16 Jan., 2025

Guidance for Using a Statistical Method to Complete a Service Line Inventory

Kentucky public water systems (PWS) may wish to supplement the Service Line Inventory required to comply with the 2021 Lead and Copper Rule Revisions with a statistical approach. The statistical approach provides a method to complete a service line inventory while reducing the need to inspect every service line of unknown material.

Division of Water (DOW) recommends consulting with Division of Water (DOW) staff or seeking external assistance with this method.

This document applies to drinking water systems that have no known lead service lines (LSLs). Those systems that have known lead service lines will require modifications to this method. Contact the Division of Water (DOW) for further assistance.

The method described in this document is used in conjunction with other methods to be used to identify service line materials. Water systems must continue to utilize other methods, including records review and materials documentation during routine operations.

If using this method to complete the Initial Service Line Inventory, note: 1) This does not eliminate the requirement to submit annual or triennial updated inventories as required by 40 CFR 141.90(e)(3); and 2) Communicate with customers as described in **Step 10** of this document.

With this guidance, water systems will assess the likelihood that unknown service lines are non-lead by physically verifying a sample of service lines. ***Unknown* service lines** are defined as lines of unknown material with no documented records.

Two key factors in the success of this strategy are:

1) the use of a ***randomly*** generated list of unknown service lines to be physically inspected; and

2) ensuring that the entire distribution system is well-represented with the service lines that are randomly chosen for physical inspection.

Disclaimer: The policies and procedures outlined in this guidance are intended to supplement existing requirements. Nothing in the policies or procedures shall affect regulatory requirements. The policies and procedures herein are not an adjudication or a regulation. Kentucky Department of Environmental Protection (DEP) does not intend to give this guidance that weight or deference. This document establishes the framework, within which DEP will exercise its administrative discretion in the future. The DEP reserves the discretion to deviate from this policy statement if circumstances warrant.

Before Getting Started

Before starting a statistical approach, it's important to establish the following:

1. A data management plan (i.e., how the data will be handled throughout the lifecycle of the project)
2. Standard operating procedures used and how it will be communicated to customers.
3. A list of any assumptions made during the statistical approach

See **Appendix A** for detailed information.

Identification Process

Before using the statistical approach to identify unknown service lines, PWS must first complete a records review. **If no LSLs** are identified during this process, the steps outlined in this document may be used. A list of records to review can be found in the SL Inventory Verification Methods guidance.

It's important for PWS to verify the accuracy of any historical records used. Field verifying a subset of sites *for each record type* and comparing with the records is one way to demonstrate accuracy. Have a well-defined procedure in place to address incomplete, inaccurate, or unreliable records.

Note: If ANY service line is found to be a lead service line, then the methodology of this document will need to be supplemented with additional steps. **Contact the DOW for further guidance.**

Key terminology:

- **LSL** = lead service line; the pipe running between the main water line and the building premise plumbing (including sections owned by the utility and owned by the customer), made primarily of lead
- **Representative sample set** = for the purpose of this document, a representative sample set is one that includes samples from all areas of the distribution system, with an emphasis on ensuring that areas most likely to have lead service lines are fully represented in the sample set (e.g., areas that are were constructed in the early 1900s or before).
- **Sample** = for this statistical method, a sample is a service line to be physically verified. A single sample includes both the utility-owned and the privately-owned sections of service line, as well as both active and inactive service lines.
- **Stratum (plural: strata)** = a group defined by certain variable(s) used to identify how to focus the selection of samples (e.g., groups of service lines, where each group is defined by the variable: age of building construction)
- **Unknown service line** = service line of unknown material with no documented records or inspections of the material composition, or where accuracy of existing records is questionable. Records to be used to identify service line materials are defined in 40 CFR 141.84(a)(3).
- **Variable** = a criterion that can be measured and that may influence the type of service line material, such as 'age of construction' or 'development zone.' Variables are used to identify strata for the selection of samples, or to ensure that the sample set represents the whole distribution system.

Note:
The guidance provided in this document may be revised based on the Lead and Copper Rule Improvements and input from the EPA.

Step-by-Step Method

Step 1: Define strata (groups) of service lines

Separate out any service lines that have been verified via records review.* These will not be part of the statistical method.

> *Service lines verified with records whose accuracy is questionable remain in the *Unknowns* category.

Starting with all the **Unknown** and **previously physically verified** service lines:

> **First**, move or filter them into a separate dataset from the service lines identified via records review.

> **Second**, organize them into groups, called strata (singular: stratum), based on date of construction (age).[1] Have at least 3 strata. The exact strata set may vary by water system, but a suggested breakdown is:

> **Stratum 1:** Buildings constructed before 1940

> **Stratum 2:** Buildings constructed 1940-1949

> **Stratum 3:** Buildings constructed 1950-1969

> **Stratum 4:** Buildings constructed 1970-present[2].

