Weight of Evidence Approach in Data Validation

Brannon Seay, EPA OAQPS

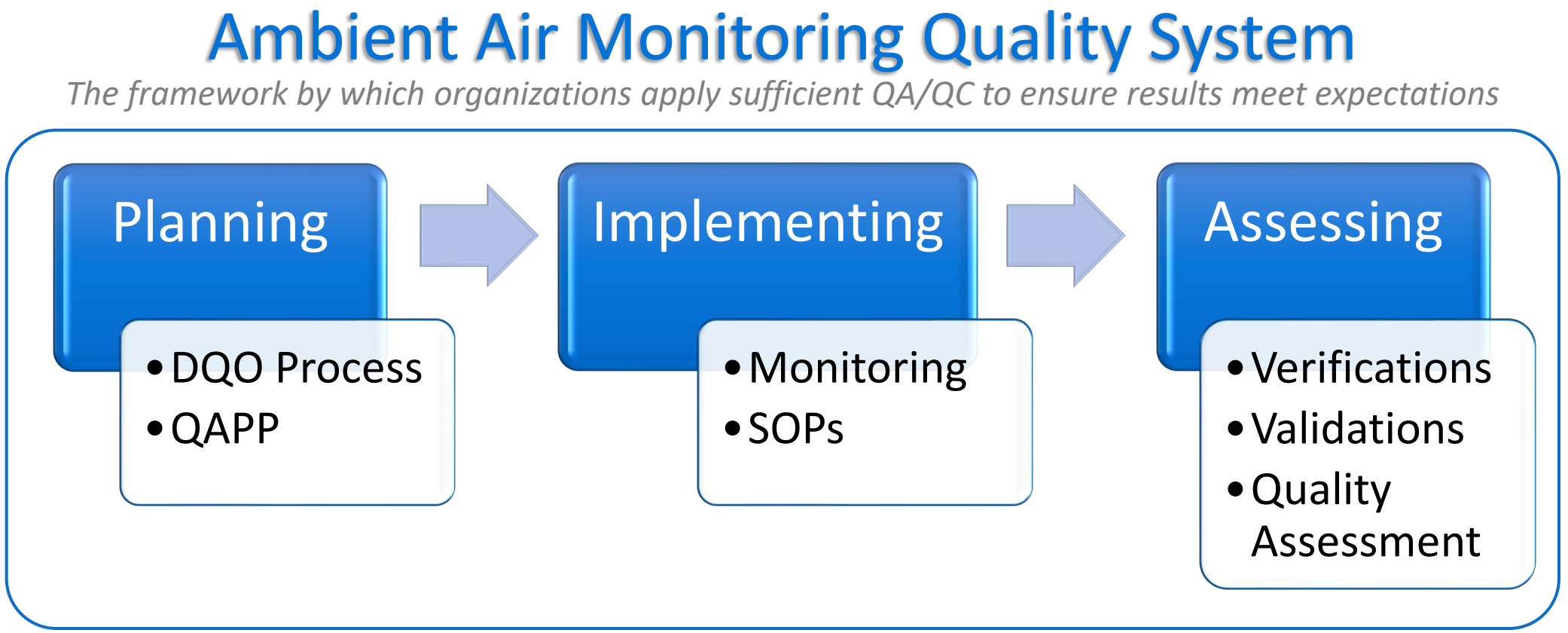
2024 NAAMC, New Orleans, LA

August 12th, 2024

Table of Contents

What is the “weight of evidence” (WoE) approach?

Elements of an Ambient Air Monitoring Quality Systems



Elements of an Ambient Air Monitoring Quality Systems

Ambient Air Monitoring Quality System

The framework by which orgs. apply sufficient QA/QC to ensure results meet expectations

Planning

- DQO Process
- QAPP

Implementing

- Monitoring
- SOPs

Assessing

- Verifications
- **Validations**
  - **WoE**
- Quality Assessment

Data validation *is a data review technique to ensure reported values meet quality goals*

*"**Weight of evidence**" is an expression used when discussing **data validation** that currently lack formal definitions in the CFR. It is an essential part of validation, and one the CFR specifically states that PQAOs and the EPA must use.* – paraphrasing EPA-454/B-21-007

40 CFR Part 58, Appendix A, Section 1.2.3

"Each PQAO is required to implement a quality system that provides sufficient information to assess the quality of the monitoring data. The quality system must, at a minimum, include the specific requirements described in this appendix. Failure to conduct or pass a required check or procedure, or a series of required checks or procedures, does not by itself invalidate data for regulatory decision making. Rather, PQAOs and the EPA shall use the checks and procedures required in this appendix in combination with other data quality information, reports, and similar documentation that demonstrate overall compliance with Part 58. Accordingly, the EPA and PQAOs shall use a "weight of evidence" approach when determining the suitability of data for regulatory decisions. The EPA reserves the authority to use or not use monitoring data submitted by a monitoring organization when making regulatory decisions based on the EPA's assessment of the quality of the data. Consensus built validation templates or validation criteria already approved in QAPPs should be used as the basis for the weight of evidence approach."

