Analytical Validation and Points for Discussion

Julia Tait Lathrop, PhD
Immunology and Hematology Branch
Division of Immunology and Hematology Devices
Office of In Vitro Diagnostics and Radiological Health
Center for Devices and Radiological Health
Food and Drug Administration

What's so hard about validating proteomic tests?

- Mass spec well understood, well validated in analyte evaluation, newborn screening tests
- High sensitivity
- High precision
- Multi-plexing with multiple analytes done all the time
- Constantly being updated and optimized
- Decades of experience with immunoassays can be applied to biomarkers using proteomic technologies

FDA: Proteomic technologies are just like every other technology

- Analytical validation demonstrates the accuracy, precision, reproducibility of the test- how well does the test measure what it claims to measure?
- Clinical validation demonstrates the effectiveness of the test- how relevant is the test measurement to the clinical condition?
- The goals for demonstrating analytical and clinical validation are no different for MS or multi-analyte arrays than they are for any other test.

FDA: Proteomic technologies are nothing like any other technologies

A number of challenges are specific to proteomic-based technologies (arrays, MS, gels)

- Immunodepletion of patient samples
- Multiple technologies
- Complex, multi-parameter algorithms
- Lack of gold standards for analytes
- Lack of reference methods for assays
- Expectations that technological optimization is directly translatable to IVDs

Outline

- Definitions
- Principles of Analytical Validation
  - *Are there elements of validation that are technology-specific?*
  - *How to control for sample preparation variability?*
  - *Sample Biorepositories-how useful are they?*
- How can a test developer know what to *do?*
- *What role can/should funding agencies play in improving validation?*
- *What should FDA do next? How can we help?*
- Points for Discussion with speakers and audience

Definitions

How FDA defines:

- Safety:
  - How accurate are the results of measuring the analyte?
  - What is the risk to patient of wrong result and of test format?
    - False positive (FP) genotype as determination of treatment vs FP for diagnosis in conjunction with signs and symptoms
    - False negative (FN) in rule out test vs. FN in aid to diagnosis
    - Incorrect result in blood typing
  - Analytical validation
- Effectiveness:
  - Is the test result relevant to the clinical condition?
  - Clinical validation

How FDA defines (colloquially)

- An “assay”
- A “run”
- A “replicate”
- A “cut-off”
- An “algorithm”
- “verification” vs “qualification” vs “validation”

How FDA defines (colloquially)

- An "assay"

Assay includes the entire test system, from sample collection to preparation to delivery of test results

- A "run"
- A "replicate"
- A "cut-off"
- An "algorithm"
- "verification" vs "qualification" vs "validation"

How FDA defines (colloquially)

- An "assay"
- A "run"

A "run" is a single processing of a sample, from sample prep to test result. Ex, 2 runs per day means 2x sample prep through reporting of results, not just MS run

- A "replicate"
- A "cut-off"
- An "algorithm"
- "verification" vs "qualification" vs "validation"

How FDA defines (colloquially)

- An "assay"
- A "run"
- A "replicate"

A replicate is a number of spots/wells/etc from a single sample preparation process, i.e., two replicates per run means two replicates from one sample

- A "cut-off"
- An "algorithm"
- "verification" vs "qualification" vs "validation"

How FDA defines (colloquially)

- An "assay"
- A "run"
- A "replicate"
- A "cut-off"

The medical decision point that determines health from disease, stable from progression, etc. Cutoff can be dichotomous (qualitative positive or negative) or continuous where the value has a specific clinical meaning, i.e. IgG light chain level in multiple myeloma

- An "algorithm"
- "verification" vs "qualification" vs "validation"

How FDA defines (colloquially)

- An "assay"
- A "run"
- A "replicate"
- A "cut-off"
- An "algorithm"

An algorithm is the mathematical process by which various variables are combined into an actionable score. It is part of the assay, determines the cutoff, and should be locked down prior to validation

- "verification" vs "qualification" vs "validation"

How FDA defines (colloquially)

"Verification" vs "validation" vs "qualification"

