Evaluating a Method: *what do we look for ?*

James Bartos

Office of Indiana State Chemist

Objectives:

- *Terms & Definition requirements are: 1) Term, 2) Efficacy Data, and 3) Laboratory Method*
- *Identify what the AAPFCO Laboratory Services Committee looks for?*
- *Is the process uniform?*
- *Do we have a justifiable and defensible reason(s) for rejecting a method?*
- *Do we have documentation we can share with the company?*
- *Can we let company know what additional information they need to obtain/provide?*

Method Development vs. Method Validation?

- ***Method Development:***
  - *Thoughtful, but somewhat arbitrary process*
  - *Generally in early, experimental, theoretical stage*
  - *Changes and improvements are often and expected*
  - *Not quite ready for "prime time" as yet unproven*
- ***Method Validation:***
  - *Highly structured process*
  - *More mature method stage*
  - *Further changes are not expected (or allowed)*
  - *AOAC, ISO, ASTM, etc. type methods*
  - *Preferred AAPFCO process*

Criteria for Method Validation

- ***What we are looking for / at:***
  - ***Validation Materials and Preparation***
  - ***Precision***
  - ***Accuracy***
  - ***Selectivity***
  - ***Ruggedness***
  - ***Calibration***
  - ***Limit of Detection / Limit of Quantification***
  - ***Final report***

Validation Materials & Preparation

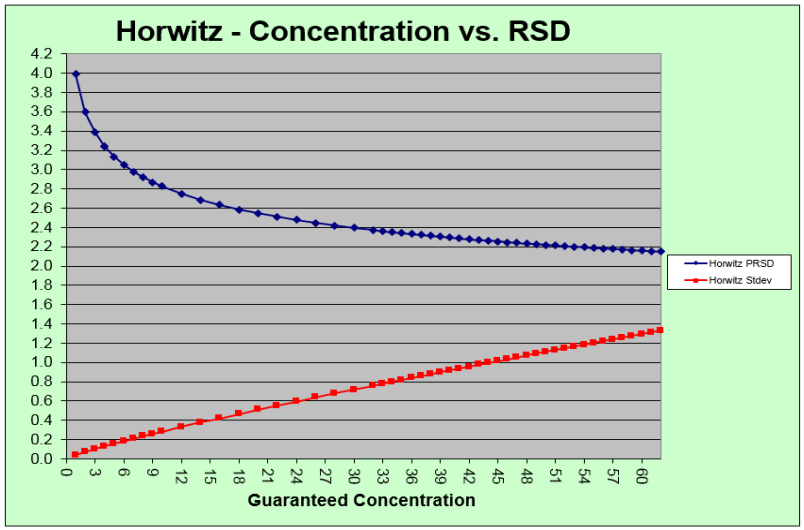
- *a.k.a. Test Samples*
- *Must include a number of different materials, matrices and concentrations*
  - *Low, medium and high concentrations, liquids, solids, slurries, inorganic, organic, blends, different products and companies, etc.*
  - *Common flaw – only tested your product(s)*
- *How were samples collected & prepared*
  - *Representation*
  - *Uniformity/homogeneity – may have to prove*
  - *Want to attribute data variability to the method and not to a poor or non-uniform sample*

Precision

- *The ability to produce the same result every time*
  - *Note: not necessarily the "true" result*
- *Why the "first" data evaluation criteria?*
  1. *Obvious – don't want inconsistent results*
  2. *Less obvious – generates much of the statistics*
     - *Mean, standard deviation, RSD/CV*
     - *Horwitz ratios – compare variability to other studies*
     - *Removal of outlier data/labs*
- *Magruder check sample program*
  - *Determining how well we agree with each other*
- *May be limitation of the method that prevents it from obtaining the true result*

Precision (continued)

- ***Repeatability (Individual Lab)***
  - ***How well a single individual is able to produce similar results***
- ***Reproducibility (All Labs)***
  - ***How well multiple individuals are able to produce similar results***
- ***Horwitz – variability is function of concentration***