40 CFR Part 58, Appendix A, Section 1.2.3

"Each PQAO is required to implement a quality system that provides sufficient information to assess the quality of the monitoring data. The quality system must, at a minimum, include the specific requirements described in this appendix. ***Failure to conduct or pass a required check or procedure, or a series of required checks or procedures, does not by itself invalidate data*** *for regulatory decision making. Rather, PQAOs and the EPA shall use the checks and procedures required in this appendix in combination with other data d similar documentation that demonstrate . Accordingly, the EPA and PQAOs shall use a when determining the suitability of data for eserves the authority to use or not use a monitoring organization when making regulatory decisions based on the EPA's assessment of the quality of the data. Consensus built validation templates or validation criteria already approved in QAPPs should be used as the basis for the weight of evidence approach."*

In many cases, validity is not a simple "yes or no" decision.

40 CFR Part 58, Appendix A, Section 1.2.3

"Each PQAO is required to implement a quality system that provides sufficient information to assess the quality of the monitoring data. The quality system must, at a minimum, include the sp

appendix. Failure to conduct or pass

of required checks or procedures, do

regulatory decision making. Rather,

and procedures required in this appendix in combination with other data quality information, reports, and similar documentation that demonstrate overall compliance with Part 58. Accordingly, the EPA and PQAOs shall ***use a "weight of evidence" approach when determining the suitability of data for regulatory decisions****. The EPA reserves the authority to use or not use monitoring data submitted by a monitoring organization when making regulatory decisions based on the EPA's assessment of the quality of the data. Consensus built validation templates or validation criteria already approved in QAPPs should be used as the basis for the weight of evidence approach."*

CFR requires the use of WoE in data validity decisions.

40 CFR Part 58, Appendix A, Section 1.2.3

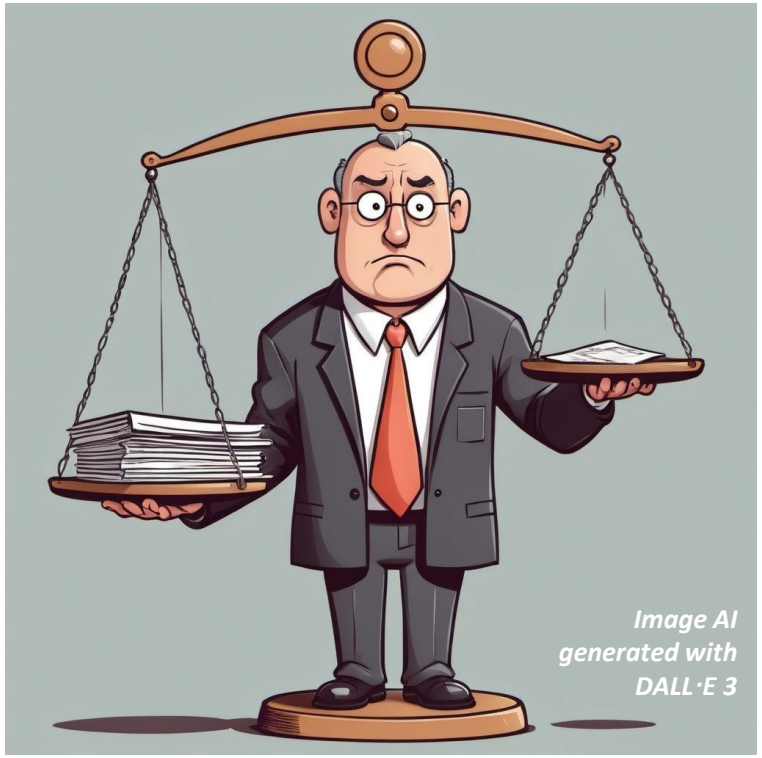
"Each PQAO is required to implement a quality system that provides sufficient information to assess the quality of the monitoring data. The quality system must, at a minimum, include the specific requirements described in this appendix. Failure to conduct or pass a required check or procedure, or a series of required checks or procedures, does not by itself invalidate data for

Validation templates, e.g., QA Handbook VII Appendix D, are the cornerstone to weight of evidence.

regulatory decisions. The EPA reserves the authority to use or not use monitoring data submitted by a monitoring organization when making regulatory decisions based on the EPA's assessment of the quality of the data. Consensus built ***validation templates or validation criteria already approved in QAPPs should be used as the basis for the weight of evidence approach****."*

