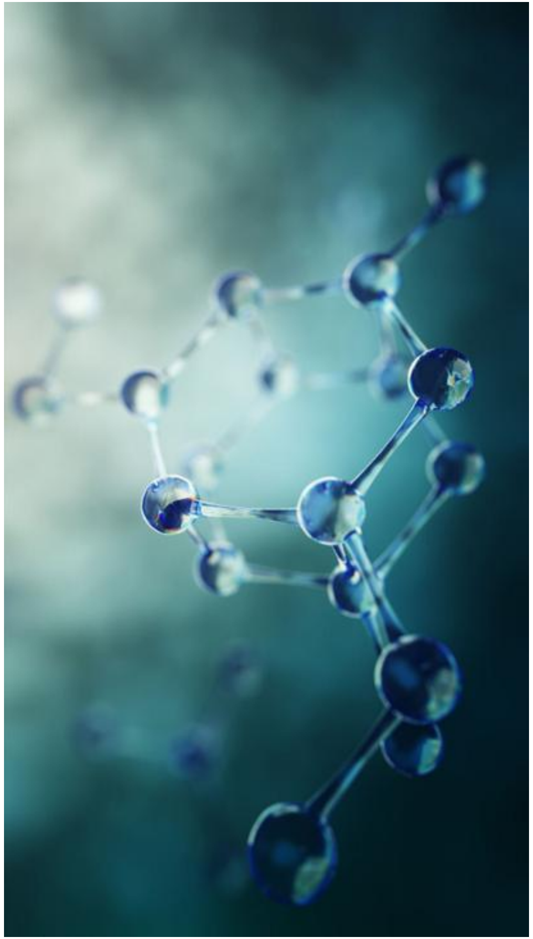
CDC’s National Center for Environmental Health

Toxicological Outbreak Investigation

Module 4: Analyzing and Interpreting Laboratory Results

Welcome

- Welcome to Module 4 of Toxicological Outbreak Investigation.
- In this module, we will discuss how to analyze and interpret laboratory results.
- This module should take about 60 minutes to complete.



Objectives

- After completing this module, you will be able to
  - Describe LOD and two constant values that can be substituted for values less than LOD
  - Identify two common examples of making adjustments to laboratory results
  - Identify and analyze data that have a log-normal distribution
  - Describe information needed to interpret laboratory results, specifically toxic levels associated with illness and comparison values

Descriptive Analysis of Laboratory Measurements

- When analyzing data from laboratory measurements for an investigation, we are often interested in describing the distribution of the values (the magnitude of the values observed and how often various values occur).
- That commonly includes graphing the values.
- It also often includes calculating values that represent the center of the distribution, called measures of central tendency.
- Two commonly used measures of central tendency are the mean and median.
  - The mean (average) is calculated by adding all values of the variable together and dividing the sum by the number of values that were added together.
  - The median is calculated by sorting the values of the variable in increasing order and taking the middle value (or the average of the two middle values if there is an even number of values).

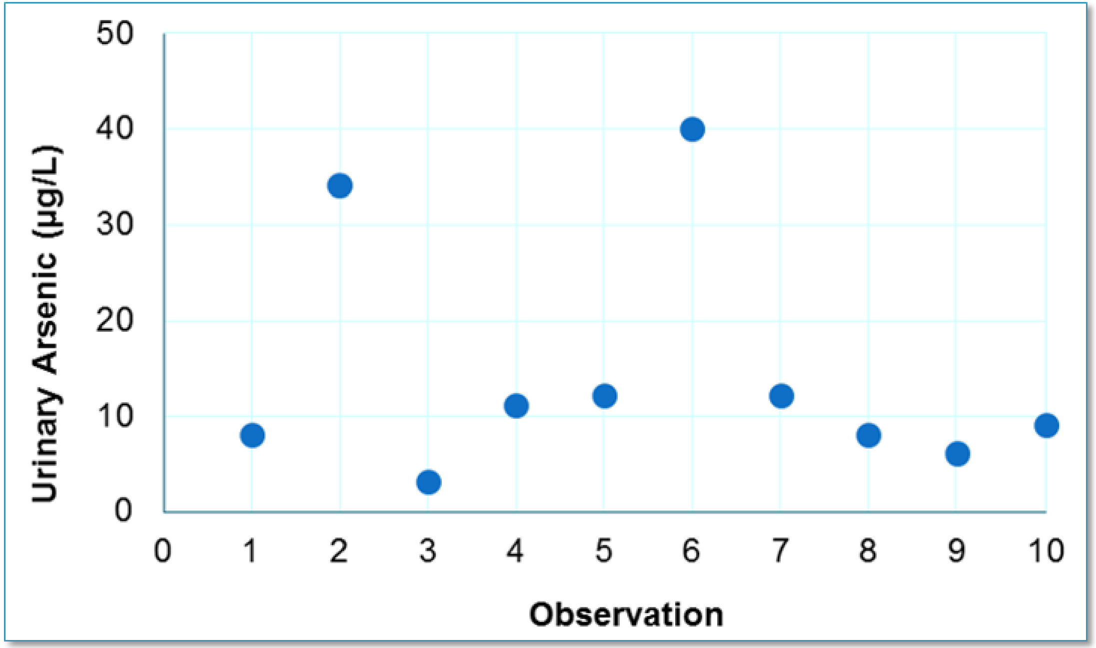
Example: Urinary Arsenic Measurements

- For example, consider arsenic concentrations in urine collected from patients involved in an acute outbreak investigation.