- Verification- small sample sets used to ensure that the lab can recapitulate manufacturer's specifications
- Validation- demonstration by a test developer that the test meets pre-specified performance criteria and is safe and effective for its intended use

- Qualification- is a separate issue and its definition varies depending on the use- efficacy of biomarker for clinical endpoint, performance as a development tool, CDER DDT

Principles of Analytical Validation

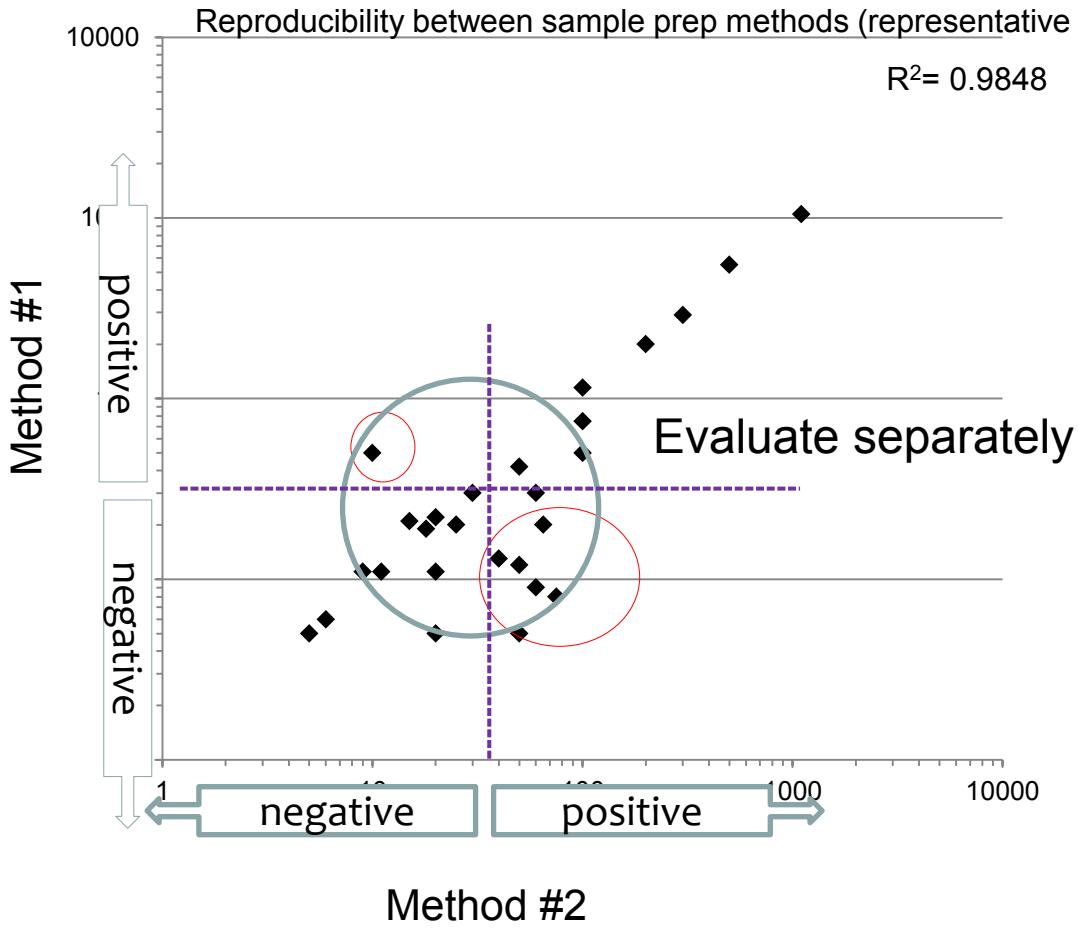
Analytical validation explores all aspects of the assay performance:

- Precision, including repeatability and reproducibility
- Analytical sensitivity: Limits of blank, detection, quantitation
- Analytical specificity and interference
- Linearity of quantitative and semi-quantitative assays
- Stability of samples, calibrators, controls; real time vs accelerated; on-board vs storage vs opened/prepared samples
- Matrix comparison
- Method comparison [for 510(k)]
- Measuring range, reference range
- Software for instrument and algorithm