Choose a stratification method that adequately represents the entire distribution system, without being overly complex.

Step 2: Identify how many service lines must be physically inspected

- PWSs with fewer than 1,500 Unknown service lines must physically verify at least 20 percent of the lines.
- PWSs with more than 1,500 Unknown service lines must physically verify enough lines to reach a minimum 95 percent confidence level that service lines are not lead. See Appendix B to determine the number of service lines requiring verification. The Appendix B table uses a confidence level of 95 percent.

Step 3: Randomly select service lines within each stratum for physical inspection

Organize the information this way:
1. Identify which service lines fall into each stratum of Step 1.
2. Based on the number of service connections in each stratum, allocate a portion of the total number of samples needed. For example, if the "1940-1949" stratum has 30% of the total service connections served, then 30% of the required physical inspections come from that stratum.
3. Selection within each stratum **must be uniformly random** and not selected based on any specific criteria which can introduce bias. In other words, each service line within a stratum

---

[1] Water systems that are not able to find reliable data on the age of construction or plumbing installation may choose to use a "simple random sample" of service connections. In this case, the randomly selected samples described in Step 3 will be chosen from the entire distribution system at once, rather than from each stratum individually. In the submission to DOW, the water system would need to demonstrate that the simple random sample adequately covers the entire distribution system (i.e., demonstrate that the service lines selected aren't clustered in one area or another, and that no area of the distribution system was unsampled).

[2] Buildings built after Kentucky's lead plumbing ban of January 1, 1988 are unlikely to contain lead service lines; however, users are urged to keep these in the statistical model to compare the model's prediction capability of service line material in older buildings (similar to a 'Control group').

must have an equal chance of being chosen for verification.

Random selection is extremely important in this step. This is key to ensure that the service lines selected are fully representative of each stratum, and are not selected because of any biases or assumptions that may be made about service line material. See Appendix C for an easy way to generate a uniformly random set of service lines for inspection.

Alternate Sites: Have a plan for choosing alternate sites randomly, in the event that a site chosen for physical inspection cannot be sampled for some reason. One solution could be to select more sites than needed in each stratum. In order to ensure randomness for alternate sites, use these two criteria: 1) outline very specific criteria about when a sample site can be skipped (including documented refusal by homeowner, extreme difficulty accessing one or both sections of the service line, etc.) and ONLY use those criteria for skipping sites, and 2) always field verify service lines in the order of the initial randomly-generated list, without skipping any unless the criteria outlined above occur; when an alternate site needs to be chosen, simply move down to the next site on the list.

Water systems could also offset this problem by field-verifying more than the minimum number of sites in each stratum. Sampling more than the minimum is also recommended because it provides added insurance that the entire distribution system will be well-represented by the sampling effort, and it reduces the need to randomly-select replacement samples later.

Step 4: Ensure each stratum, and the distribution system as a whole, are well-represented by the samples

After randomly selecting the service lines to be sampled, review where they are located within the entire distribution system to ensure that all areas are well-represented by the sample set. This is easiest done if the selected service lines are plotted on a map. Areas with a greater building density will have more samples.

Another way to test whether the sample set is representative of the distribution system is to use a table or histogram to compare the sample set to another variable. For example, if the distribution system has distinct areas that all need to be adequately sampled, like zones, census blocks, or meter-reading tracts, those can be used to assess whether the model represents all areas. Below is an example of how that could be done:

System population: 10,000
Number of Unknowns in the system: 9,000
Number of samples needed: 368
Zones: Residential (R-1 and R-2), Agricultural (A-1), Business (B-1)

Zone	Number of Unknown SLs	Percent of Unknowns in Zone	Number of Random Samples in Zone	Percent of Samples in Zone
Zone R-1	2500	**28%**	71	19%
Zone R-2	4500	**50%**	210	57%
Zone A-1	1500	17%	64	17%
Zone B-1	500	6%	23	6%
Totals	**9000**	**100%**	**368**	**100%**

Even though in Step 3, you ensured that there were sufficient samples in each of the strata chosen, it is important to also ensure that other variables that may impact service line materials are also considered. You can see from this example that two areas (the two Residential areas R-1 and R-2) are not well represented by the initial sample selection – Zone R-1 has too few samples and Zone R-2 has too many samples. While oversampling an area may result in a more robust data set for that zone, undersampling an area may reduce the confidence that the data set fully represents the conditions of the zone. In this case, choose a new random sample, and then repeat this process.