The WoE Process

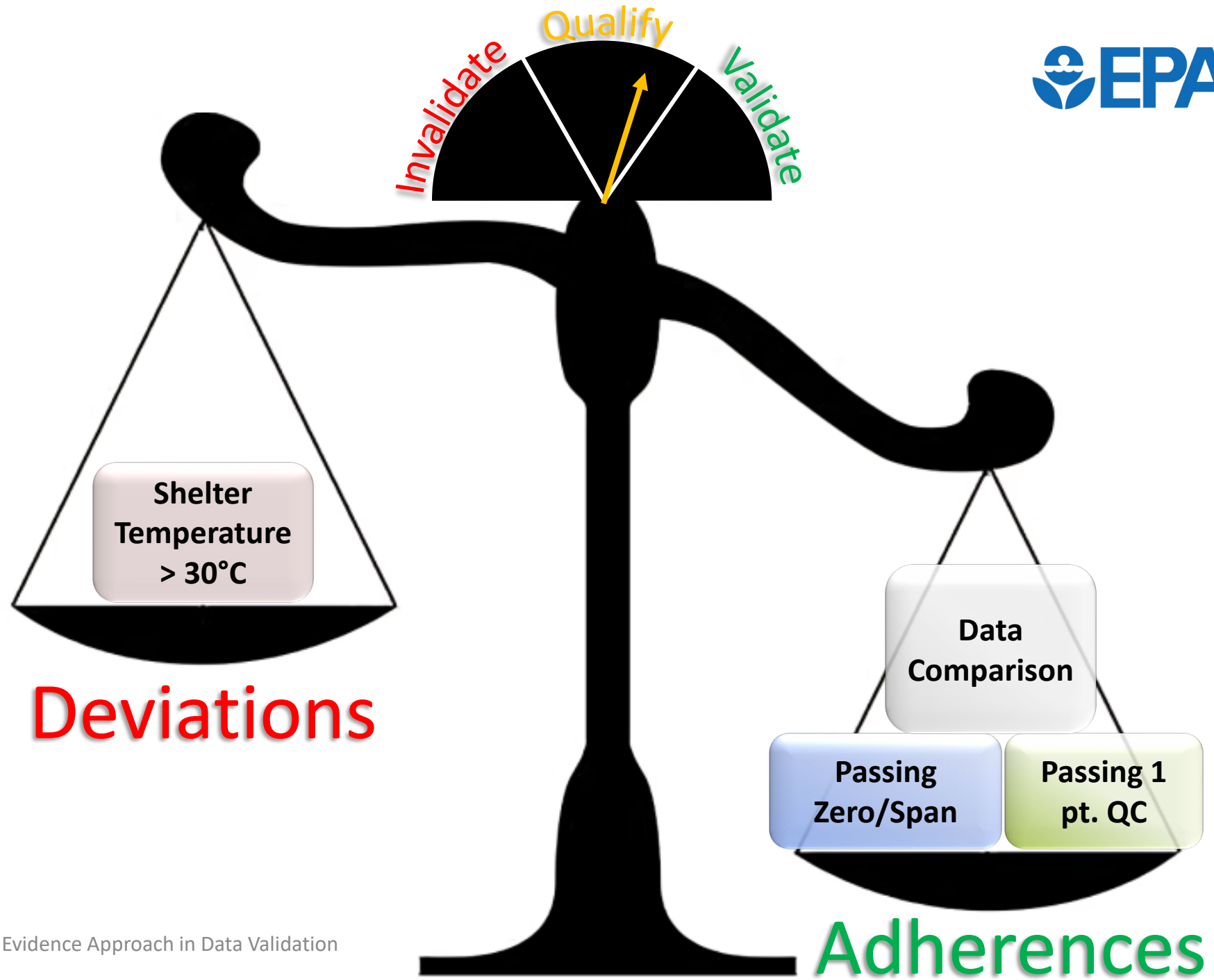
- All available information + professional judgment = decisions about data validity.
- Whether the evidence suggests the data cannot be used for its intended purpose outweighs the evidence available that suggests that it can, or vice versa.



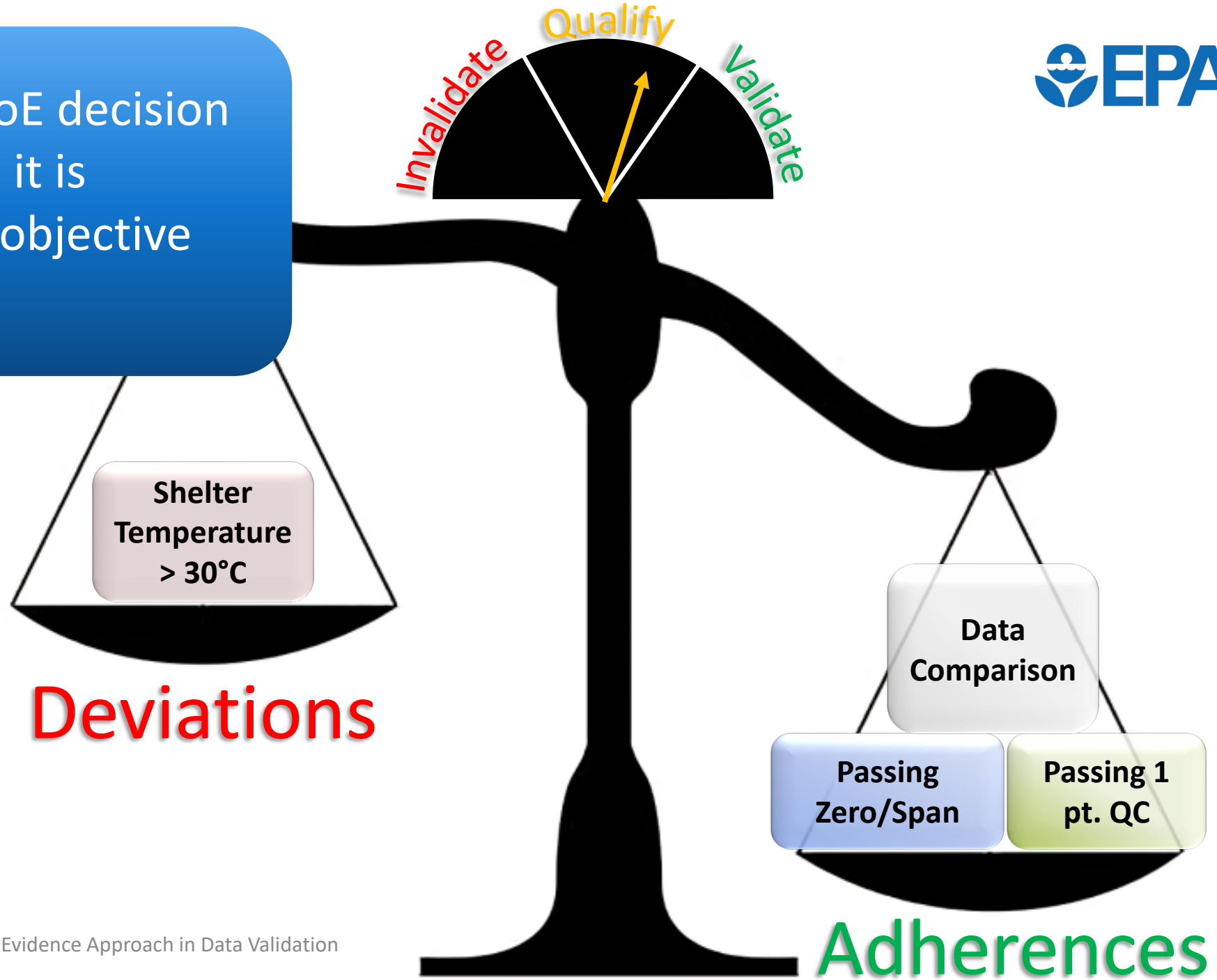
Evaluate
- Data
- Supporting documentation