Analytical sensitivity and specificity evaluate assay, not clinical, performance

- Analytical sensitivity- LoD/LoB/LoQ
  - How sensitive is the assay? False negatives
  - How to handle with MS and multiplex features?
  - Each feature or final output?
- Analytical specificity
  - How specific is the assay?
  - Evaluate interfering substances and differential diagnoses
  - Evaluate cross-reactivity in capture, identification steps

Performance around the clinical decision point is paramount



Clinical decision point

Demonstrating validity around the medical decision point is paramount

- Medical decision point is the result that directs patient care. This may be:
  - The result that differentiates "positive" and "negative" in a qualitative assay;
  - The entire measuring range in a quantitative assay;
  - A combination
- At Decision point the SD and %CV of the assay can change patient treatment
  - Statistical significance ≠ clinical significance

FDA reviews *all* of the data

- FDA reviews, re-plots, the line data
- Review includes expert consults
  - Clinical experts- Medical officers
  - Statisticians- OSB
  - Software reviewers-OIR
  - Technology experts- OSEL, CBER, CDER
  - Other experts: OIR and other Offices and Centers (all within FDA unless panel is called for PMA)

Differences in validation design b/w research and clinical use can be significant:

Research

- Samples: Pools of disease and healthy
- Precision: 2 pools, disease vs healthy , 5 replicates x 5 days
- Linearity- 3:1, 1:1, 1:3 dilutions of pools
- Stability- accelerated
- Disclosure: Public deposition of data and protocols?

FDA

- Samples: Individual patient samples except for calibrators
- Precision: individual patient samples through the AMR, 2 replicates/run, 2 runs/day, 20 non-consecutive days
- Linearity: proportional dilutions of individual samples
- Stability- real-time necessary by the end of review
- Disclosure: Decision summary and PI contain study design and results

Point for Discussion: Are elements of validation technology-specific?

Which (if any) elements of proteomic technologies have unique attributes that affect validation?

- *MRM vs MALDI-ToF- validation of a single analyte vs validation of a profile*
- *Multi-analyte microarrays*
- *2D gels/WB*
- *Reference methods vs no gold standard for outcome*
- *Quantitation requirements*

Study design requirements come from understanding failure modes

- Sources of variability
  - Sample collection, shipment, preparation
  - Controls and calibrators
  - Instrument variation
  - Patient variability
- Performance of assay components
  - Antibody cross-reactivity, protein/peptide ID
  - Identify most vulnerable features
  - Assays must have sufficient statistical power to identify failures

Point for discussion: How to control for variability stemming from sample preparation?

- *Sample prep- collection, shipping, depletion, trypsinization, sample stability- these are part of the total error of the assay.*
- *How to manage the technical variability, e.g. sample prep?*
- *How to manage the biological variability?*
- *If the score is reproducible in the end, isn't that enough?*

Point for Discussion: Sample Biorepositories

- *How useful are current biorepositories?*
- *How relevant are the samples to the IU?*
- *Use best practices: annotation, storage, collection?*
- *How many differential diagnosis samples are available? How many are being planned in current collection strategies?*
- *Can you use these samples for validation? Should FDA require samples from several different sources if they are?*

How can a test developer know what to *do*?

Several different guidelines* are available depending on the IU of the test and the level of development

*starting place, not prescriptive

Choice of guidelines depends on the intended use of the assay

	ICH Q2A/Q2B	CLIA	DDT Qualification	CLSI
Source	International Committee on Hybridization/CDER/CBER	Clinical Laboratory Improvements Amendment	CDER/FDA	Clinical and Laboratory Standards Institute
Utilized by (ex.)	ICH/CDER/CBER	FDA/CMS	FDA	FDA/CDRH
Intended Users	Regulators, QC, drug, biologics manufacturers	Clinical labs	Drug developers	Test developers
Purpose	Analytical validation of product characterization assays	Ensures proper training of lab personnel and performance of assays	Develop platform-agnostic biomarkers for use in drug development NOT for use in patient management	Analytical validation of in vitro diagnostic assays

