Validation Case studies: the good, the bad and the molecular

Patti Jones PhD
Professor of Pathology
UT Southwestern Medical Center
Director of Chemistry
Children’s Medical Center Dallas
Learning Objectives

• Through example validations:
  – Compare the assay validations required for FDA-approved tests versus Laboratory Developed Tests (LDTs)
  – Delineate the basic steps to validating an assay
  – Discuss the validation of a molecular test
Method Validation

• Clinical Laboratory Improvement Amendments of 1988 (CLIA ’88)
  
  – Subpart K – Quality System for Non-waived Testing
  – Sec. 493.1253 Standard: Establishment and verification of performance specifications

  Tells you **what** must be validated, but not **how** to do it.
Method Validation

- CLIA Regulation 493.1253(2)
  1. Accuracy – (closeness to true/comparative method)
  2. Precision – (reproducibility)
  3. Reportable range – (linearity, AMR, MD/C)
  4. Reference interval –
  5. Analytical sensitivity – (lower limit)
  6. Analytical specificity – (interferences)
  7. Other specifications
Case 1: the Good – FDA-approved methods/reagents

- CLIA Regulation 493.1253(2)
  1. Accuracy
  2. Precision
  3. Reportable range
  4. Reference interval
     Determine the assay performs in your hands the way the manufacturer says it performs.
  5. Analytical sensitivity
  6. Analytical specificity
  7. Other specifications
Case 1: FDA-approved test

• Three basic studies:
  – Precision studies
    • Reproducibility

  – Reportable Range Study
    • Linearity, analytical measurement range (AMR), maximum dilution/concentration (MD/C)

  – Assay comparison/correlation study
    • Accuracy, Reference interval
Case 1: glucose

• Precision Studies:
  – Within-run precision
    • Patient or QC samples assayed 20 times on the same day within the same run
    • If precision poor, no need to do further eval
  – Between-run precision
    • Patient or QC samples once per day for 20 days
    • Establish qc range as well as total precision
  – Samples at least 2 - 3 levels – medical decision points
# Case 1: Glucose: precision

## Table 1. Data on imprecision.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Glucose mg/dL (mmol/L)</th>
<th>Mean</th>
<th>SD</th>
<th>CV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>CV, %</td>
</tr>
<tr>
<td><strong>Within run (n = 20 replicates)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>50.4 (2.8)</td>
<td>1.4 (0.08)</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>200.6 (11.14)</td>
<td>2.7 (0.15)</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td><strong>Between run (n = 20 runs)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>51.2 (2.84)</td>
<td>2.1 (0.12)</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>202.3 (11.24)</td>
<td>3.5 (0.19)</td>
<td>1.7</td>
<td></td>
</tr>
</tbody>
</table>
Case 1: Glucose: precision

• What’s good precision?
  – Depends on the analyzer and the analyte
    • <5% CV considered good
    • Many automated analyzers (blood gases) < 1% CV
    • Tandem MS, HPLC, etc <10% CV is excellent
Case 1: Glucose: reportable range

• Validation of Reportable Range
  – Minimum of 3 test specimens (4-5 better), measured in duplicate or triplicate
  – Appropriate matrix
  – Well established target concentrations
  – Concentrations near the low, midpoint, and high values of the AMR
Case 1: Glucose: reportable range

\[ y = 1.0087x - 1.9098 \]
\[ R^2 = 0.9998 \]

<table>
<thead>
<tr>
<th>known</th>
<th>Average measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td>100</td>
<td>102</td>
</tr>
<tr>
<td>250</td>
<td>246</td>
</tr>
<tr>
<td>500</td>
<td>504</td>
</tr>
</tbody>
</table>

linearity = 25 – 500 mg/dL
CAP “AMR”
CLIA “reportable range” = CAP AMR +

Verify maximum dilution/concentration (MD/C) (CRR)?
• spike serum with high concentration (5000+)
Case 1: glucose: reportable range - MD/C

<table>
<thead>
<tr>
<th>dilution</th>
<th>theor</th>
<th>measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:100</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>1:50</td>
<td>100</td>
<td>97</td>
</tr>
<tr>
<td>1:20</td>
<td>250</td>
<td>248</td>
</tr>
<tr>
<td>1:10</td>
<td>500</td>
<td>502</td>
</tr>
<tr>
<td>1:5</td>
<td>1000</td>
<td>992</td>
</tr>
<tr>
<td>1:2</td>
<td>2500</td>
<td>2100</td>
</tr>
<tr>
<td>0</td>
<td>5000</td>
<td>2875</td>
</tr>
</tbody>
</table>
Case 1: Glucose: Accuracy - Correlation