Observation #	Arsenic Concentration (μg/L)
1	8
2	34
3	3
4	11
5	12
6	40
7	12
8	8
9	6
10	9

Example: Urinary Arsenic Measurements (cont.)

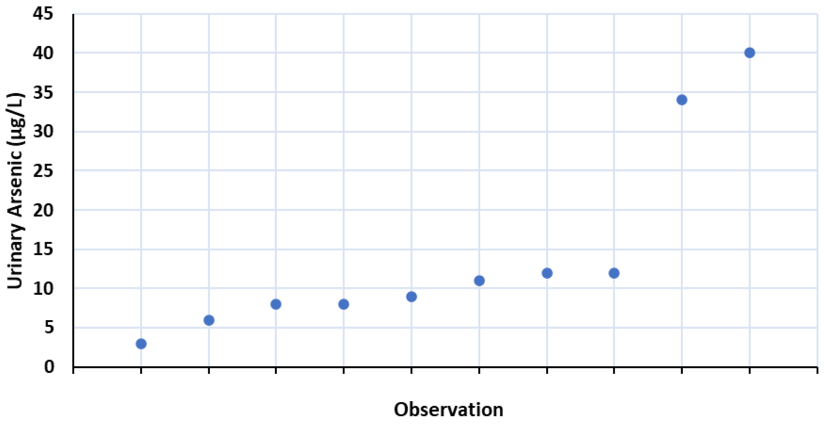
- One approach for plotting the measured concentrations on a graph might look like this.
- Here, each person is represented along the x-axis, and their corresponding urinary arsenic concentration is on the y-axis.
- ***What might be some other ways to graph these data?***



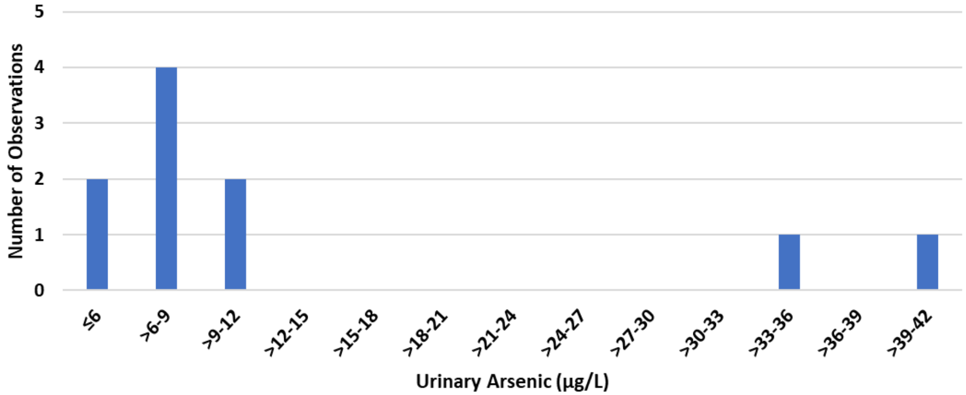
Example: Urinary Arsenic Measurements (cont.)

- Some additional approaches to graphing the data

Sorted scatter plot

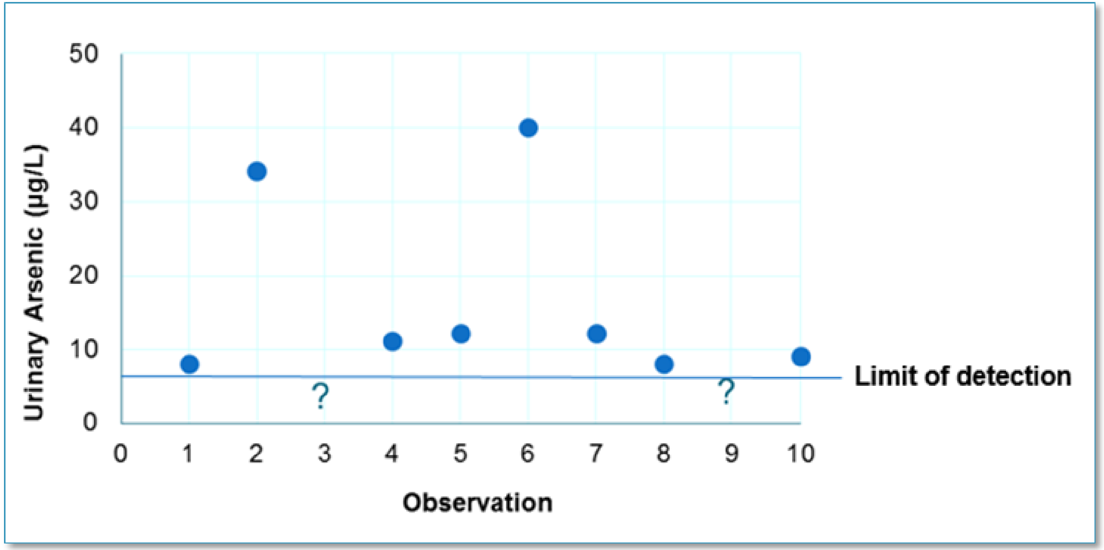


Histogram



Limit of Detection

- Most laboratory tests can only detect a toxic agent if there is a certain measurable amount present.
- This level is called the limit of detection (LOD).
- If the limit of detection for the laboratory test in the previous example is 7 µg/L, then the two observations that fall below 7 µg/L could be undetected.