Repeat this exercise with at least two variables (do not include building age if it was used for stratification already). These may include the geographic data mentioned above, or socioeconomic factors (income, home size, presence of vulnerable populations, disadvantaged community), land value, town/county/community lines, or other factors that are relevant to the area served. Some tools that may help with this are:

- U.S. Census Bureau Census Block Viewer - this is a good tool to view how many households are in each census block in a community
- EPA's EJ Screen
- The Center for Disease Control and Prevention's Social Vulnerability Index
- Climate and Economic Justice Screening Tool

Note: selecting more sites than the minimum can also help ensure the distribution system will be better represented by the sample set.

Step 5: Conduct a two-point (or more, if needed) physical inspection
Physical identification is required of at least one point of *each portion* of unknown service line:
- one point on the customer-owned section, and
- one point on the utility-owned section.

Physical identification methods include excavation, televising, in-home inspections, and other emerging methods[3] and are conducted or overseen by water system personnel. If inspecting near the meter, be sure the observed material is the actual service line and not part of the metering components or connectors. Refer to EPA's "Guidance for Developing and Maintaining a Service Line Inventory," Chapter Five, for methods of service line physical identification.

If one or more of the original randomly selected sites cannot be physically inspected, replace it by **randomly** generating a new site in the same stratum using the process described in Appendix C.

Skipping a site from the randomly-generated list: Only skip sites when absolutely necessary. Examples include: home/building owner refuses access (must be documented); site is no longer a service line connected to the distribution system; site is inaccessible or extremely difficult to access on one side or the other of the meter. Note: in some circumstances, many sites may be inaccessible (e.g., older downtown areas where the majority of surfaces are paved). Have a contingency plan to work within these constraints so any areas most likely to have lead service lines are still effectively sampled.

Step 6: Record results of the physical inspection process
Record the physical inspection results in the database. In the DOW Service Line Inventory template: record the material and verification method in the System-owned and Customer-owned Service Line columns. (Verification method: choose the "Field verified" option that

---

[3] Emerging methods must be approved by the Kentucky Division of Water before use.

applies).

Step 7: Enter results for remaining Unknown service lines

Remaining Unknowns: record the verification method as *Statistical Method* or *Predictive Model.* Record the system-owned and customer-owned service line material as *Non-lead.*

Note: If *lead* service lines are found during physical verification of service lines in any stratum, further investigation will be warranted. ***Do not* record the remaining unknowns in the system as non-lead. Consult DOW for further guidance.

Step 8: Retain identification records

Create, compile, and retain documentation of all service line identification efforts. DOW may ask PWS to produce or submit these records.

Step 9: Continuously update records with new information

Create a standard procedure to continuously document service line materials during routine operations and maintenance (O+M) and with any distribution system upgrades, and record those into the Service Line Inventory. Update any service line whose verification method has been recorded as 'Statistical method' to 'field-verified' as information is available. In addition, update entries that identify service line materials verified based on historic records alone if a field inspection identifies a material different than what the historical records stated.

If any lead service lines are identified in the future, further steps are warranted; **consult with DOW**.

Step 10: Communicate with Consumers

Because the water system does not have data about the service lines that were not identified with a records review or physically verified, it is important to communicate clearly with the consumers what level of confidence the water system has in the inventory. Plan to send a letter in the first year to each customer whose service line is identified as 'Non-lead' through the statistical method only. (Customers whose service line is identified as 'Non-lead' through records review or an alternative verification method can be exempt from this mailing). Offer to physically inspect these service lines upon request, as well as explain what information the statistical method provides. An example of an appropriate way to inform customers is:

> [Water system name] found ***n*** out of **[384]** randomly selected service lines were not lead. Therefore, we are 95% confident that fewer than 1% of the unknown service lines are lead. We are going to classify all of those unknown service lines as Non-lead. We will continue to document service line materials in the future during routine operations in the distribution system. If you would like [Water system name] to physically inspect your service line to verify the material, please contact us at [phone number].

Remember to provide definitions and explanations about service lines, as needed.

In addition to sending the letter to customers, it would be a good idea to include a statement in the Consumer Confidence Report (CCR) that states that a statistical model was used to complete the Service Line Inventory and that customers may contact the water system to request inspection of their service line material.

Submission to DOW

If choosing to use this statistical method, submit the following to DOW prior to submitting the

initial service line inventory:

1. The following, as described in Appendix A:[4]
    a. Description of the data management and storage plan
    b. Standard operating procedure(s) used and how it will be communicated to customers.
    c. A list of any assumptions made
2. Contact information of person(s) who is(are) responsible for executing this method at the utility, and a brief explanation of their qualifications to do so. This could be water system staff, a technical assistance provider, or a consultant.
3. Description of the strata used, and what data sources were used to determine the stratification scheme.
4. Documentation of physical inspections that have been/will be performed.
5. List of circumstances under which sites may be skipped and how skipped sites are replaced with new sites.
6. Procedure to be used to ensure that service line materials are documented during routine operations in the months and years after the initial Service Line Inventory is completed. Demonstrate a plan for continual updates to the Service Line Inventory database with information gathered during routine work such as leak repairs, meter replacements, main line updates, etc.
7. Example letter that will be sent to customers whose service line is listed as 'Non-lead' via the statistical method.