• Comparison of Methods – correlation
  – Select a minimum of 20 (usually 40 – 60) patient’s serum samples with analyte values as evenly distributed throughout the linear reportable range of the assay as possible
  – Assay all samples by the current method (comparative or reference method – x-axis data) and the method being evaluated (test method – y-axis data)
Case 1: Glucose - correlation

• Correlation experiment:
  – Plot \((x,y)\) pairs of values and apply appropriate regression analysis to these data
  – Obtain linear regression equation of the line (least squares line)
    • Assumption: Any errors are in the test method (TM) and not the comparative method (CM)
  – Deming Regression – assigns errors to both methods depending on their variances
  – Slope, intercept, correlation coefficient, standard error of the estimate, bias plot
Least-Squares Linear Regression Data

\[ y = 0.997x - 0.34 \]

\[ R^2 = 0.9997 \]

\[ S_{y/x} = 2.21 \]
No Errors
slope = 1.000
y-intercept = 0.000
$s_y/x = 0.000$
r$^2 = 1.000$

"Ideal" Regression Plot/Data
Narrow and Unevenly Distributed Range of Analyte Concentrations – effect on Correlation Coefficient and Regression

y = 1.029x - 25.0
$r^2 = 0.995$

y = 0.793x + 8.38
$r^2 = 0.9715$

sodium

y = 1.03x - 1.9352
$r^2 = 0.9844$

sodium

y = 0.813x + 28.43
$r^2 = 0.5567$
Case 1: Glucose: Total Error

- Method evaluation should also ensure that the magnitude of the errors affecting the results are acceptable
  - Evaluate the Total Error (TE) in the assay
    - TE = Random Error (RE) + Systemic Error (SE)
      - RE = SD_{TM} - largest component of TE
      - SE = Bias: determined by regression equation; plugging in medical decision point value

- Marginal method  2SD < TE
- Fair method  3SD < TE
- Good method  4SD < TE
- Outstanding method  6SD < TE
Allowable Error ($E_A$)

“Acceptable” analytic error is decided using a **performance standard (PS)** based on the maximum allowable error ($E_A$) at a medical decision concentration of analyte ($X_C$).

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Acceptable Performance, ±</th>
<th>$X_C$</th>
<th>$E_A$</th>
<th>Max SD, CLIA</th>
<th>Max SD, Fraser</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALT</td>
<td>20%</td>
<td>50 U/L</td>
<td>10</td>
<td>2.5</td>
<td>5.8</td>
</tr>
<tr>
<td>Calcium</td>
<td>1.0 mg/dL</td>
<td>10.8 mg/dL</td>
<td>1.0</td>
<td>0.25</td>
<td>0.10</td>
</tr>
<tr>
<td>Glucose</td>
<td>10%</td>
<td>126 mg/dL</td>
<td>12.6</td>
<td>3.2</td>
<td>2.8</td>
</tr>
<tr>
<td>pH</td>
<td>0.04</td>
<td>7.35</td>
<td>0.04</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.45</td>
<td>0.04</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>pCO₂</td>
<td>5 mm Hg</td>
<td>35 mm Hg</td>
<td>5.0</td>
<td>1.25</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Example: Glucose

PS = ±10% @ $X_C$

If $X_C = 200$ mg/dL

$E_A = 20$ mg/dL
Case 1: Glucose: total error

- Evaluate TE of Test Method (TM)
  - RE of Glucose TM
  - From an inter-assay precision study, $SD_{TM} @ [\text{glucose}] = 200 \text{ mg/dL}$ was 3.5 mg/dL
  - $RE = 4 \times SD_{TM} = 4(3.5) = 14 \text{ mg/dL}$
Case 1: Glucose: total error

• SE(bias) of Glucose TM
  – From linear regression equation for glucose
    • slope = 0.997, y-intercept = -0.34
    • \( Y_C = 0.997X - 0.34 \)
  – Therefore, @ [glucose] = \( X_C = 200 \text{ mg/dL} \)
  – \( Y_C = 0.997(200)-0.34 = 199.06 \)

• \( SE = \left| Y_C - X_C \right| = \left| 200.00 - 199.06 \right| = 0.94 \text{ mg/dL} \)