Characterize
- # of Deviations
- Their implications

Logically determine if data are:
- Undefendable
  OR
- Defendable


Invalidate
Qualify
Validate
Shelter Temperature > 30°C
Deviations
Data Comparison
Passing Zero/Span
Passing 1 pt. QC
Adherences

While the WoE decision is subjective, it is informed by objective evidence



Compelling Evidence vs. WoE

- The two terms are often used interchangeably.
- Compelling evidence defined: data that concretely establishes instrument performance or the validity of a QA/QC check.
- Some *compelling evidence* may lead to either validate or invalidate a single QC check failure, which is then considered in the overall *WoE approach* to either validate or invalidate a sample.
- Compelling evidence **informs** the WoE decision.

Data Validation Templates

The foundation of the Weight of Evidence Approach

Introducing Data Validation Templates

- Validation templates:
  - Contain the MQOs for monitoring programs
  - Promotes consistency in data quality decision-making process
  - "…*should be used as the basis for the weight of evidence approach*" (40 CFR Part 58, App. A, Sect. 1.2.3)
- QA Handbook VII App. D: Criteria pollutant validation templates

CO Validation Template

1) Requirement (CO)	2) Frequency	3) Acceptance Criteria	Information /Action
CRITICAL CRITERIA-CO			
Sampler/Monitor	NA	*Meets requirements listed in FRM/FEM designation*	1) 40 CFR Part 58 App C Sec. 2.1 2) NA 3) 40 CFR Part 53 & FRM/FEM method list
One Point QC Check Single analyzer	*Every 14 days*	< ±10.1% (percent difference)	1 and 2) 40 CFR Part 58 App A Sec. 3.1.1 3) Recommendation based on DQO in 40 CFR Part 58 App A Sec. 2.3.1. QC Check Conc range 0.5 – 5 ppm
Zero/span check	Every 14 days	Zero drift < ± 0.41 ppm (24 hr) < ± 0.61 ppm (>24hr-14 day) Span drift < ± 10.1 %	1 and 2) QA Handbook Volume 2 Sec. 12.3 3) Recommendation
OPERATIONAL CRITERIA-CO			
			1, 2 and 3) QA Handbook Volume 2 Sec. 7.2.2

Validation Criteria Classes

- Three criteria (critical, operational, and systematic), each with different degree of implication about data quality.
  - How significantly a criterion impacts resulting concentration is basis for grouping.

Validation Criteria Classes

- Three criteria (**critical**, operational, and systematic), each with different degree of implication about data quality.
  - How significantly a criterion impacts resulting concentration is basis for grouping.
- Critical criteria:
  - Needed to maintain sample integrity.
  - Observations not meeting every critical criterion should be invalidated unless compelling reason and justification.
  - Typically identify distinct measurement(s) or time periods.

