Hurdles and Benefits of Implementing a QA Program Based on Patient Moving Averages

Mark A. Cervinski, PhD, DABCC
The Geisel School of Medicine at Dartmouth
Dartmouth-Hitchcock Medical Center

Problem with Traditional QC

Quality Control regulations are actually Quality Compliance regulations!
James Westgard, PhD

They specify the minimum frequency and don’t reflect what may really be needed to ensure the system is in control.

What is a Quality Laboratory to do?

- Could run controls more frequently
  - Some run several times daily
  - Forcing a “batch-mentality” on a continuous stream of data
  - Do laboratories that run frequent controls detect more analytical shifts?
- Establish “Delta Check” rules
  - Comparing patients results to previous results
    - Timing is an issue!
    - “Free” you are already running the samples
Is There Another Way?

- Why not use patient results as supplementary QC?
- What about interindividual variability?
  - Results can span from low to high?
  - True, individual patient results vary
  - BUT! The mean of the aggregated results should not vary (much)
- Average of Normals (AON) or Moving Averages (MA)

### Moving Averages Example

<table>
<thead>
<tr>
<th>Total Protein Result (mg/dL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2</td>
</tr>
<tr>
<td>7.1</td>
</tr>
<tr>
<td>6.8</td>
</tr>
<tr>
<td>7.4</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>7.9</td>
</tr>
<tr>
<td>5.7</td>
</tr>
<tr>
<td>7.4</td>
</tr>
<tr>
<td>7.9</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>7.3</td>
</tr>
<tr>
<td>7.4</td>
</tr>
</tbody>
</table>

- N = 1
  - Every patient result generates a point

- N = 5
  - Each point equals the mean of 5 patient results
  - Goal is to monitor the process: Not the patients
AON or MA Concept

- First proposed for laboratory QC in 1965
  - Hoffman RG., AJCP 1965;43:143-41
  - Plotted mean of 10 consecutive values within reference interval (A.K.A. Normal Range)
    - No optimization of protocols

Data Innovations Moving Averages participant’s Guide, 2010

Glucose MA Chart

- Only ten consecutive samples per each point
- Mean value is highly variable!


The Right Papers/Guidance

  - Blended the ideas of previous article
    - Calculating the ratio of Pop. SD to Analytic SD (Sp/Sa) ratio
    - Tested the effect of varying the number of patient samples averaged (N)
    - Effect of truncating outliers
  - Generated power function curves representing protocol performance
Effect of Varying Number of Results

The Right Papers/Guidance

  - Refinement of the previous article
  - Demonstrated that the number of patient samples per point could be much smaller
- ANPed = Average Number of Patients samples affected until error detected
- Both papers used randomly accessioned data...

Selecting the Correct Number of Points

- ANPed = Ave. Number of affected results prior to detection
- SE = systematic error
- Potassium data
  - Large circles = 5
  - Squares = 10
  - Triangles = 20
  - Small circles = 30
- Small N = detection of large shift
- Large N = detection of small shift
Hurdles: Step by Step

- Putting Theory Into Practice
- Need an commercially available program
- Need to determine:
  - How big of an error you wish to detect
  - How many points to analyze
  - Truncation limits
- Published nomograms and laboratory experience
- Modeling of the process

Defining Control Limits

- Determine control limits – how much error is tolerable?
- Cembrowski article used multiples of analytic SD for error detection
  - Na⁺ - SD = 1.00 mmol/L at 134 mmol/L
    - 2 x 1.00 = +/- 2 mmol/L
  - Total Protein – SD = 0.07 mg/dL at 6.9 mg/dL
    - 2 x 0.07 = +/- 0.14 mg/dL
  - Very tight error limits!
- Population SD or SEM, CLIA limits or RCV/LSC
  - Na⁺ Sp = 2.5 mmol/L x 2 = +/- 5 mmol/L
  - Na⁺ CLIA = +/- 4 mmol/L

Step 1: Population Mean & SD

- How do you determine the N?
- Each analyte needs to be considered individually
- Determine patient population mean and SD (Sp)
  - May or may not approximate reference interval…
  - Which population?
    - Ambulatory mean value ≠ inpatient mean value
    - Exclude certain populations?
      - ED, Dialysis, NICU, PICU, CCU, etc.
Population Mean and SD