Please submit these documents by uploading via eForm 169: Drinking Water Information and Data Submittal.

---

[4] Upon request, DOW can pre-review these documents before the water system moves on to full implementation of the model.

Appendix A
Guidelines for Documents to be Submitted to DOW

Data Management Plan:
- Describe the organizational system that has been developed to store all the data that will be accumulated through the process. Each property / connection has its own row in a spreadsheet, file, attribute table, or other organization. The data management system must differentiate between the utility-owned section and the customer-owned section of the service line. In addition, have a method to refer or link to any digitized records, such as photos, tap cards, or maps to enable easy access. Include a "data dictionary" that defines and explains the data management system components (e.g., defines column headers, worksheets, or layers).

Standard Procedure and Customer Communication:
- PWS shall provide a summary to DOW of how the statistical method is implemented, including a description of any assumptions made, data verification methods used, the process and results of field-testing, and any other descriptions of how the method is implemented. In addition, list any consultants or experts that assisted with the process, along with their credentials.

- A standard operating procedure describing how the statistical method would be carried out, and including all information in the bullet above, is acceptable for submitting to DOW.

- Describe how the water system ensured that the entire distribution system was adequately represented. Be able to demonstrate that all areas/populations served by the system are included in the sampling method.

- When the initial inventory is completed, send a letter to customers with a description of the statistical method. Description of process must contain phrasing like: "[Water system name] found n out of [384] randomly selected service lines were not lead. Therefore, we are 95% confident that fewer than 1% of the unknown service lines are lead. We are going to classify all of those unknown service lines as Non-lead. We will continue to document service line materials in the future during routine operations in the distribution system. If you would like [Water system name] to physically inspect your service line to verify the material, please contact us at [phone number]."

- It is strongly suggested to add a statement in the CCR that this method was used. A statement could say: "[Water system name] completed the service line inventory using a statistical method. For more information about how this method was implemented, contact us."

List of Assumptions:
- Describe any assumptions made during this process, including any made when selecting strata or assessing whether the sampled service connections adequately represent the distribution system.

- Examples of assumptions include: 1) based on records review, we presume there are no lead service lines served by the system; 2) presence of lead goosenecks or connectors does not correlate with the potential presence of lead service lines; or 3) homes of undocumented age located in subdivisions with homes of known age are likely within the same age range as surrounding homes.

- List any factors within the distribution system that were noted which could influence the likelihood for lead service lines to be present.

Appendix B
Minimum number of service lines requiring physical inspection

Number of Unknown Service Lines*	Number to Physically Inspect
1,500 or fewer	20% of service lines
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
more than 90,000	384

*For the purposes of this process, this number represents the number of service lines that cannot be categorized from records, installation date, diameter, or customer data. If the number of service lines falls between two values on the chart, round up to the higher number.

Appendix C
Generating a uniformly random set of service lines for inspection

You can use a spreadsheet (such as Microsoft Excel or Google Sheets) to generate a uniformly random set of locations of unknown service lines for inspection using the following Microsoft Excel steps (the same formulas and method work for Google Sheets):

1. In the first column of a spreadsheet, list every unique service line of unknown material. They can be listed by address, service line ID, or other identification method.

2. In the second column, generate uniformly random numbers, so that each service line is associated with a randomly generated number. Follow these steps:

   a. Enter the formula =RAND() into the second column next to each location and press Enter. This generates a number between 0 and 1 for each service line.
   b. Select the second column (the column with the random values) and copy it, using the spreadsheet's Copy feature.
   c. With the second column still selected, use the Paste Special option to Paste Values Only into that same column. This will ensure your random numbers remain static.
   d. Use the Sort feature to list the randomly generated numbers from lowest to highest. If the Sort Warning appears, select Expand the Selection, then Sort.

3. Select only the top *n* service lines, where *n* is the number requiring inspection. For example, if you need to inspect 20 service lines, select the first 20 service lines on the list. These are the 20 uniformly random service lines to be inspected.

See the brief online tutorial for generating random samples in Microsoft Excel.

Appendix D
Invalidation Criteria

Ensure that the statistical method adequately and accurately represents the service lines in the distribution system. Avoid the following situations:

1. Inaccurate records to verify building age, installation date, or service line material.
2. Physical verification by customer with no supporting documentation. (Supporting documentation may include photo of the service line clearly demonstrating the material, receipt of new service line installation, etc.)
3. Non-random selection of service lines that are physically verified.
4. Sites used to verify the method were physically verified before deciding to use a statistical approach. (Always use a random selection of sites.)