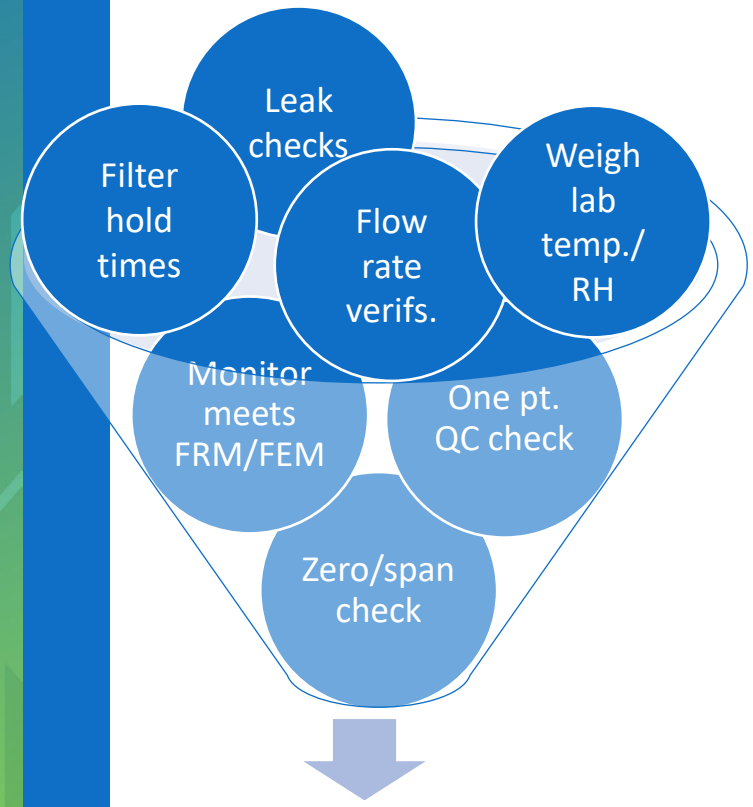
Validation Criteria Classes

- Three criteria (critical, **operational**, and systematic), each with different degree of implication about data quality.
  - How significantly a criterion impacts resulting concentration is basis for grouping.
- Operational criteria:
  - Important for maintaining and evaluating the quality of the data collection system.
  - Violation of 1 or more criterion may be cause for invalidation.
  - Should consider other quality control information that may or may not indicate the data are acceptable.
  - Samples are suspect unless other QC info demonstrates otherwise.

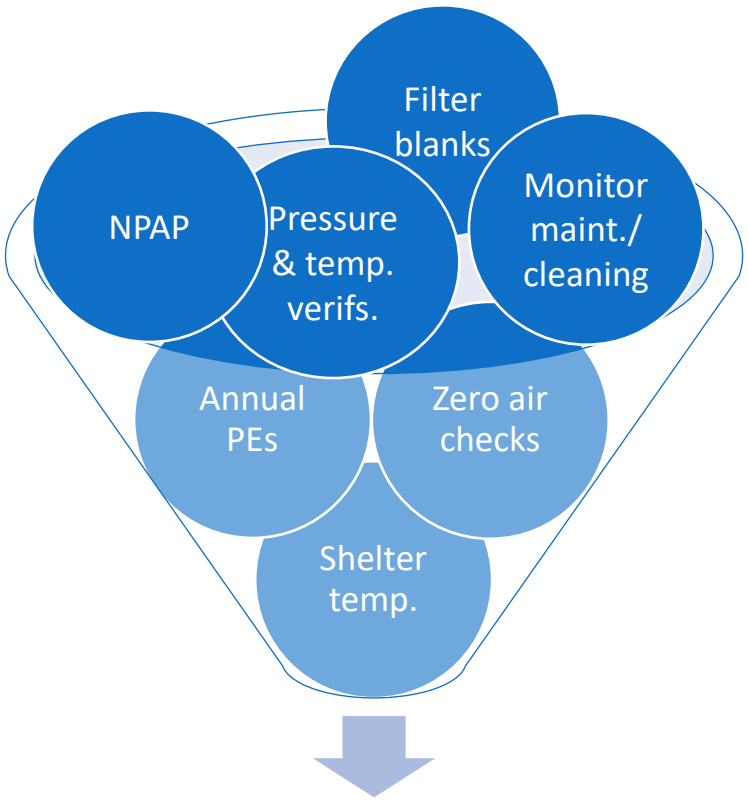
Validation Criteria Classes

- Three criteria (critical, operational, and **systematic**), each with different degree of implication about data quality.
  - How significantly a criterion impacts resulting concentration is basis for grouping.
- Systematic criteria:
  - Important for the correct interpretation of the data but do not usually impact the validity of a sample(s).
  - DQOs included as systematic criteria.