Example: Urinary Arsenic Measurements (cont.)

- Observations that are below the LOD are usually reported by the lab as <LOD, or "Not Detected".
- This means the true value is somewhere from 0 up to just under the LOD.

Observation #	Arsenic Concentration (µg/L)
1	8
2	34
3	<LOD
4	11
5	12
6	40
7	12
8	8
9	<LOD
10	9

Challenge #1

What is the median of the values listed in the table?

Select the best response.

A. 16.75

B. 12

C. 10

D. Not possible to calculate

Observation #	Arsenic Concentration (μg/L)
1	8
2	34
3	<LOD
4	11
5	12
6	40
7	12
8	8
9	<LOD
10	9

Challenge #1 Answer

- **The correct answer is C: 10.**

- The correct answer was calculated by ordering the values as shown to the right, then taking the average of the numbers in the middle (#5 and #6 spots). In this case, (9+11)/2 = 10. Half of the values on the list are above 10 and half are below 10.
- The two values of <LOD are included here.
- Although those exact values are unknown, we do know that the values are less than 7 (in this example where LOD=7), so they fall in the lower half of the distribution.

1. <LOD
2. <LOD
3. 8
4. 8
5. 9
6. 11
7. 12
8. 12
9. 34
10. 40

Challenge #2

- **What is the average of the values listed in the table?**

Select the best response.

A. 16.75
B. 12
C. 11.5
D. Not possible to calculate

Observation #	Arsenic Concentration (µg/L)
1	8
2	34
3	<LOD
4	11
5	12
6	40
7	12
8	8
9	<LOD
10	9

Challenge #2 Answer

- **The correct answer is D: Not possible to calculate.**

  The value of “less than LOD” cannot be summed or averaged.

Substituting for Values <LOD

- To calculate a mean or perform statistical tests, investigators need to assign a numeric value to levels that are <LOD; these values should not be ignored.
- The choice of substitution will affect any subsequent descriptive or statistical analyses of the data.
- Two constant values are commonly substituted for values that are <LOD:
  - LOD/2
  - $LOD/\sqrt{2}$

Substituting for Values <LOD (cont.)

- In this example, where the LOD = 7 µg/L:

$$\frac{LOD}{2} = \frac{7}{2} = 3.5\ \mu g/L$$

$$\frac{LOD}{\sqrt{2}} = \frac{7}{\sqrt{2}} = 4.9\ \mu g/L$$

Take a moment to calculate the mean value using each of these substitutions.

Observation #	Arsenic Concentration (µg/L)
1	8
2	34
3	<LOD
4	11
5	12
6	40
7	12
8	8
9	<LOD
10	9

Substitution Differences

Observation #	Original Data	Ignore Values <LOD
1	8	8
2	34	34
3	<LOD	
4	11	11
5	12	12
6	40	40
7	12	12
8	8	8
9	<LOD	
10	9	9
Mean	**NA**	**16.8**

Ignoring values <LOD would result in a mean of 16.8.

Substitution Differences

Observation #	Original Data	Substitute 0 for Values < LOD
1	8	8
2	34	34
3	**<LOD**	**0**
4	11	11
5	12	12
6	40	40
7	12	12
8	8	8
9	**<LOD**	**0**
10	9	9
Mean	**NA**	**13.4**

Substituting 0 for values <LOD would result in a mean of 13.4.

Substitution Differences

Observation #	Original Data	Substitute LOD/2 for Values < LOD
1	8	8
2	34	34
3	**<LOD**	**3.5**
4	11	11
5	12	12
6	40	40
7	12	12
8	8	8
9	**<LOD**	**3.5**
10	9	9
Mean	**NA**	**14.1**

Substituting LOD/2 for values <LOD would result in a mean of 14.1.

Remember that LOD = 7 µg/L, so

$$\frac{LOD}{2} = \frac{7}{2} = 3.5\ \mu g/L$$

Substitution Differences

Observation #	Original Data	Substitute $\mathrm{LOD}/\sqrt{2}$ for values < LOD
1	8	8
2	34	34
3	**<LOD**	**4.9**
4	11	11
5	12	12
6	40	40
7	12	12
8	8	8
9	**<LOD**	**4.9**
10	9	9
Mean	**NA**	**14.4**

Substituting LOD/$\sqrt{2}$ for values <LOD would result in a mean of 14.4.

Remember that LOD = 7 µg/L, so

$$\frac{LOD}{\sqrt{2}} = \frac{7}{\sqrt{2}} = 4.9\ \mu g/L$$

Compare Results of Substitution Differences

Observation #	Original Data	Ignore Values <LOD	Substitute 0 for Values < LOD	Substitute LOD/2 for Values < LOD	Substitute $\mathrm{LOD}/\sqrt{2}$ for values < LOD
1	8	8	8	8	8
2	34	34	34	34	34
3	**<LOD**		**0**	**3.5**	**4.9**
4	11	11	11	11	11
5	12	12	12	12	12
6	40	40	40	40	40
7	12	12	12	12	12
8	8	8	8	8	8
9	**<LOD**		**0**	**3.5**	**4.9**
10	9	9	9	9	9
Mean	**NA**	**16.8**	**13.4**	**14.1**	**14.4**

Selecting a Method for Handling Values <LOD

- Ignoring values that are <LOD or substituting the value 0 are not recommended.
- The most commonly used substitutions are LOD/2 and LOD/$\sqrt{2}$.
- There are other ways to account for values that are <LOD that are more advanced than simple substitution.
  - One method is to use survival analysis procedures that account for left censoring (as opposed to the right censoring that is more commonly encountered).
  - Note: you might need to consult with a statistician or other scientist familiar with such methods.
  - For further reading to that topic:
    - Helsel, D.R. (2005). Nondetects and Data Analysis: Statistics for Censored Environmental Data. New York: John Wiley & Sons.
    - Helsel, D.R. (2012). Statistics for Censored Environmental Data Using Minitab and R. New York: John Wiley & Sons.