Choice of guidelines depends on the intended use of the assay

example of differences in study design

	ICH Q2A/Q2B	CLSI
Title	Guidance for Industry Q2B Validation of Analytical Procedures: Methodology	EP05-A2 Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline—Second Edition
Repeatability/Precision	6 replicates at 100% strength or 3 replicates of 3 doses	Samples at clinical decision point, +/- 25%, high and low range and in between: run in duplicate/2 runs/day for 20 days. Compare between run/within run/between day, etc

CLSI guidelines can be useful for planning study design

- Clinical and Laboratory Standards Institute
- Different recommendations for test developers and for labs implementing developer-validated tests
- Standards are recommendations, not regulations
- Not all guidelines are sufficient for every assay
- Not all assays have to adhere to every guideline
- Overall goal is determination that the device is safe and effective for its intended use

Why Won't FDA commit to Criteria? How does a developer know what to aim for?

Acceptance criteria and study design depend on the type of device and the Intended Use:

- Genotyping vs autoimmune
- Screening vs monitoring
- Stand alone diagnostic vs adjunct to signs and symptoms
- Risk to patient of wrong result- adjunct vs blood typing
- Prevalence of analyte and existence of gold standard

Why Won't FDA commit to Criteria?

	Genotyping for FVIII	Ro60 in Systemic Lupus Erythematosus (SLE)
Prevalence of analyte	Homozygotes 2% in population	~30% in SLE
Sensitivity Acceptance criterion	100.0% (95% CI 96.3-100.0%)	24.0% (95% CI 17.4-31.6%)
Specificity Acceptance criterion	99.3% (95% CI 98.6-99.8%) 5 no calls, 0 wrong calls	97.6% (95% CI 94.8-99.1%)
Study design	80 hetero, 8 homo, 750 WT	150 SLE+, 240 SLE Differential Diagnosis (e.g., SS), 0 NHS
Precision	4 samples, 12 of each, 0 wrong calls	12 samples spanning AMR, 80-120 replicates of each, <10% total imprecision
Reference method for test	Sanger sequencing	None

510(k) decision summaries and PMA SSED are publically available

- Provide an outline of what FDA has required for previous approvals/clearances
- Describes the studies that were done, the data that was submitted and reviewed
- Review most recent clearances
- www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/510kClearances/default.htm

Point for Discussion

What role can or should funding agencies play in improving validation?

Point for Discussion: What should FDA do next?

How can FDA help move these tests into the clinic?

Next steps?

Point for Discussion (1): Are elements of validation technology-specific?

Which (if any) elements of proteomic technologies have unique attributes that affect validation?

- MRM vs MALDI-ToF- validation of a single analyte vs validation of a profile
- Multi-analyte microarrays
- 2D gels/WB
- Reference methods vs no gold standard for outcome
- Quantitation requirements

Point for discussion (2): How to control for variability stemming from sample preparation?

- Sample prep- collection, shipping, depletion, trypsinization, sample stability- these are part of the total error of the assay.
- How to manage the technical variability, e.g. sample prep?
- How to manage the biological variability?
- If the score is reproducible in the end, isn't that enough?

Point for Discussion (3): Sample Biorepositories

- How useful are current biorepositories?
- How relevant are the samples to the IU?
- Use best practices: annotation, storage, collection?
- How many differential diagnosis samples are available? How many are being planned in current collection strategies?
- Can you use these samples for validation? Should FDA require samples from several different sources if they are?

(Moore et al, J. Proteome Res. (2011) 10:3429)

Point for Discussion (4)

What role can or should funding agencies play in improving validation?

Point for Discussion (5): What should FDA do next?

How can FDA help move these tests into the clinic?

Next steps?