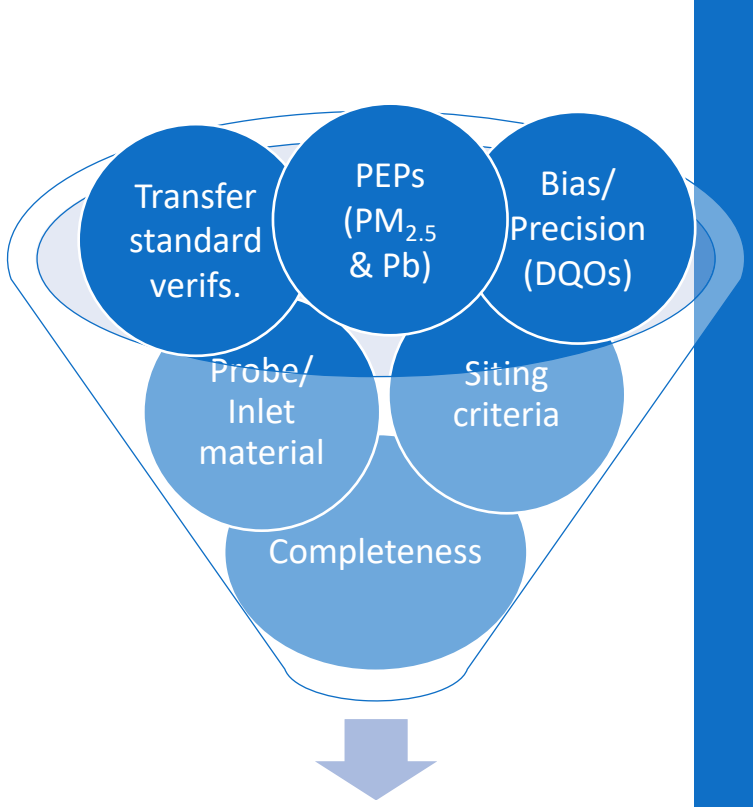
Criteria Examples, per Class



Critical



Operational



Systematic

Validation Criteria Classes

- How much ‘weight’ does each criteria class carry during WoE?

Validation Criteria Classes

- How much ‘weight’ does each criteria class carry during WoE?

Generally, violations of critical criteria result in data invalidation, whereas violations of operational/systematic criteria typically result in data qualification (flagging).

Validation Criteria Classes

- How much ‘weight’ does each criteria class carry during WoE?

Validation Criteria Classes

- How much ‘weight’ does each criteria class carry during WoE?

Validation Criteria Classes

- How much ‘weight’ does each criteria class carry during WoE?

A weight of evidence approach should be taken when assessing the data and the number/type of violations observed.

WoE Case Studies

Ex. 1: PM_{10} filter fingerprint

- **Scenario**: Site operator observes fingerprint on exposed 8x10" PM_{10} filter. Recommends invalidating.
- **Evidence**:
  - Lab measured PM_{10} concentration = 165 µg/m^3
  - Continuous PM_{10} monitor 10 miles away recorded 187 µg/m^3
  - Photo shows small imperfection on one of filter's corners
  - Site passed all QC checks surrounding event
- **Options**:
  - Invalidate
  - Validate w/ QA qualifier 'FX' (filter integrity issues)
  - Validate

Ex. 1: PM_{10} filter fingerprint

- **Scenario**: Site operator observes fingerprint on exposed 8x10" PM_{10} filter. Recommends invalidating.
- **Evidence**:
  - **Lab measured PM_{10} concentration = 165 µg/m³**
  - **Continuous PM_{10} monitor 10 miles away recorded 187 µg/m³**
  - Photo shows small imperfection on one of filter's corners
  - Site passed all QC checks surro
- **Options**:
  - Invalidate
  - Validate w/ QA qualifier 'FX' (
  - Validate

Evidence that the high conc. measured true to ambient conditions and not from contamination.

Ex. 1: PM_{10} filter fingerprint

- **Scenario**: Site operator observes fingerprint on exposed 8x10" PM_{10} filter. Recommends invalidating.
- **Evidence**:
  - Lab measured PM_{10} concentration = 165 µg/m^3
  - Continuous PM_{10} monitor 10 miles away recorded 187 µg/m^3
  - **Photo shows small imperfection on one of filter's corners**
  - Site passed all QC checks surrounding event
- **Options**:
  - Invalidate
  - Validate w/ QA qualifier 'FX' (filter integrity issues)
  - Validate

Evidence of the fingerprint; illustrates only minor impact to overall filter integrity.

Ex. 1: PM_{10} filter fingerprint

- **Scenario**: Site operator observes fingerprint on exposed 8x10" PM_{10} filter. Recommends invalidating.
- **Evidence**:
  - Lab measured PM_{10} concen
  - Continuous PM_{10} monitor 10 miles away recorded 187 µg/m³
  - Photo shows small imperfection on one of filter's corners
  - **Site passed all QC checks surrounding event**
- **Options**:
  - Invalidate
  - Validate w/ QA qualifier 'FX' (filter integrity issues)
  - Validate

No other QC concerns outside of the fingerprint.

Ex. 1: PM_{10} filter fingerprint

- **Scenario**: Site operator observes fingerprint on exposed 8x10" PM_{10} filter. Recommends invalidating.
- **Evidence**:
  - Lab measured PM
  - Continuous PM_{10} monitor
  - Photo shows small imperfection on one of the filter's edges
  - Site passed all QC checks surrounding event

As a data validator, what option would you choose and why?

- **Options**:
  - Invalidate
  - Validate w/ QA qualifier 'FX' (filter integrity issues)
  - Validate

Ex. 1: PM_{10} filter fingerprint

- **Scenario**: Site operator observes fingerprint on exposed 8x10" PM_{10} filter.
- **Evidence**:
  - Lab meas
  - Continuou
  - Photo sho
  - Site passed all QC checks surrounding event

Validator rationalized that while fingerprint can add/remove mass, impact on large 8x10" filter less significant. This, coupled with other compelling evidence/no other QC failures, validate with flag.

