Appendix D: Statistical Metadata for Results Derived from Random Inspection Data

The tables on the following pages present statistical metadata that participating states should strive to provide when reporting core measures derived from random inspection data (such as the baseline or post-certification random inspections typically associated with ERP). Providing metadata adds to the credibility of reporting and enhances users' ability to interpret the data. Failure to provide metadata for results based on statistical random samples may lead some reviewers to question the validity of the analysis.

The types of metadata requested for each measure vary somewhat from measure to measure, depending upon two characteristics of the measure being reported: (1) whether the measure reflects a single point in time or tracks change over time, and (2) whether the measure is based upon yes/no answers (e.g., an achievement rate for an EBPI) or upon continuous numeric answers (e.g., an average facility score for EBPIs or an average amount of hazardous waste being generated).

Please note that the light gray rows in the tables indicate information, like the confidence level, that states would only need to provide once. Italicized text in the tables indicates information that states should provide if available.

Measures Based on Yes/No Questions (also known as dichotomous variables)

For Results from a Single Random Sample
Table 1 identifies statistical information to be provided with measures that identify the proportion or percentage of facilities achieving a particular measure at a single point in time (e.g., baseline random inspections). Specifically, follow the guidelines in Table 1 when reporting statistical metadata for the following core measures:

2. Rate of facility certification "Yes"/"NA"/"Blank" and inspector "No" responses on EBPIs;
6. Achievement rate for each EBPI;
8. Achievement rate across all compliance-related measures;
9. Media-specific achievement rates across all compliance-related measures; and
16. Aggregate achievement rate for all EBPIs.

Table 1: Yes/No, One Sample

Item Reported	Notes
Confidence level	Minimum 90%
Population size	As of that particular sample (may vary among samples)
Overall sample size	E.g., number of valid inspections in that round
Effective sample size for the particular measure in question	I.e., the number of facilities for which the particular measure was relevant and for which valid data were collected
Number of facilities achieving the particular measure in question	Data should always be reported so that yes/achievement = good
Observed achievement rate	Reported in whole numbers if reported as a percentage, such as 90% of facilities achieving the measure; to two decimal places if reported as a proportion, such as 0.90
Confidence interval	Lower and upper bounds reported in whole numbers if reported as a percentage, such as 90% to 98% of facilities achieving the measure; to two decimal places if reported as a proportion, such as 0.90 to 0.98.
Type of confidence interval calculated	E.g., Wald, Score

For Results Comparing Two Random Samples
Table 2 identifies statistical information to be provided with measures that identify changes in the proportion or percentage of facilities achieving a particular measure time over time (e.g., change in performance between baseline random inspections and post-Round 1 random inspections). Specifically, follow the guidelines in Table 2 when reporting statistical metadata for the following core measures:

2. Rate of facility certification "Yes"/"NA"/"Blank" and inspector "No" responses on EBPIs;
6. Achievement rate for each EBPI;
8. Achievement rate across all compliance-related measures;
9. Media-specific achievement rates across all compliance-related measures; and
16. Aggregate achievement rate for all EBPIs.

Table 2: Yes/No, Two Samples

Item Reported	**Notes**
Confidence level	Minimum 90%
Population size	As of that particular sample (may vary among samples)
Overall sample size	E.g., number of valid inspections in that round
Observed difference between the rates	• Calculation is always post-certification result minus pre-certification result • Report whole numbers if reported as a percentage, two decimal places if reported as a proportion
Whether a significant difference was detected	
One-sided or two-sided test?	• *If relevant*
P-value	• *If relevant* • *To at least three decimal places*
Confidence interval for the difference between the rates	*Report lower and upper bounds as` whole numbers if reporting performance levels as percentages (e.g., confidence interval for the difference of 10 percentage points to 20 percentage points). Report to two decimal places if reporting as proportions.*

Measures Based on Numeric Data, Rather than Yes/No Questions
Such measures are also known as continuous variables. In ERP, these will typically be reported primarily as means (commonly called "averages").

For Results from a Single Random Sample
Table 3 identifies statistical information to be provided with reported results for continuous variables at a single point in time (e.g., an average facility score or average amount of hazardous waste generated, based upon baseline inspections). Specifically, follow the guidelines in Table 3 when reporting statistical metadata for the following core measures:

10. Average facility score for all EBPIs;
12. Average facility score for compliance-related EBPIs; and
14. Average facility score for all compliance-related measures.

Note that, while the above measures are reported as percentages, the actual measure is a mean. For instance, an average facility score for all EBPIs of 78.0% is calculated by adding up each individual facility score for all EBPIs and dividing by the number of facilities. The reported percentage is a continuous variable, bounded at 0% and 100%.

Table 3: Continuous Variables, One Sample

Item Reported	Notes
Confidence level	Minimum 90%
Population size	As of that particular sample (may vary among samples)
Overall sample size	E.g., number of valid inspections in that round
Effective sample size for the particular measure in question	I.e., the number of facilities for which the particular measure was relevant and for which valid data were collected. (Note that for the average facility scores in the core measures , the effective sample size will typically be equal to the overall sample size.)
Mean*	To an appropriate number of decimal places
Confidence interval for the mean	Lower and upper bounds each to an appropriate number of decimal places
Standard deviation for the mean	To an appropriate number of decimal places
Distribution of values*	Minimum, each decile, median, maximum

* Organizationally, in the template, the mean and the distribution of values might be presented as the measure, rather than the supporting information. Both are listed here for clarity.

For Results Comparing Two Random Samples
Table 4 identifies statistical information to be provided with measures that identify changes in a continuous variable over time (e.g., change in performance between baseline random inspections and post-Round 1 random inspections). Specifically, follow the guidelines in Table 4 when reporting statistical metadata for the following core measures:

10. Average facility score for all EBPIs;
12. Average facility score for compliance-related EBPIs; and
14. Average facility score for all compliance-related measures.

Note that, while the above measures are reported as percentages, the actual measure is a mean. For instance, an average facility score for all EBPIs of 78.0% is calculated by adding up each individual facility score for all EBPIs and dividing that total by the number of facilities. The reported percentage is a continuous variable, bounded at 0% and 100%.

Table 4: Continuous Variables, Two Samples

Item Reported	**Notes**
Confidence level	Minimum 90%
Population size	As of that particular sample (may vary among samples)
Overall sample size	E.g., number of valid inspections in that round
Observed difference between the means	• Calculation is always post-certification result minus pre-certification result • To an appropriate number of decimal places
Whether a significant difference was detected	
One-sided or two-sided test?	• *If relevant*
P-value	• *If relevant* • *To an appropriate number of decimal places*
Confidence interval for the difference between the means	• *If calculated* • *Lower and upper bounds each to an appropriate number of decimal places*