- Inpatient values have a wide spread and tend to have lower concentration
- V. little difference between inpatient and ambulatory populations

Step 1: Analytic SD

- Determine analytical SD (Sa)
- QC value nearest the mean population value
- Determine the Sp/Sa ratio
- Apply to nomogram

Determination of N

- Sp/Sa ratio to nomogram
- 50% probability of detecting a 2SD shift
- Sodium Sp/Sa = 2.8
- Albumin Sp/Sa = 4.0

Mean = 7.1
Mean = 7.9
Total Protein, Ambulatory Patients
mg/dL
Truncation Limit Effect
Truncation Limits = 4.6 & 9.6 mg/dL; N = 12
ANPed SD ANPed SD
+0.8 mg/dL 22.00 16.48
-0.8 mg/dL 15.32 9.79

Truncation Limit Effect
Truncation Limits = 5.1 & 9.1 mg/dL;
ANPed SD
+2.7 mg/dL 30.67 17.39
-2.7 mg/dL 10.01 3.95

Narrow Limits = Reduced Performance
• Decreasing truncation limits reduces power of the protocol to detect shifts!
• Default exclusion limits in our program = +/- 4 SD
• Most analytes are not normally distributed
Ambulatory vs. Inpatient Populations

- Frequency Distribution of calcium data
- Light blue = all patients
- Orange = Ambulatory
- Green = Inpatients
- Two distinct patient populations!
- Optimized setting
  - \( N = 23 \)
  - \( TLH = 15.2 \)
  - \( TLL = 6.2 \)
  - Mean = 8.9 mg/dL

Two Populations, One Protocol?

- How do we account for this shift?
- Initial MA rules derived from literature
  - In paper - random selection of points from database
  - In reality – Inpatients vs. Outpatients
    - Essentially two separate patient populations with separate means
    - May require separate inpatient vs. outpatient protocols
Random Patients vs. Unit Draws

- Samples typically come down en masse
  - Morning draws, rounds, etc.
  - Not randomly selected!
- May need to exclude patient groups with values that deviate significantly from mean
- Exclude:
  - Pediatric pts, ICU, Heme-Onc, ED, etc.

Moving Averages for All Analytes?

- Will MA work for all analytes?
  - NO!
- Low volume tests
  - Amylase ~ 15 samples daily
    - N = 100+
  - Large gender differences
    - CK ~ 100 samples daily
    - Large differences between men and women

Protocol Validation and Potential Benefits

Does it work?
Protocol Validation

- Validation of simple protocols
  - Accumulated time stamped data in spreadsheet
  - Added incremental systematic error to data
    - Sent raw and altered data through a “shadow” program

Total Protein With Artificial Error

- Error induced at a fixed point
- Positive shift (0.8 mg/dL) detected in 12 samples
- Negative shift (0.8 mg/dL) not detected
  - 1.2 mg/dL detected in 22 samples (not shown)

Validation – Calculating ANPed

- Collected 96 days patient data
- Tested initial protocols - Matlab
  - Added artificial error every 10 samples
  - Calculated average number of patients affected until error detected (ANPed)

<table>
<thead>
<tr>
<th>Total Protein</th>
<th>ANPed</th>
<th>SD</th>
<th>Mode</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0 mg/dL</td>
<td>1452</td>
<td>1433.3</td>
<td>25</td>
<td>935</td>
</tr>
<tr>
<td>+1.0 mg/dL</td>
<td>18</td>
<td>24.1</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>
Optimization Process

- Simulated Annealing Process
  - Stochastic process: varying of truncation limits and N
  - Added a fixed error equal to control limit every 100 pts
- Cost = β(false positive rate) + ANPed
- More efficient optimization than total enumeration

Total Protein, All Pts

- Original Protocol settings:
  - Blue Diamonds
  - N = 12; TLH = 8.7 mg/dL; TLL = 5.7 mg/dL;
  - Mean = 7.2 mg/dL;
- Optimized Protocol settings:
  - Red squares
  - N = 37; TLH = 18.2 mg/dL; TLL = 2.4 mg/dL;
  - Mean = 6.9 mg/dL;