- **Options**:
  - Invalidate
  - **Validate w/ QA qualifier 'FX' (filter integrity issues)**
  - Validate

Ex. 2: Rural O_3 monitor exceeds standard

- **Scenario**: Rural O_3 monitor unexpectedly exceeds 8-hr standard; recent TSA uncovered some issues.
- **Evidence**:
  - Urban site 20 miles away with similar trends, but lower concentrations.
  - Site TSA week earlier found monitor in an office space and temperature controlled by office thermostat (not NIST-traceable/certified).
  - Instrument operated with cover removed.
  - Performance audit passed at all concs w/ avg. PD = 3%.
- **Options**:
  - Invalidate
  - Validate w/ QA qualifiers ‘1’ (Deviation from CFR/Critical Criteria Deviation), ‘2’ (Operational Deviation), and/or ‘3’ (Field Issue)
  - Validate

Ex. 2: Rural O_3 monitor exceeds standard

- **Scenario**: Rural O_3 monitor unexpectedly exceeds 8-hr standard; recent TSA uncovered some issues.
- **Evidence**:
  - **Urban site 20 miles away with similar trends, but lower concentrations.**
  - Site TSA week earlier found monitor in an office space and temperature controlled (
    traceable/certified).
  - Instrument operated w
  - Performance audit pas
- **Options**:
  - Invalidate
  - Validate w/ QA qualifiers '1' (Deviation from CFR/Critical Criteria Deviation), '2' (Operational Deviation), and/or '3' (Field Issue)
  - Validate

Similar concentration trends from a nearby monitor provides some confidence in the rural monitor's operation. However, doesn't answer potential high bias issue.

Ex. 2: Rural O_3 monitor exceeds standard

- **Scenario**: Rural O_3 monitor unexpectedly exceeds 8-hr standard; recent TSA uncovered some issues.
- **Evidence**:
  - Urban site 20 miles away with similar trends, but lower concentrations.
  - **Site TSA week earlier found monitor in an office space and temperature controlled by office thermostat (not NIST-traceable/certified).**
  - Instrument operated within expected range.
  - Performance audit passed.
- **Options**:
  - Invalidate
  - Validate w/ QA qualifier ('1' (Operational Deviation), '2' (Operational Deviation), and/or '3' (Field Issue))
  - Validate

CFR requires NIST-traceable temperature device. However, FEM allowable range for operation is 41-104 °F. Would an occupied office exceed this range?

Ex. 2: Rural O_3 monitor exceeds standard

- **Scenario**: Rural O_3 monitor unexpectedly exceeds 8-hr standard; recent TSA uncovered some issues.
- **Evidence**:
  - Urban site 20 mi
  - Site TSA week ea
    - **Instrument operated with cover removed.**
    - **Performance audit passed at all concs w/ avg. PD = 3%.**
- **Options**:
  - Invalidate
  - Validate w/ QA qualifiers '1' (Deviation from CFR/Critical Criteria Deviation), '2' (Operational Deviation), and/or '3' (Field Issue)
  - Validate

Excellent audit results (acceptance threshold set at <7.1%) indicate performance likely not impacted by lid removal.

Ex. 2: Rural O_3 monitor exceeds standard

- **Scenario**: Rural O_3 monitor unexpectedly exceeds 8-hr standard; recent TSA uncovered some issues.
- **Evidence**:
  - Urban site 20 miles away with similar trends but lower concentrations.
  - Site TSA: work surface temperature controlled by office thermostat (not NIST traceable/certified)
  - Instrument operated with cover removed.
  - Performance audit passed at all concs w/ avg. PD = 3%.

As a data validator, what option would you choose and why?

- **Options**:
  - Invalidate
  - Validate w/ QA qualifiers ‘1’ (Deviation from CFR/Critical Criteria Deviation), ‘2’ (Operational Deviation), and/or ‘3’ (Field Issue)
  - Validate

Ex. 2: Rural O_3 monitor exceeds standard

- **Scenario**: Rural O_3 monitor unexpectedly exceeds 8-hr standard; recent TSA uncovered some issues.
- **Evi**
  - Validator rationalized that temps almost certainly w/in appropriate range, passing audit and similar nearby concs. provides evidence of proper monitor function, audit results also indicate lid missing not impacting operation. Validate w/ flags 2 and 3.
  - Performance audit passed at all concs w/ avg. PD = 3%.
- **Options**:
  - Invalidate
  - **Validate w/ QA qualifiers** '1' (Deviation from CFR/Critical Criteria Deviation), **'2' (Operational Deviation), and/or '3' (Field Issue)**
  - Validate

Ex. 3: Issues at a $PM_{2.5}$ Weigh Lab

- **Scenario**: TSA of 2-month-old weigh lab indicates multiple "operational criteria" non-conformances.
- **Evidence**:
  - All field parameter QC checks met over 2-month period.
  - Microbalance had no calibration/certification documentation.
  - Microbalance not grounded.
  - Lab blanks out of spec (ranged from -477 to +98 µg).
  - RH/temp datalogger doesn't meet accuracy specifications.
- **Options**:
  - Invalidate all samples w/in 2-month timeframe
  - Validate w/ QA qualifiers '2' (Operational Deviation), and/or 'LB' (Lab Blank Value Above Acceptable Limit)
  - Validate all samples w/in 2-month timeframe

Ex. 3: Issues at a $PM_{2.5}$ Weigh Lab

- **Scenario**: TSA of 2-month-old weigh lab indicates multiple "operational criteria" non-conformances.
- **Evidence**:
  - **All field parameter QC checks met over 2-month period.**
  - Microbalance had no calibration/certification documentation.
  - Microbalance not grounded.
  - Lab blanks out of spec (ranged f
  - RH/temp datalogger doesn't m
- **Options**:
  - Invalidate all samples w/in 2-month timeframe
  - Validate w/ QA qualifiers '2' (Operational Deviation), and/or 'LB' (Lab Blank Value Above Acceptable Limit)
  - Validate all samples w/in 2-month timeframe

No observed field issues to consider.

