Tools to Calculate and Evaluate Measurement Performance

Marlene Moore
Advanced Systems, Inc.
mmoore@advancedsys.com

Objective

- Review methods for calculation and evaluation of measurement performance
- Topics
  - common terms for data quality indicators
  - assess measurement performance within batches, between batches and for reported results
  - formulas and calculations used
  - methods for evaluating data quality indicators
- Calculate and evaluate data

Agenda

- Data Quality Indicators (DQI)
  - What are DQI's
- Calculations
  - Methods of Calculation
- Evaluations
  - Methods of Evaluation
- Questions/Answers
  - Summary, Evaluation Forms

What are DQI's?

- Precision
- Bias
- Accuracy
- Representativeness
- Comparability
- Completeness
- Sensitivity

Ways of monitoring and expressing measurement imperfections

EPA QA/G-5 2002

Measurement Imperfections

- Random Effects (Precision)
  - Analyst
  - Environmental factors
  - Short term fluctuations
- Systematic Effects (Bias/Accuracy)
  - Calibration drifts
  - Personal bias in reading instrument
  - Uncertainty value of a reference standard
- Blunders (OOPs)
  - Quality system
    - Quality assurance
      - Quality control

Precision

- The measure of agreement among repeated measurements of the same property under identical, or substantially, similar conditions; calculated as either the range or as the standard deviation
- May also be expressed as a percentage of the mean of the measurements, such as relative range or relative standard deviation (coefficient of variation)

EPA QA/G-5 2002

Precision

- Use the same analytical instrument to make repeated analyses on the same sample. *(Instrument Duplicate)*
- Use the same method to make repeated measurements of the same sample within a single laboratory *(Lab Duplicate)* or have two or more laboratories analyze identical samples with the same method. *(Split Sample)*
- Split a sample in the field and submit both for sample handling, preservation and storage, and analytical measurements. *(Field Duplicate)*
- Collect, process, and analyze collocated samples for information on sample acquisition, handling, shipping, storage, preparation, and analytical processes and measurements. *(Collocated Sample)*

EPA QA/G-5 2002

Bias

- The systematic or persistent distortion of a measurement process that causes errors in one direction.
- Use reference materials for method performance by lab operations
- Use spiked matrix samples for method performance in sample

EPA QA/G-5 2002

Accuracy

- Measure of the overall agreement of a measurement to a known value; includes a combination of random error (precision) and systematic error (bias) components of both sampling and analytical operations.
- Analyze a reference material or reanalyze a sample to which a material of known concentration or amount of pollutant has been added.
- Usually expressed either as percent recovery or as a percent bias.

EPA QA/G-5 2002

Representativeness

- Qualitative term that expresses the degree to which data accurately and precisely represent a characteristic of a population, parameter variations at a sampling point, a process condition, or an environmental condition.
- Evaluate whether measurements are made and physical samples collected in such a manner that the resulting data appropriately reflect the environment or condition being measured or studied.

EPA QA/G-5 2002

Comparability

- Qualitative term that expresses the measure of confidence that one data set can be compared to another and can be combined for the decision(s).
- Compare sample collection and handling methods, sample preparation and analytical procedures, holding times, stability issues, and QA protocols.
- Not calculated

EPA QA/G-5 2002

Completeness

- Measure of the amount of valid data needed to be obtained from a measurement system.
- Compare the number of valid measurements completed (samples collected or samples analyzed) with those established by the project's quality criteria (performance/ acceptance criteria).
- Expressed as percentage.

EPA QA/G-5 2002

Sensitivity

- Capability of a method or instrument to discriminate between measurement responses representing different levels of the variable of interest.
- Determine the minimum concentration or attribute that can be measured by a method (detection limit), by an instrument (instrument detection limit), or by a laboratory (quantitation limit).
- Compare the appropriate limit to the action level established during project planning.

EPA QA/G-5 2002

Laboratory Performance

- Contamination Monitors
  - Blanks
- Response Monitors
  - Control Samples
  - Limit of Detection
  - Limit of Quantitation
  - Critical Value
- Other
  - Repeatability
  - Reproducibly
  - Uncertainty

Blanks

- Types
  - Method
  - Reagent
  - Instrument
- Calculations
  - Depends on Technology
  - Method
  - Manufacturer's Program

Control Samples

- LCS, LOD, LOQ, CV
- MS, MSD
- Surrogates
- Internal Standards
- External Standards
- Second Source
- Technology specific control samples

Other

- Repeatability (Precision)
  - Same person, instrument, sample, lab, etc
  - Different person, instrument, sample, lab, etc.
- Reproducibility (Bias/Accuracy)
  - Same person, instrument, sample, lab, etc
  - Different person, instrument, sample, lab, etc.
  - Same true value - National standard (?)

