the inventory, including how the modeling team maps the service line "ranking" into a material classification for the LSL Inventory reporting.
12. Explain the thresholds between designations of "Lead", "Unknown" and "Non-Lead" and how they were set.
13. Explain how the model addresses galvanized requiring replacement.
14. Explain the next steps, including additional or ongoing work on the LSL Inventory following the submittal of the initial LSL inventory. Explain if predictive modeling is continuing or ending.
15. Provide a detailed listing of the model output, including the predicted likelihood of lead for each service line (utility side and customer side) and the resulting material assignments.

Report Review

Waterworks must provide the report to VDH ODW for review prior to final submission of the inventory spreadsheet. VDH ODW will review the results of statistical analysis and predictive modeling and retains the authority to reject the thresholds and associated service line material determinations, particularly those that were predicted to be "non-lead". Water systems may also be required to, at the request of VDH ODW, conduct additional field investigation, including possible excavation, of these service lines, particularly those that were predicted to be "Non-lead."

General Project Sequence