Making Adjustments to Lab Results

- It might be necessary to adjust toxicological laboratory results to account for factors such as how concentrated or diluted the specimen is.
- These are two common examples:
  - Adjusting for creatinine in urine specimens (i.e., creatinine corrected results)
  - Adjusting for proteins or lipids in serum specimens (e.g., lipid corrected results)

Example: Accounting for Urine Concentration Variability

- Urine dilution can vary between people and within a single person during the day.
- The picture shows how the color of urine changes with various levels of dilution (lighter color = more dilution).
- Factors that affect this dilution include water consumption and general health status.
- The dilution can affect the concentration of toxic agents in a urine specimen.
- Levels of a toxic agent in urine may decrease if a person drinks a lot of water and increase if the person is dehydrated.

Example: Accounting for Urine Concentration Variability

- Adjusting for urine creatinine can help account for differences in urine dilution.
- Common formula for determining the adjusted analyte concentration (µg/g creatinine):

$$\frac{Concentration\ of\ toxic\ agent\ (\mu g/L)}{Creatinine\ concentration\ (g/L)}$$

Observation #	Concentration of Toxic Agent in Urine (µg/L)	Creatinine Concentration (g/L)	Adjusted Measurement (µg/g Creatinine)
1	7	1.1	6.4
2	34	0.7	49
3	26	1.3	20

- Analyte and creatinine levels are measured in the same urine specimen.
- The amount of creatinine excreted is relatively constant across and within individuals, so the creatinine concentration reflects the degree of dilution of the sample, and the ratio reflects the level of the toxic agent in a way that is independent of the degree of urine dilution.
- More complex methods have also been proposed that might be useful in some situations
  - O'Brien, et al. (2017). Lipid and creatinine adjustment to evaluate health effects of environmental exposures. Curr Envir Health Rpt. 4:44–50.

Example: Accounting for Lipid Variability in Serum

- If someone has high amounts of lipids in their serum, serum levels may be higher for a substance that binds to lipids (lipophilic).
- A serum level of an analyte can be divided by a lipid measurement in the same specimen to obtain a "lipid adjusted" result.
- When a lipid adjustment is made, results are expressed as the analyte units per gram of lipid.
- Examples of toxicants that are lipophilic and require lipid adjustment include
  - Polychlorinated biphenyls (PCBs)
  - Organochlorine pesticides

Example: Serum Protein Levels

- Some substances bind to proteins (e.g., albumin) in the blood and levels of those proteins can affect the biologic activity of a given total concentration of the substance in blood or serum.
- Some examples are calcium, phenytoin, and valproic acid.
- Concentrations of such substances must be interpreted with consideration of serum protein levels and sometimes other clinical factors.
- Consultation with a toxicologist and the laboratory may be needed for interpretation of tests for such substances.

Making Adjustments to Laboratory Results

- Adjustment is a way to standardize lab results.
- Adjustments are not always needed. For example, for toxic agents that do not bind to lipids in the blood, you would not adjust for lipid concentration.
- Consult with a laboratorian or a toxicologist for guidance on whether and how adjustments should be made for your type of specimen and the analyte you are considering.

Challenge #3

- **Which of the following are common substitutions for values that are <LOD? Select all that apply.**

A. Zero

B. LOD/2

C. LOD/$\sqrt{2}$

D. No substitution is needed. These values can be ignored in calculations.

Challenge #3 Answer

- **The correct answers are B and C.**

  Ignoring values below the LOD or substituting 0 for these values will result in incorrect calculations.

Challenge #4

- **True or false. Adjustments to lab results are always necessary. Select the best response.**

  - ❑ True
  - ❑ False

Challenge #4 Answer

- **The correct answer is False.**

  Adjustments to lab results are not always needed. For example, for toxic agents that do not bind to lipids in the blood, you would not adjust for lipid concentration.

Examining Distribution of Laboratory Measurements: Infectious vs. Toxicological Outbreaks

- Infectious disease outbreak investigations often result in dichotomous (e.g., present vs. absent) laboratory data.

Observation #	Salmonella	Shigella
1	Present	Absent
2	Present	Absent
3	Absent	Absent
4	Present	Absent
5	Absent	Absent

Examining Data Distribution: Infectious vs. Toxicological Outbreaks

- Toxicological outbreak investigations often result in continuous laboratory data.