Sources of Imprecision

- Instrumental instability
- Environmental fluctuations
- Operator skill
- Reagent control
- Variability of blank, sample
- Variable contamination, losses
- Faulty technique
- Maintenance of tolerances

Sources of Bias

- Calibration
- Operator bias
- Uncorrected blank
- Inefficiencies or losses
- Tolerances adjustments
- Interference resolution
- Contamination gains
- Instrumental shifts
- Matrix effects
- Theoretical

Random Effects

$$s(\bar{q}) = \frac{s(q_k)}{\sqrt{n}}$$

Standard Deviation- mean

$$u(x_i) = s(\bar{q})$$

Standard Uncertainty

Adding Imperfections

- Combining Standard Uncertainties
- Must be in terms of the measured quantity
  - assume linear summation
  - assume uncorrelated input quantities
  - sensitivity coefficient = 1 = (ci) = $\partial f / \partial x$
    - (partial derivative)

$$u_i(y) = c_i u(x_i)$$

Combined Components

- Square root of the sum of the squares of the separate standard uncertainty components (Root sum square method)
  - Must all be in same units
  - Often done using relative standard uncertainty values
- Combined uncertainty * coverage factor = Expanded Uncertainty
- Coverage factor = 2 for 95% confidence level

Combined Standard Uncertainty

$$u_c(y) = \sqrt{\sum_{i=1}^{N} u_i^2(y)}$$

(y)= estimate of the measurand Y, output estimate

Expanded Uncertainty

$$U = k\, u_c(y)$$

$u_c(y)$= Combined Standard Uncertainty

k = Coverage Factor

Estimation of Uncertainty

$$\overline{X} \pm U$$ 95% confidence level, k- 2

Where,

$\overline{X}$ = the mean of *n* measurements
U = the uncertainty
k = the coverage factor for a confidence level of approximately 95%

Uncertainty Procedure

- Specify the measurand
- Determine measurement conditions
- Identify Uncertainty sources
- Consolidate Uncertainty components
- Quantify Uncertainty components
- Convert components to Standard Uncertainty
- Calculate Combined Uncertainty and Expanded Uncertainty

www.a2la.org
www.navylabs.navy.mil

Methods of Calculation

- Precision
- Bias/Accuracy
- Demonstration of Capability
- Detection Limits
- Control Limits

Precision

- Range

$$\overline{R} = \frac{R_1 + R_2 + \ldots + R_n}{n}$$

- Standard Deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

x = each value

$\bar{x}$ = mean of values

n = # of values

Precision

- Relative Standard Deviation

$$RSD = 100 * \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \Big/ \bar{x}$$

%RSD

- Coefficient of Variation

$$CV = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \Big/ \bar{x}$$

Let's Practice

$$? = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$? = \frac{|x_1 - x_2|}{(x_1 + x_2)/2} * 100$$

Bias/Accuracy

- Percent Recovery or Percent Bias

$$\%R = \frac{|x_1 - x_T|}{x_T} * 100$$

x_1 = Measured value

x_T = Known value

Demonstration of Capability

- QC Sample Outside Source
- Four aliquots
  - concentration 1-4 times the LOQ
- Same day or over several days
- All results to be used
- Compare mean recovery and standard deviation to method or lab acceptance criteria to evaluate performance

Appendix C NELAC 2003

LOD

- MDL - 40CFR 136 or Lab Defined
- Verification required if results reported to LOD
  - Initial and Annual
  - All steps of method
  - 2-3 times the reported LOD - must be measured
  - Each instrument

Appendix C NELAC 2003

LOQ

- Reporting Limit or Quantitation Limit
  - Lab procedure defines how LOQ determined
  - Reporting results only to LOQ
  - Must have procedure to define relation of LOD with LOQ
  - Confirm validity (Annual if not done for LOD)
    - Spike 1-2 times claimed LOQ
    - Recovery must be within method or client or lab acceptance criteria

Appendix C NELAC 2003

Critical Value

- Critical Value (Lc): the statistically estimated measured value of a substance that has less than a 1% probability of being exceeded in a method blank.
- False Negative Quality Sample (FNQS): a low level laboratory control QC sample (LCS) used to establish the recovery and reliability of a measurement near the detection limit. The concentration of the FNQS is set at 2-5 times Lc but preferably no more than 2 times the regulatory reporting limit, if one exists.