Ex. 3: Issues at a $PM_{2.5}$ Weigh Lab

- **Scenario**: TSA of 2-month-old weigh lab indicates multiple "operational criteria" non-conformances.
- **Evidence**:
  - All field parameter QC checks met over 2-month period.
  - **Microbalance had no calibration/certification documentation.**
  - Microbalance not grounded.
  - Lab blanks out of spec (ranged
  - RH/temp datalogger doesn't m
- **Options**:
  - Invalidate all samples w/in 2-month timeframe
  - Validate w/ QA qualifiers '2' (Operational Deviation), and/or 'LB' (Lab Blank Value Above Acceptable Limit)
  - Validate all samples w/in 2-month timeframe

Microbalance lacks NIST-traceability

Ex. 3: Issues at a $PM_{2.5}$ Weigh Lab

- **Scenario**: TSA of 2-month-old weigh lab indicates multiple "operational criteria" non-conformances.
- **Evidence**:
  - All field parameter QC checks met over 2-month period.
  - Microbalance had no calibration/certification documentation.
  - **Microbalance not grounded.**
  - **Lab blanks out of spec (ranged from -477 to +98 µg).**
  - RH/temp datalogger doesn't meet accuracy specifications
- **Options**:
  - Invalidate all samples w/in
  - Validate w/ QA qualifiers
    Blank Value Above Acceptable Limit)
  - Validate all samples w/in 2-month timeframe

No ground to balance and severe LB swings indicates potential static electricity impacting weighings

Ex. 3: Issues at a $PM_{2.5}$ Weigh Lab

- **Scenario**: TSA of 2-month-old weigh lab indicates multiple "operational criteria" non-conformances.
- **Evidence**:
  - All field parameter QC checks m
  - Microbalance had no calibration
  - Microbalance not grounded.
  - Lab blanks out of spec (ranged from -477 to +98 µg).
  - **RH/temp datalogger doesn't meet accuracy specifications.**
- **Options**:
  - Invalidate all samples w/in 2-month timeframe
  - Validate w/ QA qualifiers '2' (Operational Deviation), and/or 'LB' (Lab Blank Value Above Acceptable Limit)
  - Validate all samples w/in 2-month timeframe

Adds suspicion to filter conditioning environment meeting requirements.

Ex. 3: Issues at a $PM_{2.5}$ Weigh Lab

- **Scenario**: TSA of 2-month-old weigh lab indicates multiple "operational criteria" non-conformances.
- **Evidence**:
  - All fie
  - Micro
  - Micro
  - Lab blanks out of spec (ranged from -477 to +98 µg).
  - RH/temp datalogger doesn't meet accuracy specifications.

As a data validator, what option would you choose and why?

- **Options**:
  - Invalidate all samples w/in 2-month timeframe
  - Validate w/ QA qualifiers '2' (Operational Deviation), and/or 'LB' (Lab Blank Value Above Acceptable Limit)
  - Validate all samples w/in 2-month timeframe

Ex. 3: Issues at a $PM_{2.5}$ Weigh Lab

- **Scenario**: TSA of 2-month-old weigh lab indicates multiple "operational criteria" non-conformances
- **Evide**
  - All
  - Mi
  - Mi
  - Lab
  - RH/temp datalogger doesn't meet accuracy specifications.

Simply cannot trust the microbalance readings given the compounding evidence. With 4+ flags, even though operational, validator invalidated all samples in which either pre- or post-sampling weighings occurred in the 2-month window.

- **Options**:
  - **Invalidate all samples w/in 2-month timeframe**
  - Validate w/ QA qualifiers '2' (Operational Deviation), and/or 'LB' (Lab Blank Value Above Acceptable Limit)
  - Validate all samples w/in 2-month timeframe

References

- Best Practices for Review and Validation of Ambient Air Monitoring Data. August 2021 (EPA-454/B-21-007)
- QA Handbook Volume II, Appendix D: Measurement Quality Objectives and Validation Templates. Jan. 2017 (EPA-454/B-17-001)
- Steps to Qualify or Validate Data After an Exceedance of Critical Criteria Checks. Jan. 2022 EPA technical memo
- 40 CFR Part 58, Appendix A
- AQS Qualifiers: https://aqs.epa.gov/aqsweb/documents/codetables/qualifiers.html

Contact Information:

Brannon Seay
Office of Air Quality Planning and Standards
seay.brannon@epa.gov

Backup slides

Compelling Evidence Consideration for a Critical Criteria Failure