Observation #	Arsenic (µg/L)	Lead (µg/dL)
1	8	3
2	34	4
3	10	4
4	11	3
5	12	5

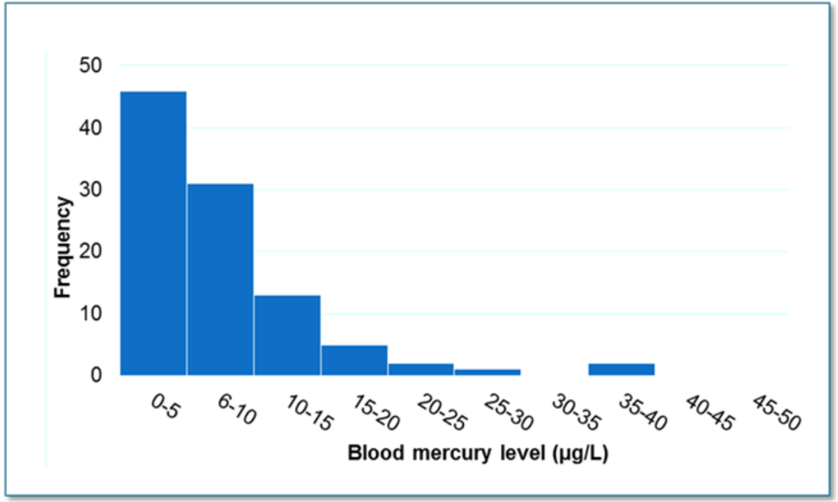
Normal Distribution

- Before analyzing continuous variables, the distribution of the measurement values must be checked for normality.
- If data are normally distributed, you can calculate means and compare groups using t-tests.
- The normal distribution is also called
  - Bell-shaped curve
  - Gaussian distribution
- A normal distribution is symmetric.

Challenge #5

- **Are the values shown in the chart normally distributed? Select the best response.**

  ❑ Yes
  ❑ No



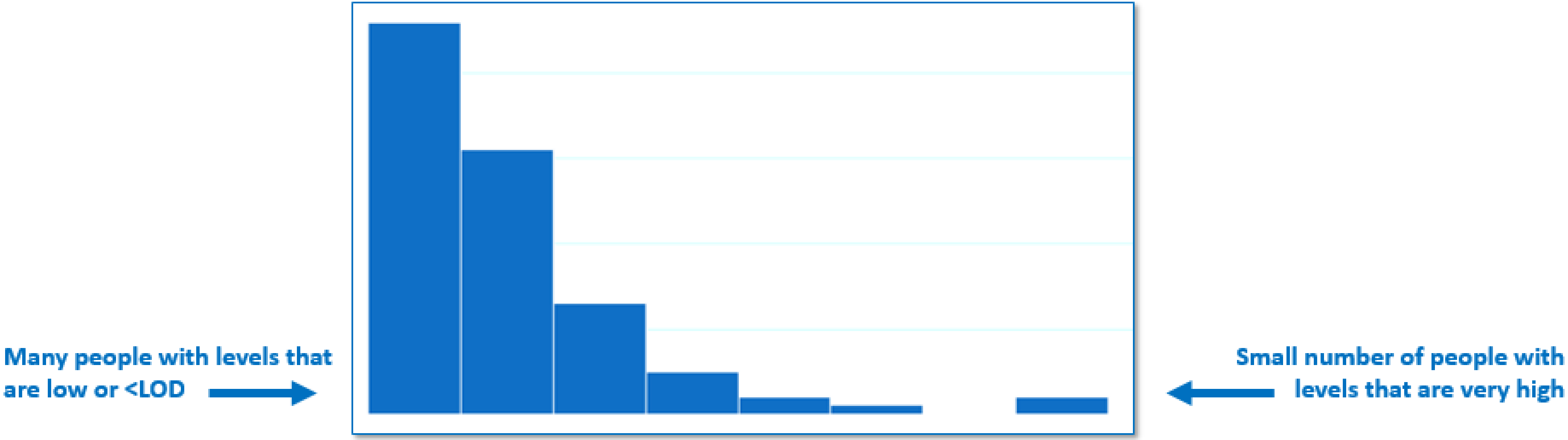
Challenge #5 Answer

- **The correct answer is No.**

  Most participants have low levels and a small number have high levels.

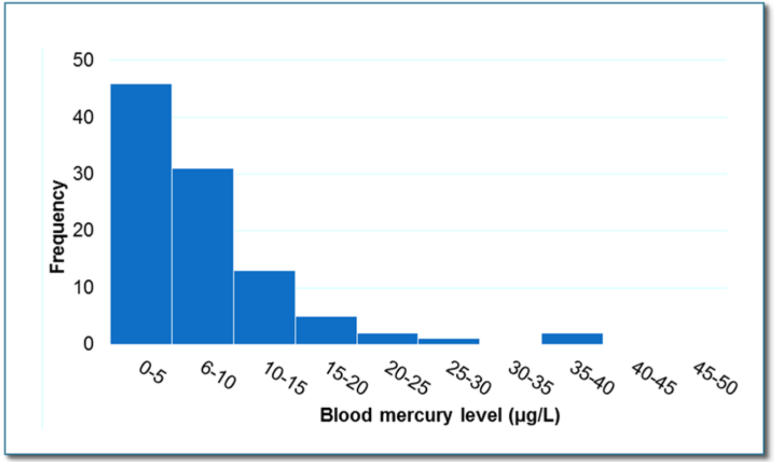
Log-Normal Distribution

- When toxic agents are measured in people, they often have a log-normal distribution.
- Log-normal distributions can become normally distributed when you take the natural log of each of the measurements.