• \( TE = RE + SE = 14 + 0.94 = 14.94 \text{ mg/dL} \)
Case 1: glucose: method decision charts

Bias = 0.94 = 4.5%
SD = 3.5 = 17.5%
Case 2: The **Bad**: LDTs

- Modified FDA-approved, In-house, Scratch, Homebrew – Laboratory Developed Test

1. Accuracy
2. Precision
3. Reportable range
4. Reference interval
5. Analytical sensitivity – (lower limit)
6. Analytical specificity – (interferences)
7. Other specifications
Case 2: The Bad: LDTs

- Precision studies
- Analytical Measurement Range study

- Accuracy
  - Assay comparison/correlation study
  - Recovery
  - Analyte identity

- Lower limit of detection/quantitation
  - How low can you go

- Interferences Study
  - No manufacturer’s data

- Reference Interval Study
  - How much should be there

- Clinical validity - sensitivity, specificity, predictive values
3-Hydroxy-fatty acids – (3OHFA)

• Intermediates of Mitochondrial fatty acid oxidation
  – Serum build-up of 3-OHFAs indicates deficiencies of 3-hydroxy-acyl-CoA dehydrogenases
  – Diagnosis of LCHAD, SCHAD and TFP deficiencies
  – Stable-isotope dilution, electron impact ionization Gas Chromatography-Mass Spectrometry
3-Hydroxy-fatty acids

• Background:
  – Assay measures 6 different chain length 3OHFAs
  – 6 different Isotope-labeled standards, 6C – 16C
  – C\textsuperscript{13} in place of C\textsuperscript{12} in two places in the compound
  – example:
Case 2: 3OHFA: reportable range

• Analytical measurement range
  – tried 0.001 - 500 μM
  – best range 0.2 - 50 μM (range of linearity)
  – LOQ 0.2μM

• Concentration of isotope-labeled standards
  – Internal Standard ≅ Analyte concentration
  – tried 0.1 - 100 μM 10μM
Case 2: 3OHFAs: analytical sensitivity: lower limit

• Limit of detection (limit of absence)
  – Assay patient sample with no measurable analyte present multiple (20 times) – LOD = mean ± 3SD
  – Assay zero calibrator
  – 3OHFA - 0.05 calibrator
    • Mean + 3SD = 0.0195 + 0.078 = 0.0975 = 0.1 µM

• Limit of quantitation (functional sensitivity)
  – Minimum concentration that can be reproducibly measured at an acceptable CV – CV based on biological variation, i.e. CV < 20% TSH
  – CV at 0.2 µM = 5 – 10%
Case 2: 3OHFAs: precision

• Precision
  – Within-assay (20 points)
    • low (0.3 μM): cv 4 - 8 %
    • high (30 μM): cv 1 - 4 %
  – Between-assay (25 points)
    • low (0.3 μM): cv 4 - 15 %
    • high (30 μM): cv 1 - 2 %
Case 2: 3OHFAs: Accuracy

• Correlation?

• Analyte Identity

• Recovery
Case 2: 3OHFAs – analyte identity

- Analyte Identity - analyzed each chain length native and isotopic compound via GC/MS to determine fragmentation pattern and appropriate ions to use for identification and quantitation.
**Case 2: 3OHFAs: analyte identity**

– Stable isotope standards

<table>
<thead>
<tr>
<th>3-OHFA</th>
<th>Native MW Quant</th>
<th>Isotope MW Quant</th>
<th>Native MW confirm</th>
<th>Isotope MW confirm</th>
<th>Retention time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-OH-C6</td>
<td>261</td>
<td>263</td>
<td>233</td>
<td>235</td>
<td>16 – 19</td>
</tr>
<tr>
<td>3-OH-C8</td>
<td>289</td>
<td>291</td>
<td>233</td>
<td>235</td>
<td>22 – 25</td>
</tr>
<tr>
<td>3-OH-C10</td>
<td>317</td>
<td>319</td>
<td>233</td>
<td>235</td>
<td>28 – 32</td>
</tr>
<tr>
<td>3-OH-C12</td>
<td>345</td>
<td>347</td>
<td>233</td>
<td>235</td>
<td>36 – 39</td>
</tr>
<tr>
<td>3-OH-C14</td>
<td>373</td>
<td>375</td>
<td>233</td>
<td>235</td>
<td>43 – 47</td>
</tr>
<tr>
<td>3-OH-C16</td>
<td>401</td>
<td>403</td>
<td>233</td>
<td>235</td>
<td>48 – 51</td>
</tr>
</tbody>
</table>
Case 2: 3OHFAs: recovery