Accuracy *(Trueness)*

- ***The ability to produce the true result***
- ***Sometimes a method limitation(s) prevents the recovery of true result, but universally accepted***
  - ***Total vs. acid-soluble metals (e.g. SUIP-25)***
  - ***Total vs. Kjeldahl Nitrogen***
- ***Use Certified Reference Materials, pure chemical reagents, consensus materials, etc. to evaluate if the method recovers true/consensus result***
- ***Associated with the term "bias"***
  - ***e.g. – recoveries of 90%, 96%, 94% and 92% would suggest a slightly low method bias***
  - ***Recoveries of 97%, 101%, 98%, 103% would suggest no particular method bias***

Selectivity *(Specificity)*

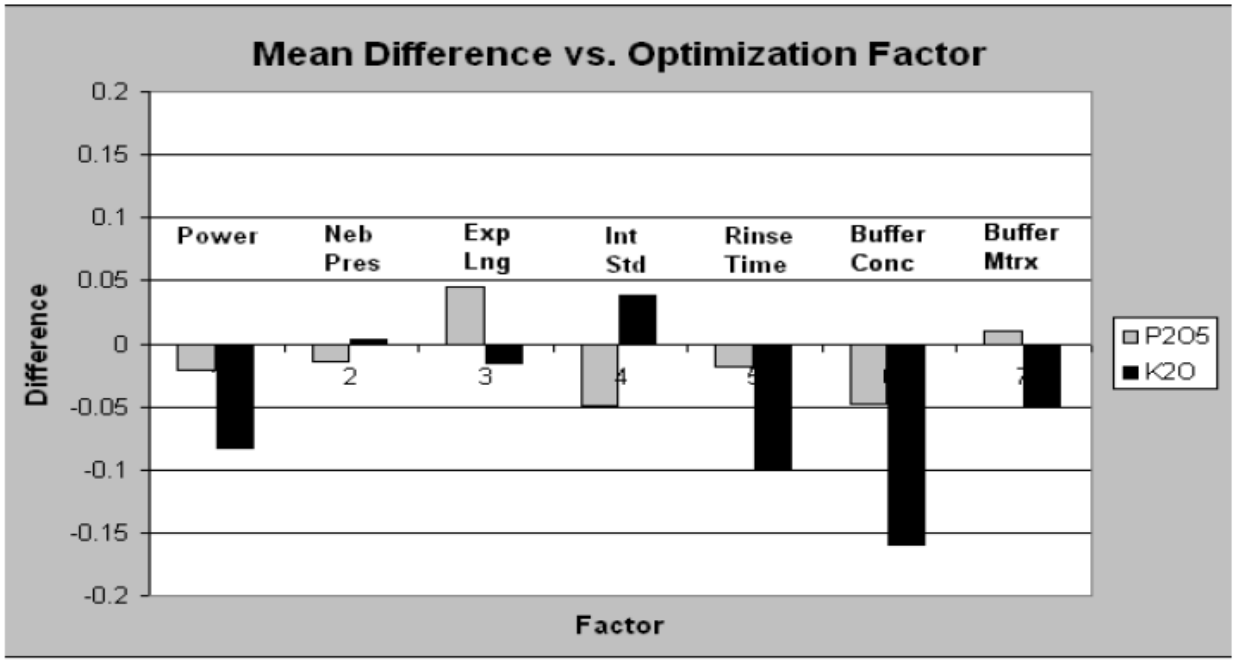
- *"The degree to which the method can quantify the target analyte"*
- *Examples:*
  - *PO4 (phosphate) vs. PO3 (phosphite)*
  - *Cadmium at wavelength 226.502 nm vs. Iron at 226.505*
- *False positive – presence indicated, when it doesn't exist (e.g. above)*
- *False negative – no presence indicated, when it does exist (e.g. sulfate and elemental, but not organic S)*
- *Spike samples and matrices with standards; test many different products; history & experience*
  - *% Recoveries*
- *List limitations in method Scope*

Ruggedness *(Robustness)*

- ***Ability of method to tolerate minor (unintended) deviations from the required conditions***
- ***Different operators, different instruments, slight differences in temperatures/thermometers, instrument condition, day-to-day, etc.***
- ***Evaluate by intentionally deviating slightly from the prescribed conditions and observing data***
  - ***Options:***
    - ***Pick a setting in the "robust" middle***
    - ***May see a tolerance:***
      - ***1.00 g +/- 0.10***
      - ***65^{o}C +/- 3***
    - ***If no tolerance, then clearly state***