Examining Data Distribution

- The histogram displayed below shows mercury levels in blood samples from 100 case-patients from an outbreak investigation.
- Values that were below the limit of detection received a substitution of LOD/$\sqrt{2}$.
- These data are not normally distributed because a lot of people have low concentrations and only a few have high concentrations.
- We say that these data are "right skewed" because there is a long tail to the right.



Challenge #6

- **If you calculate a mean on data that are right skewed (as shown on the previous slide), what is the typical result? Select the best response.**

A. The mean will be higher than the median.

B. The mean will be lower than the median.

C. The mean will be about the same as the median.

Challenge #6 Answer

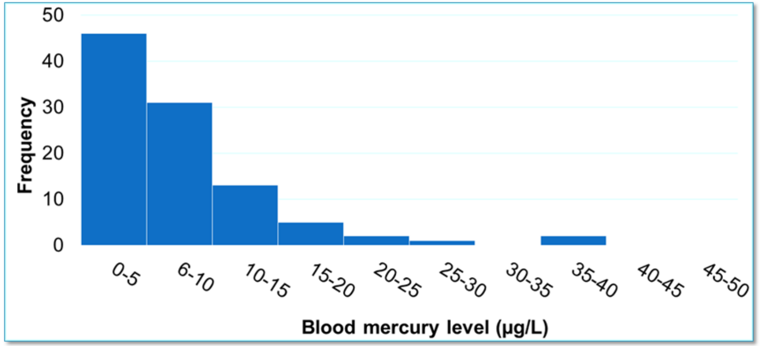
- **The correct answer is A. The mean will be higher than the median.**
  For right-skewed data, the mean will be higher than the median.

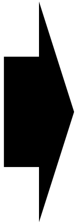
Analyzing Log-Normal Data

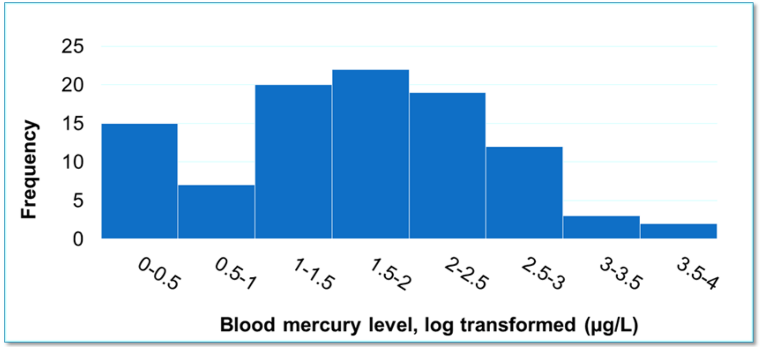
- There are two common ways to analyze log-normal data:
  - Transform the data to make them normally distributed. Then use parametric statistics (for example, the t-test) to compare levels between groups.
  - Keep the data as they are and use nonparametric statistics (for example, the Kruskal-Wallis test) to compare levels between groups.

Log Transformation

- Log transformation means calculating the natural log of each value.
- For example
  - 0.8 --> Ln(0.8) = -0.2
  - 10 --> Ln(10) = 2.3
  - 167 --> Ln(167) = 5.1
- Log transforming values pulls in very high values, resulting in a distribution that is closer to normal.







Geometric Mean

- Log-transforming toxicological laboratory data usually results in data that are approximately normally distributed. This allows you to calculate a geometric mean.
  - Geometric mean = Exponentiated average of log transformed data
  - Arithmetic mean = Average of non-transformed data
- Steps to calculate a geometric mean on laboratory data with a log-normal distribution and values <LOD
  1. Perform substitution for values that are <LOD.
  2. Perform adjustments (such as accounting for creatinine or lipids), if necessary.
  3. Log-transform the data.
  4. Calculate the mean of the log-transformed data.
  5. Exponentiate the mean.

Activity: Calculate Geometric Mean

- *Calculate the geometric mean of the data to the right.*
- *Note that the LOD was 7 µg/L.*

Observation #	Arsenic (µg/L)
1	8
2	34
3	<LOD
4	11
5	12
6	40
7	12
8	8
9	<LOD
10	9

Activity: Calculate Geometric Mean

Observation #	Arsenic (µg/L)	Log-transformed concentrations
1	8	ln(8) = 2.1
2	34	ln(34) = 3.5
3	<LOD	ln(4.9) = 1.6
4	11	ln(11) = 2.4
5	12	ln(12) = 2.5
6	40	ln(40) = 3.7
7	12	ln(12) = 2.5
8	8	ln(8) = 2.1
9	<LOD	ln(4.9) = 1.6
10	9	ln(9) = 2.2

- Step 1: Substitute values that are <LOD with values equal to LOD/$\sqrt{2}$

  LOD/$\sqrt{2}$ = 7 µg/L /$\sqrt{2}$ = 4.9 µg/L

- Step 2: Log-transform the data
- Step 3: Calculate the mean of the log-transformed data

  2.1 + 3.5 + 1.6 + 2.4 + 2.5 + 3.7 + 2.5 + 2.1 + 1.6 + 2.2 = 24

  24/10=2.4

- Step 4: Exponentiate the mean

  Exp(2.4)=11

- Geometric mean=11

Examining Data Distribution

- The geometric mean (11 µg/L) is closer to the median (10 µg/L) than the arithmetic mean (14.4 µg/L) calculated earlier.
- This pattern in biologic data is common. When data are log-normal, the arithmetic mean is skewed higher.
- For log-normal data, the geometric mean is a better indicator of the central tendency of the data than the arithmetic mean.