• Accuracy - recovery

  – Target          % recovery
   (0.3 µM):      87 - 103
   (8.0 µM):      89 - 108
   (30 µM):       94 - 100
Case 2: 3-OHFA: method decision charts

Total Error = 0.5
Bias = 0.04 = 8%
SD = 0.1 = 20%
Case 2: 3OHFAs: specificity (interferences)

Interference Experiment Notes:
1. Volume of interferent should be ≤10% of total volume
2. Use several samples per interferent
3. At a minimum, test samples in duplicate
3-Hydroxy-fatty acids

• Interferences
  – No:
    • Hgb, Bilirubin, Lipemia
    • Spike with badly hemolyzed sample
      – Get same result – hemolysis does not affect assay
      – Get a different result – begin spiking with increasingly dilute hemolyzed sample
  – Yes:
    • Citrate large 347 ion in citrate
Case 2: 3-OHFAs: reference interval

• Reference Interval Study
  – **ESTABLISH** an Interval, rather than validate one
  – Typically, require a minimum of 120 specimens from “healthy” individuals
  – Specimens from “healthy” individuals – medically well; taking no drugs, herbals, vitamins, or other substances likely to affect analyte values
  – Pediatrics!
Case 2: 3OHFAs: reference intervals

- Reference intervals

<table>
<thead>
<tr>
<th>3-OH-FA</th>
<th>normal range (µmol/L)</th>
<th>MCT range (µmol/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-OH-C6</td>
<td>0.4 - 2.3</td>
<td>1.6 - 18.9</td>
</tr>
<tr>
<td>3-OH-C8</td>
<td>0.2 – 1.0</td>
<td>0.4 - 8.3</td>
</tr>
<tr>
<td>3-OH-C10</td>
<td>0.2 - 0.6</td>
<td>0.2 - 2.3</td>
</tr>
<tr>
<td>3-OH-C12</td>
<td>0.2 - 0.6</td>
<td>0.2 - 1.4</td>
</tr>
<tr>
<td>3-OH-C14</td>
<td>&lt; 0.4</td>
<td>&lt; 0.9</td>
</tr>
<tr>
<td>3-OH-C16</td>
<td>&lt; 0.5</td>
<td>&lt; 0.8</td>
</tr>
</tbody>
</table>
3-Hydroxy-fatty acids

3-OH-FATTY ACIDS

Concentration (umol/L)

Carbon chain length

C6 C8 C10 C12 C14 C16

LLN
ULN
S1
S2
L1
L2
Case 2: 3O HFAs

• Clinical Validity
  – Sensitivity
    • For LCHAD = 100%
    • For SCHAD = 100%
  – Specificity
    • For LCHAD = 100%
    • For SCHAD = 89.8%
  – Positive predictive value
    • For LCHAD = 100%
    • For SCHAD = 26.8%
  – Negative predictive value
    • For LCHAD = 100%
    • For SCHAD = 100%
Case 2: 3OHFAs

<table>
<thead>
<tr>
<th>VALIDATION</th>
<th>Accept</th>
<th>Comments</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linearity (analytical measurement range)</td>
<td>✔️</td>
<td>see paper</td>
<td>Jones</td>
<td>2/2000</td>
</tr>
<tr>
<td>Precision (Reproducibility)</td>
<td>✔️</td>
<td></td>
<td>Jones</td>
<td>2/2000</td>
</tr>
<tr>
<td>Sensitivity (Lower limit of detection)</td>
<td>✔️</td>
<td></td>
<td>Jones</td>
<td>2/2000</td>
</tr>
<tr>
<td>Accuracy (Comparison / Correlation)</td>
<td>✔️</td>
<td></td>
<td>Jones</td>
<td>2/2000</td>
</tr>
<tr>
<td>Reference range (normal patients)</td>
<td>✔️</td>
<td></td>
<td>Jones</td>
<td>2/2000</td>
</tr>
<tr>
<td>Specificity (Interferences)</td>
<td>✔️</td>
<td></td>
<td>Jones</td>
<td>2/2000</td>
</tr>
</tbody>
</table>

New method/instrument is acceptable and put in use on 2/2000 (date).