Ruggedness *(examples)*

#	Factor	Minor Value	Major Value	"Final"
1	Power	A = 1.40 kW	a = 1.50 kW	a = 1.45 kW
2	Nebulizer Pressure	B = 0.7 L/min	b = 0.8 L/min	b = 0.7 L/min
3	Exposure Length	C = 10 sec	c = 20 sec	c = 10 sec
4	Internal Standard	D = Sc	d = Be	d = Sc
5	Rinse Time	E = 30 sec	e = 40 sec	e = 35 sec
6	CsCl Buffer Strength	F = 0.018M	f = 0.024M	f = 0.018M
7	Buffer Matrix	G = 4% Nitric Acid	g = water	g = 2% Nitric



Calibration:

- *Number of calibration points (at least 5)*
- *Even distribution of standards*
- *Curve fit (linear or "other")*
- *Required correlation coefficient (e.g. $R^2 > 0.997$)*
- *Avoid manipulations*
  - *Weighting or forcing through particular points*
- *Source(s) of stock standards*
- *Recipes – how to make*
- *Special handling – refrigerate, avoid light, etc.*
- *Expiration dates*
- *These are "knowns" / samples are "unknowns"*
  - *If don't get standards right, won't get samples right*

Limits of Detection & Quantification

- *An easy way and a hard way to do this*
- *Hard way – trial-and-error until close in on values*
- *Easy way (most common) – estimate from a blank or sample containing no analyte of interest*
- *Measure blank at least 7 times, determine the average and 3 standard deviations (SD).*
- *At 3SD, you have a 99% statistical certainty that you detected something different from zero, so = the Limit of Detection (LOD; not a reliable result)*
- *10SD = Limit of Quantification (LOQ)*
  - *(arbitrary) point at which the data is becoming reliable and reportable*

Final Report

- ***Include:***
  - ***Standard Operating Procedure (SOP)***
  - ***Validation materials used***
  - ***Precision***
    - ***Repeatability, reproducibility (if possible), prefer different days, different instruments, different operators, which makes the data more "real-world"***
    - ***Horwitz repeatability and reproducibility ratios***
  - ***Accuracy***
  - ***Selectivity/Specificity***
  - ***Ruggedness***
  - ***Calibration***
  - ***Limit of Detection & Limit of Quantification***

Final Thoughts

- ***Many companies will have a method, fewer will have a have validated method***
- ***Validation is time consuming and costly***
- ***Different validation protocols available, with many similarities, but takes a "skill" person with some statistical knowledge to perform***
  - ***Some spreadsheets & programs developed/in-development***
- ***Recommend a team of reviewers to ensure objectivity and confidence in recommendation***
- ***With rapid pace of technology, method deviations are becoming more common and frequent, without validation and documentation***

Common Weaknesses

- *Just have method or SOP, with limited validation data or proof of data quality*
- *Missing pieces*
- *Only tested on their product(s)*
- *Only internal, no external/outside data*
- *Provide best data vs. "real-world"*
  - *not different days, instruments, operators, etc.*
- *Proprietary*
- *Too costly for small companies*
- *Approved method may be impractical or too costly for Regulatory Labs*

Future Lab Challenges & Concerns

- *High level of staff turn over and/or retirements*
- *Limited training opportunities*
  - *SOP is the starting point, not the finish line*
  - *Vendors talk about their instruments, not specific methods*
  - *Graduates not ready on Day 1*
  - *Increasing "virtual" lab training*
- *Nutrient testing history/mentality*
- *New active areas are beneficial substances enhanced efficiency, biostimulants, chelates, etc.*
- *Limited microbiological experience/resources*
- *Some instrumentation is very expensive and highly technical & only applies to few products*

Summary

- *The LSC uses defined criteria for evaluating a proposed method*
- *These criteria are consistent with AOAC, ASTM, ISO, etc.*
- *Shortcomings can be identified and generally corrected*
- *Cost and practicality of performing the method receive some consideration, but aren't "deal-breakers"*
- *Many of the new products pose some challenge as far as expertise, equipment, costs, etc.*
- *Great interest, but limited support for these efforts*