Challenge #7

- **Toxicological outbreak investigations often result in what type of laboratory data? Select all that apply.**

A. Dichotomous
B. Continuous
C. Normally distributed
D. Log-normal

Challenge #7 Answer

- **The correct answer is B and D.**

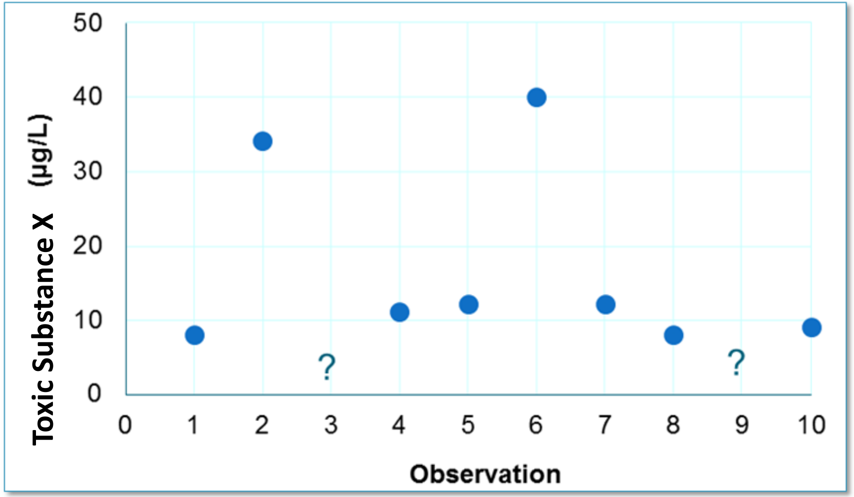
  Toxicological outbreak investigations often result in continuous data with a log-normal distribution.

Interpreting Laboratory Results

- Biologic specimens and environmental samples might contain multiple potentially toxic agents.
- Detecting a toxic agent in a biological or environmental sample does not necessarily mean that the toxic agent caused the outbreak. The toxic agent might be present at a normal background level.
- We need to interpret the levels to determine whether they are high enough to have likely caused the outbreak.

Interpreting Laboratory Results (cont.)

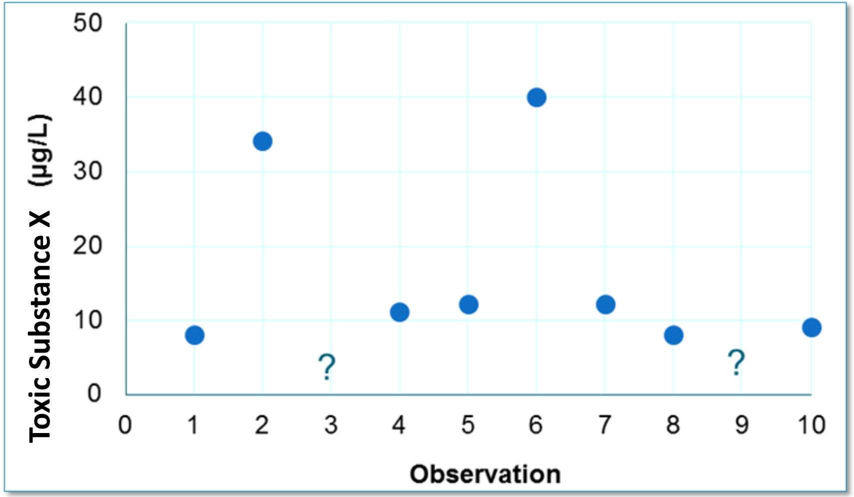
- Consider this hypothetical scenario of concentrations of toxic substance X in urine specimens collected from case-patients in an outbreak:
  - ***Based on these urinary levels, do you think that toxic substance X exposure could have caused the outbreak?***
  - ***What information do we need to know?***



Interpreting Laboratory Results (cont.)

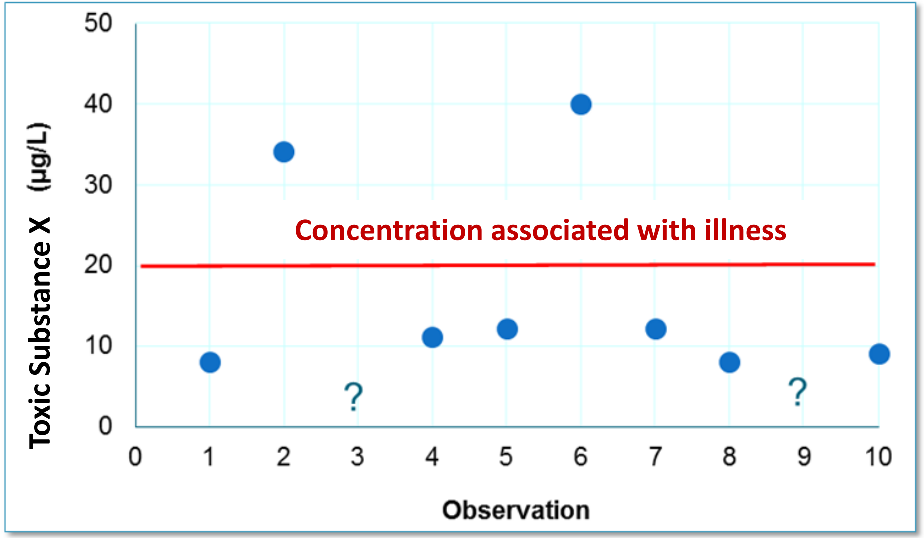
The answers to the following questions might help us interpret these values:

- What is the lowest concentration of the toxic agent associated with the type of illness being observed?
- When were the specimens collected?
- Are these levels associated with the level of the toxic agent in something that the people were exposed to?
- Did these individuals share a common exposure?



