Frequently Asked Questions (FAQ) about MassDEP DWP Statistical Analysis and Predictive Modeling Requirements

Updated September 2025

1. **Does MassDEP DWP allow PWS to do different levels of models for statistical analysis and predictive modeling?**

Yes, *MassDEP DWP allows the following levels of models:*

- System Wide Level
- Neighborhood Wide Level
- Water Main/block level

Systems are reminded to discuss with their contractor/persons completing the analysis/model what the best model may be for their service area, and which has the most representative results. Systems with lead congregated in certain areas of the service area/town may benefit from a neighborhood level approach, to focus on areas with a higher likelihood of lead, while others may prefer a system wide level.

PWS should discuss the planned procedure for their model with MassDEP DWP, should they plan to create an analysis or model, when meeting with MassDEP DWP to discuss analysis/model approval.

2. **How should PWS account for possible biases in their predictive model?**

It is important that the model is used in a way that prevents biases. Biases might appear when specific home or neighborhood types show up too frequently or not at all in the data used for prediction. For instance, if a city's historical records are concentrated in one neighborhood, the model may perform well there but fall short elsewhere.

It is also possible to introduce biases when predicting service line materials by using only tie cards or only housing age or building codes.

PWS should plan to address biases by doing the following:

To avoid neighborhood bias:

- Gather representative data to feed the model
    - Service line data of all expected materials.
    - Service line data from multiple regions of the PWS service area.

To avoid tie card bias:

- Provide numerous inputs into the model.
    - Tie cards.
    - Building age
    - Customer self-identification
    - Construction codes

PWS are encouraged to discuss all concerns with their contractor and continue to train their model with multiple iterations to strengthen the results.

3. Can a PWS use their entire service area to pull from their investigation pool?

PWS may use their entire service area installed before 1986 of known and unknown service lines to pull from for their investigation pool. This can then allow PWS to use service lines where a service line material is already known, instead of requiring immediate field inspections. However, this pool of service lines must meet the required numbers in Appendix A, Table A to be **statistically significant.**

Please note that service lines known to be installed after 01/01/1986 cannot be included in your statistical analysis. This is because these service lines may already be classified as non-lead due to the installation date, and the purpose of the statistical analysis/predictive model is to identify Lead Status Unknown service lines installed prior to 1986.

4. How is the GRR acceptance limit calculated, i.e., what is the GRR acceptance process?

This acceptance limit is generated for PWS by doing the following:

- Accounting for the total number of service lines inspected/included in your pool of randomly chosen service lines, which was used to create your statistical model **(Total # of Service Lines Identified to Create your Statistical Model).**
- We then review the information included in your report and SLI to calculate the total number of galvanized service lines found during inspections/included in your random group of service lines which are identified to create the statistical analysis. If this number is not provided in the report, it may be calculated based on the number of Field Inspected service lines included in the SLI, and the total number of Galvanized service lines discovered with this verification method **(Total # of Private Galvanized Service Lines Identified in your Random Service Line Pool)** *(If this information is unclear, MassDEP DWP can reach out for further clarification).*
- We then divide the total number of galvanized service lines found by the total number of inspections**: (Total # of Private Galvanized Service Lines Identified in your Random Service Line Pool) / (Total # of Service Lines Identified to Create your Statistical Model) = GRR %. GRR %** is your PWSs percentage (%) of the Highest Number of Estimated GRRs, it must be equal to or lower than the acceptance limit of 2.5% to be accepted.
- We then use that percentage and multiple it by the total number of service lines in your SLI which could be GRR if a galvanized service line was found on the private side **(Total Number of Service Lines which Could Be GRR).** Meaning, the total number of service lines which have been classified as NON-LEAD due to statistical analysis, which are: UNK-LG on the public side, were lead previously, or it is unknown if they were ever lead previously. **GRR % * (Total Number of Service Lines which Could Be GRR) = GRR #. GRR #** is the highest estimated number of GRR service lines we could expect in your PWS. This number must be 25 or lower to be accepted by the GRR acceptance limit and process.
- We then compare these numbers (**GRR % and GRR #)** to the acceptance **limit (2.5% & 25).** If <u>either</u> of the numbers are above the acceptance limit, the analysis cannot be accepted at this time.