References

- 21 CFR§ 800 et seq
- 510(k) decision summaries
- FDA Guidances:
  - Device Advice: www.fda.gov/MedicalDevices/DeviceRegulationandguidance/default.htm.
  - PreSubmission process: Draft Guidance for Industry and FDA Staff Medical Devices: The Pre-Submission Program and Meetings with FDA Staff": http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm310375.htm
- Mock 510(k)s: Regnier et al (2010) Clinical Chemistry 56(2):165-171 and supplementary material
- Li et al, (2011), J. Proteomics 74:2682
- CLSI Standards, especially EP09, EP17, EP06
- IoM report, NCI checklist

Thank you!

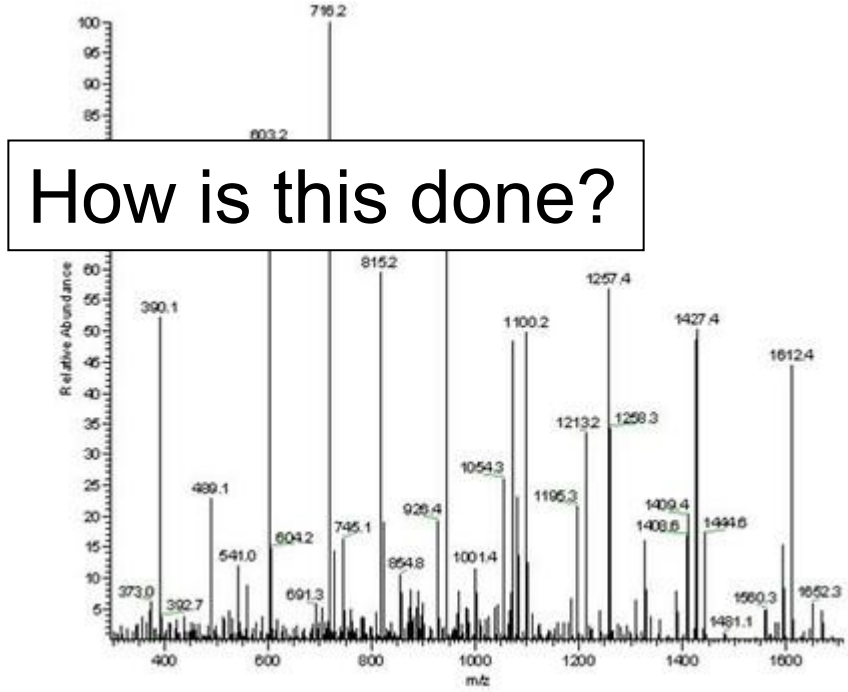
Julia.lathrop@fda.hhs.gov

Sidebar: When do you need an IDE?

- It is not always obvious, but FDA has the last word
- Risk: Based on the effect on a patient of a wrong result- i.e., denied treatment when needed, given wrong treatment
- BUT:
  - There is NO case in which it can be assumed that an IDE is never needed = i.e., Phase I trial, Stage IV disease
  - IDEs allow investigators to use an investigational device in patients whether or not it's a CDx

Point for Discussion: Do you need to validate each individual feature?

- *It depends on impact, number, variability, scientific, technical, and practical concerns
- Stems from identifying and addressing failure modes
  - Which features are key to making a clinical decision?
  - Which features are most sensitive to assay conditions?



*http://bioinformatics.uncc.edu/research-areas/computational-mass-spectrometry

The FDA PreSubmission Process

Precision study design depends on # sites, etc

	One Site		Multiple Sites	
Instruments	One instrument	Multiple instruments/ models	One instrument	Multiple instruments
Operators	One operator	Multiple operators	One operator/site	Multiple operators/site
One lot	One lot	Multiple lots	One lot	Multiple lots
Study Design	2 replicates/2 runs/day for 20 days	2 replicates/2 runs/day for 20 days per instrument	3 sites/1 instruments/1 operators: > 20 days, all components tested	3 sites/3 instruments/3 operators: > 20 days, all components tested
Results	Total Precision, repeatability at every level tested	Total Precision, repeatability Reproducibility at every level individually and pooled	Total Precision, repeatability Reproducibility at every level, individually and pooled	Total Precision, repeatability Reproducibility at every level, individually and pooled

See CLSI EP05 §§ 10 AND 11 (for manufacturers)