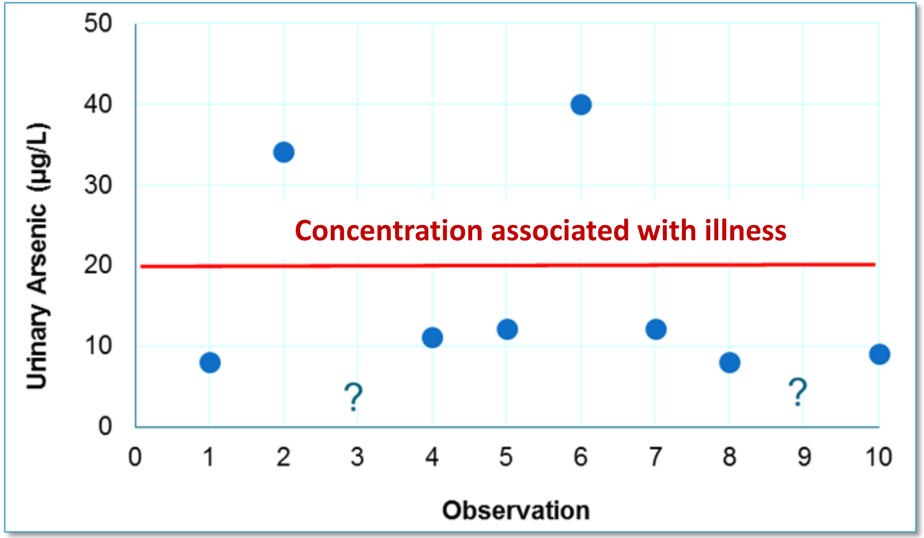
Concentration Associated with Toxicity

- If the concentration associated with illness is 20 µg/L, then you might conclude that toxic substance X exposure could be causing sickness in individuals #2 and #6.
- ***What about the other individuals? Is it possible that the toxic agent might have made them sick as well?***



Concentration Associated with Toxicity (cont.)

- It is still possible that this toxic agent might have caused health effects in people with levels below 20 µg/L.
- The toxic agent might have been excreted to a greater extent in these people (e.g., because of later sample collection).
- Those individuals might have been more susceptible, such that they experienced health effects at lower levels.



Comparison Values

- For many toxic agents, the levels that cause illness might not be known. Other comparison values are needed, such as these:
  - Concentrations of the toxic agent in comparison samples collected from your investigation (e.g., samples collected from people who were not ill)
  - Published values from previous outbreaks or research studies

  * See Module 3 for more information about comparison values.
- Here are some examples of toxic agents where good data on toxic concentrations exist or do not exist:
  - **Good data exist for**
    - Carbon monoxide
    - Cyanide
    - Toxic alcohols (e.g., ethylene glycol, methanol)
    - Lead
  - **Good data do not exist for**
    - Organophosphates
    - Pyrethroids
    - Hydrofluoric acid

Considerations in Communicating Laboratory Results

- The hardest part of communicating toxicological outbreak findings can be correctly interpreting and communicating the laboratory data.
- Reasons for not finding a particular toxic agent in samples from an investigation might include
  - The agent might not have caused the outbreak.
  - For biological samples, the agent might have been metabolized or excreted before the samples were collected.
  - For environmental samples, the agent might have broken down before the samples were collected, or the agent might have been missed because of a heterogeneous distribution in the environmental medium that was sampled.
  - The agent could have been at a level <LOD, yet still high enough to cause illness.
- Identifying a toxic agent in samples does not mean it caused the outbreak.
  - People are continuously exposed to toxic agents.
  - Typical exposures might vary from region to region and between different groups of people.

Challenge #8

- **For many toxic agents, we do not know the level that causes illness. What other information would be helpful in determining whether a particular agent might have caused the outbreak? Select all that apply.**

A. Information about whether individuals shared a common exposure

B. Information about when samples were collected

C. Information about the levels of the toxic agent in other samples collected from the investigation, such as samples from people who were not ill

D. Published values from previous outbreaks or research studies

Challenge #8 Answer

- **The correct answers are A, B, C, and D.**

  All of the responses are things that you might need to know.

Module Summary

- This concludes Module 4.
- In this module, we defined limit of detection (LOD) and identified the constant values that can be substituted for values lower than the LOD. We looked at examples of how adjustments may be made to laboratory results. We also discussed data with a log-normal distribution and described the information needed to interpret laboratory results.

Module Summary (Cont.)

- You should now be able to
  - Describe LOD and two constant values that can be substituted for values less than LOD.
  - Identify two common examples of making adjustments to laboratory results.
  - Identify and analyze data with a log-normal distribution.
  - Describe information needed to interpret laboratory results, specifically toxic levels associated with illness and comparison values.

For more information, contact NCEH
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov
Follow us on X @CDCEnvironment