Total Protein Distribution

- Protocol settings:
  - Original: (blue arrows)
    - N = 12
    - TLH = 8.7 mg/dL
    - TLL = 5.7 mg/dL
    - Mean = 7.2 mg/dL
  - Optimized: (red arrows)
    - N = 37
    - TLH = 18.2 mg/dL
    - TLL = 2.4 mg/dL
    - Mean = 6.9 mg/dL
Widening the truncation limits improved performance
- Especially the detection of a negative error
- Allowed for tighter control limits

### Total Protein Performance

<table>
<thead>
<tr>
<th></th>
<th>Original Protocol</th>
<th>Optimized Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANPed</td>
<td>SD</td>
<td>Mode</td>
</tr>
<tr>
<td>-1.0 mg/dL</td>
<td>1452</td>
<td>1433.3</td>
</tr>
<tr>
<td>-0.8 mg/dL</td>
<td>6073.6</td>
<td>3675.9</td>
</tr>
<tr>
<td>+0.8 mg/dL</td>
<td>28.6</td>
<td>31.5</td>
</tr>
<tr>
<td>+1.0 mg/dL</td>
<td>18</td>
<td>24.1</td>
</tr>
</tbody>
</table>

### Total Protein, Ambulatory Pts

- **Original Protocol settings:**
  - Blue Diamonds
  - $N = 12$; $TLH = 9.1$ mg/dL; $TLL = 5.1$ mg/dL;
  - Mean = 7.2 mg/dL;

- **Optimized Protocol settings:**
  - Red squares
  - $N = 8$; $TLH = 9.6$ mg/dL; $TLL = 4.6$ mg/dL;
  - Mean = 7.1 mg/dL;

### Total Protein, Ambulatory Patients

- Protocol settings:
  - Original: (blue arrows)
    - $N = 12$
    - $TLH = 9.1$ mg/dL
    - $TLL = 5.1$ mg/dL
    - Mean = 7.2 mg/dL
  - Optimized: (red arrows)
    - $N = 8$
    - $TLH = 9.6$ mg/dL
    - $TLL = 4.6$ mg/dL
    - Mean = 7.1 mg/dL
• Original settings for Total Protein ambulatory patients protocol were a significant improvement on the all patients protocol
• Optimized setting again improved performance

Two Population Distributions
- Frequency Distribution of total protein data
  - Light blue=all patients
  - Orange=Ambulatory
  - Green=Inpatients
- Optimized protocol settings:
  - All: N = 37, TLH = 18.2 mg/dL, TLL = 2.4 mg/dL; Mean = 6.9 mg/dL
  - Amb: N=8; TLH=9.6 mg/dL, 4.6 mg/dL; Mean = 7.1 mg/dL
  - Inpt*: N=17; TLH=13 mg/dL, 2.0 mg/dL; Mean=6.4 mg/dL
  - Control limits - 1.0 mg/dL

Total Protein Performance
- Negative error detection in our original “all patients” total protein protocol was limited
- In the absence of ability to perform modeling I suggest separate inpatient and ambulatory protocols

<table>
<thead>
<tr>
<th>Total Protein, Ambulatory Patients</th>
<th>Original Protocol, FPR = 0</th>
<th>Optimized Protocol, FPR = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANPed</td>
<td>SD</td>
<td>Mode</td>
</tr>
<tr>
<td>-0.8 mg/dL</td>
<td>122.6</td>
<td>109</td>
</tr>
<tr>
<td>+0.8 mg/dL</td>
<td>66.8</td>
<td>71.2</td>
</tr>
</tbody>
</table>

Total Protein Performance
- Negative error detection in our original “all patients” total protein protocol was limited
- In the absence of ability to perform modeling I suggest separate inpatient and ambulatory protocols

<table>
<thead>
<tr>
<th>Total Protein, All Data</th>
<th>All Patients</th>
<th>Ambulatory</th>
<th>Inpatients</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANPed</td>
<td>SD</td>
<td>ANPed</td>
<td>SD</td>
</tr>
<tr>
<td>-0.8 mg/dL</td>
<td>8073</td>
<td>3076</td>
<td>122.6</td>
</tr>
<tr>
<td>+0.8 mg/dL</td>
<td>29.6</td>
<td>31.5</td>
<td>66.8</td>
</tr>
</tbody>
</table>
Potassium ANPed