Equations:

$$\frac{\text{Total \# of Private Galvanized Service Lines Identified in your Random Service Line Pool}}{\text{Total \# of Service Lines Identified to Create your Statistical Model}}$$
$$= GRR\ \%$$

$$GRR\ \% * \text{ Total Number of Service Lines which Could Be GRR } = GRR\ \#$$

(GRR % and GRR #) is then compared to the acceptance limit (2.5% & 25)

For example:

- The system has **500** service line inspections for their statistical analysis**. (Total # of Service Lines Identified to Create your Statistical Model).)**
- The system found **15** galvanized service lines during these inspections/in your random service line pool. **(Total # of Private Galvanized Service Lines Identified in your Random Service Line Pool)**
- Total service lines that are predicted through statistical analysis and could be GRR if the private service line was galvanized is **2,000** service lines. **(Total Number of Service Lines which Could Be GRR)**
- 15 / 500 = 0.03 or **3%** *(3% = GRR %)*
- 2,000 * 0.03 = **60** *(60 = GRR #)*

This systems number and percentage of possible estimated GRR service lines is **60 & 3%,** compared to the GRR acceptance limit of **25 and 2.5%,** this PWSs analysis cannot be accepted.

5. **If our PWS does not meet the GRR acceptance limit and process, can our PWS conduct a predictive model for Galvanized pipes (on the private side)?**

Running a predictive model on the private side to find the likelihood of GRR on the private side is acceptable. Since this would be a predictive model created in addition to the original statistical analysis, the GRR acceptance limit would be bypassed, as the predictive model would instead provide the percentage of likelihood of GRR per each service line. If doing so, PWS may use a standard of 20% to determine the likelihood of galvanized. Meaning, anything that is 20% or less likelihood of galvanized can be classified as non-lead/non-galvanized (UNK-NOLG), and anything that is greater than 20% likelihood, can be called galvanized (G) or UNK-LG (unknown, may be lead or galvanized).

6. **My PWS did not meet the GRR acceptance limit and process, what should we do now?**

There are multiple ways to meet the GRR acceptance limit, depending on what your percentage and number of possible GRR service lines is determined to be.

These methods are as follows, and multiple methods can be used:
- Investigate more service lines that are unknown to increase your number of inspections, which if not galvanized, can reduce your GRR %.
- Identify more service lines are and never were lead on the public side. Reducing this number can reduce your GRR #.

- Go forward with a predictive model to determine if your public service lines were ever lead, a statistical analysis to determine if the public service lines are/were ever lead, or if the private service lines could be galvanized.

7. **Why are possible GRR service lines a concern now, when the non-lead validations will require some of these service lines to be inspected?**

The point of this strategy is to ensure systems do not classify many service lines as NON-LEAD now, when, due to their SLI composition and the number of galvanized service lines in their SLI, we expect them to find many more GRR service lines in the future (above the acceptance limit). While the NON-LEAD Validations would find some of these service lines, the chances of the system finding all of them are unlikely.

To give an example, if a system estimates they could have over 200 GRR service lines using the GRR acceptance limit and process, but have 3,000 non-lead service lines and only have to validate 341 lines, the chance of the PWS finding all 200 possible GRR service lines is slim, and there will be customers who expect that their service line is NON-LEAD, that could be unknowingly consuming drinking water which is at a higher risk for lead.

8. **If a PWS performs a statistical analysis on the public side and the results support the conclusion that the unknown lines are not lead, can this analysis be used to classify remaining public side unknowns as non-lead, AND associated private side galvanized lines as not requiring replacement, under the "never known to be served by lead" criteria?**

If the system does not find any lead on the public side, has no history of lead service lines, and believes the system has never had lead, based on institutional knowledge and available records, t**he PWS may use this statistical analysis along with records and institutional knowledge**, and submit a **certification statement** explaining the reasoning why this PWS is believed to never have lead. Based on this, MassDEP can review this statement and if accepted, the PWS can use this statement as evidence that no lead was installed on the public side, recategorize all service lines to "No" never lead, and the GRR acceptance limit would therefore be bypassed.

If there is a way for the statistical analysis to take into consideration the history of lead while it is created, PWS are welcome to provide this information to MassDEP prior to submitting their SLI, for MassDEP review. Based on previous meetings regarding statistical analysis, it has been stated that the "ever lead" status was not determined by using a statistical analysis. If there is now a method in place to analyze this, PWS can submit information on their method of analyzing lead history for review and approval.

9. **If my PWS has determined that my PWS does not have any lead, and has never used lead (on the public side), does my private statistical analysis still need to meet the GRR acceptance limit?**

If your PWS determined that your (public) service lines were never lead due to institutional knowledge, records review, your public statistical analysis, and any other relevant information, your PWS may do the following:

Please provide a certification statement, signed by all relevant parties, which states that your PWS certifies that the service lines are not lead and were never lead to the best of your knowledge, based on the listed resources and evidence, and provide your resources and evidence. Please make sure, if using institutional knowledge, to include a list of all operators with years of experience who determined this, with their signatures. If all service lines could not be GRR, as the public side is not and was never lead, the change of finding GRR in the future would be 0, and therefore the GRR acceptance limit and process would be bypassed.