Technologist: [Signature] date 2/2000

Clinical Consultant: [Signature] date 2/2000

Case 3: the **molecular**: CMV by real time PCR

• Start out by setting up assay and optimizing
  – Selection of primers
  – Tweaking master mix
  – Optimizing temperatures and times in each cycle

After optimization – validation!
Case 3: CMV by real time PCR

• Follow CLIA guidelines - LDTs
  1. Accuracy
  2. Precision
  3. Reportable range
  4. Reference interval
  5. Analytical sensitivity
  6. Analytical specificity
  7. Other specifications
Case 3: CMV by real time PCR

- **Reportable range**
  - AMR/linearity

- Testing $10^2$ copies per mL to $10^5$ copies per mL – determining cycle time (CT) equivalent for viral DNA copies
Case 3: CMV by real time PCR

• Reportable range
  – 1000 copies/mL to 5,000,000 copies/mL.
  – Linear in this range of quantitation with a correlation coefficient of 0.998. (graphing CT vs viral copies)

• Precision
  • **Within-assay variation**: Ranged from 0.5% to 4.3% by CT value, and from 11% to 56% by viral copy number
  • **Between-assay variation**: Ranged from 1.7% to 3.9% by CT value, and from 23% to 70% by viral copy number
Case 3: CMV by real time PCR

• Accuracy

• 20 paired whole blood samples encompassing the linear range were assayed at CMC and Focus Dx. Correlation coefficient between the two assays was 0.93.
Case 3: CMV by real time PCR

• Analytical sensitivity – lower limit
  – Limit of Quantitation: 1000 copies/mL

• Analytical specificity – interferences
  • No cross-reactivity detected with other Herpes viruses or organisms common in transplant recipients.

• Reference interval –
  – Undetectable
Case 4: Glutaric Acidemia, type 1 (GA1)-GCDH gene sequencing

- Extraction of DNA from fibroblast cell culture
- Amplification of DNA via RT-PCR
- Purification of PCR product DNA
- Sequencing of DNA
- Reviewing data and identification of gene sequence mutations
Case 4: Glutaric Acidemia, type 1 (GA1)-GCDH gene sequencing

• Follow CLIA guidelines - LDTs
  – Accuracy
    • 10 fibroblast cultures purchased from Coriell Institute for Medical Research, each harboring a specific change in the GCDH gene. The assay detected all changes in 100% of the cell lines, and also detected some changes that were not specified by Coriell.
  – Precision
    • Same samples assayed repeatedly – same results
  – Reportable range
    • Can detect Heterozygous wild type, heterozygous mutant, homozygous wild type, homozygous mutant
Case 4: Glutaric Acidemia, type 1 (GA1)-GCDH gene sequencing

• Follow CLIA guidelines – LDTs
  – Analytical sensitivity
    • Limit of Detection (LOD): 50 ng of DNA per PCR reaction
  – Analytical specificity
    • No cross-reactivity with other regions of the genome was detected. PCR products produced single amplification bands of the expected sizes. Sequencing primers used in this assay were confirmed to be complementary to regions of the GCDH gene.
  – Reference interval
    • No mutations detected – homozygous wild type
Case 4: Glutaric Acidemia, type 1 (GA1)-GCDH gene sequencing

• Follow CLIA guidelines – LDTs
  – Other specifications
    • Clinical Validity:
      – GA-1 is an autosomal recessive disorder of lysine, hydroxylysine, and tryptophan metabolism caused by deficiency of glutaryl-CoA dehydrogenase enzyme.
      – Numerous publications have correlated GCDH mutations with the onset of Glutaric Acidemia type I.
      – Recognition of this disorder before onset of neurological symptoms is essential for treatment and prevention of permanent damage.
      – Standard acylcarnitine and organic acid assays may miss this disorder in the case of “low-excretors”. 
Self Assessment Questions

1. FDA-approved tests require validation of the following parameters:
   a) Precision, accuracy, interferences, lower limit of quantitation
   b) Precision, interferences, reference interval, lower limit of quantitation
   c) AMR, precision, accuracy, lower limit
   d) Precision, AMR, reference interval, accuracy
Self Assessment Questions

2. LDTs:
   a) Should be validated for total error and for clinical validity before use
   b) Require validation of precision, accuracy, AMR and reference intervals only
   c) Do not need validation of interferences or lower limit of quantitation
   d) Do not include modified FDA-approved tests
Self Assessment Questions

3. Molecular tests
   a) Do not require CLIA validation before clinical use
   b) Only require validation of precision, accuracy, AMR, reference interval
   c) Require the same criteria be validated as any LDT
   d) Cannot be offered by a CLIA certified lab
Answers

1. D
2. A
3. C