- Significant shift
  - $= 0.5$ mmol/L
- Original Protocol
  - Blue Diamonds
  - $N=10$; $TLH = 6.1$ mmol/L; $TLL = 2.1$ mmol/L
- Modified original protocol ($N=50$)
  - Green Triangles
  - $N=50$; $TLH = 6.1$ mmol/L; $TLL = 2.1$ mmol/L
- Optimized protocol
  - Red Squares
  - $N=50$; $TLH = 6.8$ mmol/L; $TLL = 0.5$ mmol/L

Potassium Truncation Limits

- Performance of optimized and “best-guess” protocols similar
- Truncation limits quite different
- No appreciable differences in performance except for very large errors

Theoretical vs. Reality

- In Ye, Ingles and Parvin article a 0.5 mmol/L shift detected in ~25 samples
  - Random selection from a database – simulates an in control assay
- Our experience: a 0.5 mmol/L shift may be detected in between 70 – 90 samples
  - Real stream of data
    - Batches of inpatients followed by stream of ambulatory patients
### MA Benefits

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>AM</th>
<th>Inpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcium</td>
<td>ANPed SD</td>
<td></td>
<td>ANPed SD</td>
</tr>
<tr>
<td>-1.0 mg/dL</td>
<td>100.3</td>
<td>134</td>
<td>16.4</td>
</tr>
<tr>
<td>+1.0 mg/dL</td>
<td>32.8</td>
<td>37.5</td>
<td>17.5</td>
</tr>
<tr>
<td>Creatinine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.3 mg/dL</td>
<td>34.9</td>
<td>37.8</td>
<td>13.4</td>
</tr>
<tr>
<td>+0.3 mg/dL</td>
<td>44.7</td>
<td>35.4</td>
<td>36.5</td>
</tr>
<tr>
<td>Sodium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4.0 mg/dL</td>
<td>52.7</td>
<td>60.5</td>
<td>21.2</td>
</tr>
<tr>
<td>+4.0 mg/dL</td>
<td>46.7</td>
<td>59.8</td>
<td>23.6</td>
</tr>
<tr>
<td>Glucose</td>
<td>ANPed SD</td>
<td></td>
<td>ANPed SD</td>
</tr>
<tr>
<td>-10.5 mg/dL</td>
<td>156.5</td>
<td>113.3</td>
<td>63.8</td>
</tr>
<tr>
<td>+10.5 mg/dL</td>
<td>435.3</td>
<td>435.6</td>
<td>84.3</td>
</tr>
<tr>
<td>Total Cholesterol</td>
<td>ANPed SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-18 mg/dL</td>
<td>77.0</td>
<td>61.1</td>
<td></td>
</tr>
<tr>
<td>+18 mg/dL</td>
<td>74.8</td>
<td>47.1</td>
<td></td>
</tr>
<tr>
<td>TSH</td>
<td>ANPed SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.0 mIU/L</td>
<td>93.0</td>
<td>87.9</td>
<td></td>
</tr>
<tr>
<td>+1.0 mIU/L</td>
<td>50.0</td>
<td>40.2</td>
<td></td>
</tr>
</tbody>
</table>

**Final Challenges**

Don't Get Too Excited When a Protocol Errors out!
Analytical Shift?

Analytical Shift? – Ambulatory Pts

Analytical Shift? Multiple Modules
Summary – MA Not for All Tests

• Not every analyte is amenable to MA
  • If daily test volume < N needed for average point
    – no near real time data
  • Analytes with great variability
    • CK, LD
  • Analytes such as troponin
    • Essentially undetectable concentration for most
      patients with intermittent high values
      • May be more amenable to moving median analysis

Summary – Starting MA

• Can set up protocols without in-depth modeling
• Trial and error
  • Set up one or two protocols as outlined
  • If you see frequent false errors, adjust N or truncation
    limits, one at a time
  • If the data line “hugs” the mean line tightly –
    insensitive protocol
    • Widen truncation limits and/or narrow your control limits
• Some trial and error is inevitable, even with
  modeling

Final Thoughts

• MA – Continuous Assay Monitoring (real-time)
• Benefits: Earlier detection of analytic shift
• Hurdle: Data is “free” – just need to analyze it
• Remember: All protocols will have occasional
  false detection events
• Moving Averages is only a QA tool
  • How you choose to follow up on a shift is up to
    you
Acknowledgements

DHMC
• Frank Polito
• Middleware configuration
• David Ng
  • Matlab configuration

Data Innovations
• Hunter Bagwell
• Don Jarvis