Rev 8, March 25, 2007
FACA website

Critical Value

- FNQS Recovery: percent recovery of the FNQS result calculated as 100*(Measured Value)/(True Value) and includes both random and systematic error.
- The relative standard deviation (RSD) at the FNQS should fall between the limits of 10% and 25%.
- The average FNQS Recovery when N> 20 provides an estimate of measurement bias at the FNQS concentration. Individual FNQS recoveries should fall within the limits of 50%-150%.

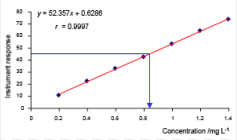
Rev 8, March 25, 2007
FACA website

Calibration Reference

- ***Preparation of Calibration Curves, A Guide to Best Practices,*** September 2003
  LGC/VAM/2003/032

Pitfalls during Studies

- Common pitfalls were identified
- Linear Systems
  - Response to concentration



Pitfalls

- Concentration range is not adequate to cover the range of sample concentrations
- Calibration standards concentration are not evenly spaced across the calibration range
- Uncertainty associated with the calibration standards concentration is too large due to preparation practices or standards purity
- Wrong regression formula is applied

Pitfalls

- Calibration line is fitted through zero even though the intercept is not zero
- Instrument software is used to carry out the regression without looking at the plot of the data
- Full standard error of prediction calculation is not performed
- Performance of the instrument is not within specification

Avoid Problems By:

- Plan the calibration study
- Analyze a standard with zero analyte concentration (i.e., method blank or calibration blank)
- Use appropriate materials and apparatus for preparation of calibration standards
- Define the equipment specifications appropriate for the calibration

Avoid Problems By:

- Plot and examine calibration results and specify the acceptable residual;
- Use validated software to perform the linear regression;
- Define when to set the intercept to zero (e.g. when is the intercept and zero insignificant);
- Calculate the uncertainty for test sample concentrations from the calibration curve as one component of the estimated uncertainty.

Plan the Study

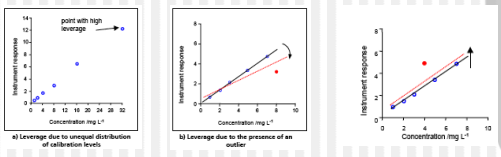
- Number of calibration standards
- Concentration of each standard
- Number of replicates of each measurement
- Preparation of calibration standards

Making Measurements

- Equipment Qualification
  - Fit for purpose
- Standards in a random order
  - Not decreasing or increasing

Plotting the Results

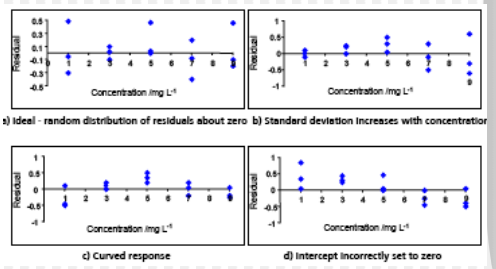
- View in a plot
- Evaluate the scatter plot
- Points of influence
  - Leverage or bias



Regression Analysis

- y = mx+b
  - x = concentration
  - y = response
- Residuals
  - y measured versus y calculated
  - How well the line fits the data
  - Least squares regression - minimize the sum of squared residuals
  - Normal distribution, equal weight to all points
    - Standard deviation is the same across all points

Analyze Residual Plots

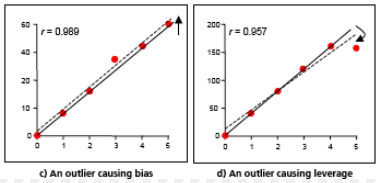


a) Ideal - random distribution of residuals about zero b) Standard deviation increases with concentration

c) Curved response d) Intercept incorrectly set to zero

Correlation Coefficient, r

- Measure of correlation not linearity
- Closer to 1 the better the correlation



c) An outlier causing bias d) An outlier causing leverage

Residual Standard Deviation

- Deviation of data from fitted regression line
- ANOVA table to assess regression
- Fitting line through origin

$$s(r) = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}}$$

Regression Analysis

Regression Statistics	
Multiple R	0.999955883
R Square	0.999911768
Adjusted R Square	0.999889709
Standard Error	0.005164622
Observations	6

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1.2091	1.2091	45330.79	2.93×10^{-9}
Residual	4	0.00010669	2.67×10^{-5}		
Total	5	1.2092			

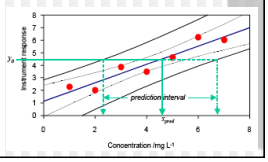
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.0021129	0.0037548	0.56270	0.60368	-0.008312	0.012538
X Variable 1	0.10441	0.00049038	212.91	2.92×10^{-9}	0.10304	0.10577

Estimate Sample Values

- Satisfactory regression analysis
- m and y used to calculate test sample results
- Samples analyzed multiple times
- Same conditions as standards

Estimate Uncertainty

- Confidence interval for regression line
- Less certain near extremes
- Prediction interval calculation provides estimate of uncertainty associated with predicted values of x



Calibration Today

- Robust statistical techniques not always used in the method validation when test developed
- Laboratory method validation data does not present study of calibration process to support laboratory method

Method Validation

- Reference test methods must define
  - specific conditions of test or define acceptable statistical parameters of performance
  - selection of number of data points
  - calibration range
  - origin
  - uncertainty
  - number of standard and sample runs to reduce standard error of prediction

Methods of Evaluation

- Sampling QC
  - Equipment Blanks
  - Field Blanks
  - Trip Blanks
  - Cooler Temperature
  - Field Duplicate Pairs
  - Collocated Samples
  - Field Splits

Methods of Evaluation

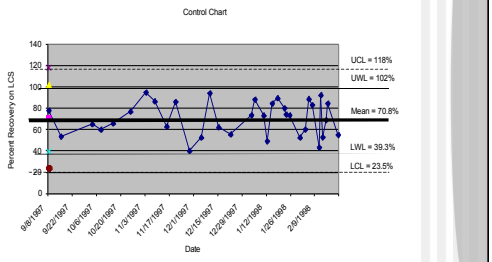
- Analytical QC
  - Method Blank (MB)
  - Laboratory Control Sample (LCS)
  - Reagent Blank
  - Storage Blank
  - Internal Standards
  - Second Source Standard
  - Proficiency Testing (PT)
  - Surrogates
  - Lab Duplicate
  - Lab Matrix Spike
  - Matrix Spike Duplicate

Control Charts

- Mean Chart
  - Standard Deviation, Average
  - Warning Limits ($\pm 2s$), Control Limits ($\pm 3s$)
  - Number of points for calculating s
    - depends on use
- Range Chart
  - Standard Deviation, Average
  - Upper WL and CL only
  - Baseline = zero
  - Factor used for duplicates =
    - 1.128 mean
    - 3.267 control limit

SM 20th edition

Control Chart



Control Charts

- Evaluation *SM 20th edition*
  - 1 measurement > CL
    - Analyze another
    - Stop test if > CL
  - 2 of 3 successive point > WL
    - Analyze another
    - Stop test if > WL evaluate bias and correct
  - 4 out of 5 points exceeds 1*s or* decreasing or increasing order on same side of the central line
    - Analyze another
    - Stop test if exceeds 1s or same pattern and correct
  - 7 successive points on same side of central line
    - Stop test and correct

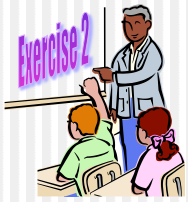
Significant Figures

- Digits known definitively, except for last digit. - last digit is in doubt
- Do not carry more than one doubtful digit in the final result
- Use standard deviation (or uncertainty) as guide for number of significant figures.
- Round only at the end of all operations
- How do you deal with zero?

SM 20th edition

Let's Practice

0.0007
12.065
4.078
25.9
4885
4927.0437



$35.9802824 = \left(56 * 0.025235 * 58.56 / 2.3\right)$

Rules for Rounding

- Drop off digits that are not significant
  - If digit > 5 increase preceding digit
  - If digit < 5 do not alter preceding digit
  - If digit = 5, round preceding digit to even number
- Practice: (Round to two sig figs)

2.77 = ?	2.30= ?	2.25 = ?
2.73 = ?	2.35 = ?	2.89 = ?

SM 20th edition

Any Questions?

Summary

- Place Information in your Quality Manual
  - **Definition of Terms**
  - **Formulas**
  - **Procedures for Data Evaluation**

Thank you

Marlene Moore
Advanced Systems, Inc.
302 368 1211
mmoore@advancedsys.com