Qualification and validation of analytical and bioanalytical methods

Yossi Levy

Method life cycle

Development → Qualification → Validation
References

- ICH guidelines for validation of analytical methods:
  - Q2A: Text on validation of analytical procedures
  - Q2B: Validation of analytical procedures: methodology
- ICH draft guideline 1033: Biological assay validation

Method development

- Exploratory process
- Early development phase – check possibility of developing the method
- Optimization phase – once development of method is possible, fine tuning of method’s parameters is needed for efficient implementation
- Examples of parameters: temperature, incubation time, type of equipment, etc.
- Statistical support is needed at the optimization phase
- Main statistical tool is DOE
- Usually, a series of controlled experiments is needed
Qualification and validation

- Qualification and validation are two steps in testing the performance of a (bio)analytical procedure/method and ensuring its quality.
- **Qualification**: A documented testing that demonstrates with a high degree of assurance that a specific process will meet its pre-determined acceptance criteria.
- **Validation**: A documented testing, performed under highly controlled conditions, which demonstrates a process consistently produces a result meeting pre-determined acceptance criteria.

What is the difference?

- Key difference: whether or not the process under review operates under 'highly controlled' conditions.
- Qualification can be viewed a less extensive form of validation.
- Less parameters are checked.
- Acceptation criteria are less strict.
- In some cases, qualification is part of the method development process. Method can be modified if necessary.
**The role of the statistician**

- To provide, in cooperation with the development team, the experimental design for the qualification/validation.
- To develop and write the statistical methods section or a statistical analysis plan as required for the qualification/validation.
- To analyze and report the qualification/validation results according to the predefined statistical methods.
- To review and approve the qualification/validation report.

---

**Selectivity/Specificity**

- Selectivity/Specificity - the ability of an analytical method to differentiate and quantify the analyte in the presence of other components in the sample

This includes:

- Identification – ensuring the identity of the analyte
- Purity – ensuring an accurate statement of the content of impurities of an analyte, i.e. related substances test, heavy metals, residual solvents content, etc.
- Assay (content or potency) - providing an exact result which allows an accurate statement on the content or potency of the analyte in a sample
### Accuracy

- The accuracy of a (bio)analytical procedure expresses the closeness of agreement between the value which is accepted either as a conventional true value or an accepted reference value and the value found.
- This is sometimes termed as “trueness”.
- Accuracy is related to systematic error or bias.

### Precision

- The precision of a (bio)analytical procedure expresses the closeness of agreement between a series of measurements obtained from multiple sampling of the same homogeneous sample under the prescribed conditions.
- Precision is related to noise or variation.
Accuracy vs. precision

Accuracy = Bias
Precision = Variance

Levels of precision

- **Repeatability** expresses the precision under the same operating conditions over a short interval of time.
- **Intermediate** precision expresses within-laboratories variations: different days, different analysts, different equipment, etc.
- **Reproducibility** expresses the precision between laboratories.
Other quality parameters

- Detection limit - the lowest amount of analyte in a sample which can be detected but not necessarily quantified as an exact value
- Quantification limit - the lowest amount of analyte in a sample which can be quantitatively determined with suitable precision and accuracy
- Linearity – the ability (within a given range) to obtain test results which are directly proportional to the concentration (amount) of analyte in the sample

Other quality parameters

- Range - the interval between the upper and lower concentration (amounts) of analyte in the sample (including these concentrations) for which it has been demonstrated that the analytical procedure has a suitable level of precision, accuracy and linearity
- Robustness – measuring the method’s capacity to remain unaffected by small, but deliberate variations in method parameters
"The way they do it at Chemistry “:

1. Measure accuracy and repeatability using 6 runs by the same analyst on the same day – report CV.

2. Measure reproducibility using another 6 runs by another analyst on another day – report "Reproducibility Difference "

"The way they do it" advantage

- No experimental design
- No modeling
- No complex calculations
- Simple reporting
"The way they do it" problems

- Biological methods are more complicated to implement, therefore the numbers of possible runs in a single day is limited
- Variation of biological methods is generally higher compared to chemical methods
- Measuring intermediate precision is not enabled
- No statistical sense

"The way they do it" Experimental Design
### Example - Biological data

<table>
<thead>
<tr>
<th>Day/ Analyst</th>
<th>Run 1</th>
<th>Run 2</th>
<th>Run 3</th>
<th>Mean</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.768</td>
<td>0.601</td>
<td>0.887</td>
<td>0.752</td>
<td>0.144</td>
</tr>
<tr>
<td>2</td>
<td>0.460</td>
<td>0.398</td>
<td>0.519</td>
<td>0.459</td>
<td>0.061</td>
</tr>
</tbody>
</table>

Accuracy = \(100 \cdot \frac{0.752}{0.7} = 107.4\%\)

Repeatability = \(100 \cdot \frac{0.144}{0.752} = 19.1\%\)

Reproducibility Difference = \(100 \cdot \frac{|0.752 - 0.459|}{0.752 + 0.459} = 48.4\%\)

---

### The way we would do it, at Statistics – Mixed Model!

\[ Y_{ij} = \mu + b_i + c_j + \varepsilon_{ij} \]

Signal = fixed parameter + random effects + random error

Assumptions:
- Independence
- Normal distribution
- Zero mean deviations
- STDs: \(\sigma_b, \sigma_c, \sigma\)
Why use Mixed Models?

- Classical statistics assumes that observations are independent and identically distributed (iid)
- Often, data have a clustered structure
- When applied to clustered data, iid assumption may lead to false results
- Mixed Effects Model treats clustered data assumes two sources of variation, within cluster and between clusters
- This is the typical situation in biological data, when, observations are of the same biological category but individuals differ

Basic principles

- Two types of coefficients are distinguished in the mixed model
  - population-averaged: same meaning as in classical statistics
  - Cluster/subject-specific: random; estimated as posteriori means
Formal modeling

\[ Y = X\beta + Z\gamma + \varepsilon \]
\[ \gamma \sim N(0, G) \]
\[ \varepsilon \sim N(0, R) \]
\[ \text{cov}(\gamma, \varepsilon) = 0 \]

The matrices \( G \) and \( R \) are covariance matrices for the random effects and the random errors, respectively. As a result:

\[ V(Y) = V = ZGZ' + R \]

The trick is to find a good model for \( G \).

SAS syntax

```sas
data example1;
  input day y @@;
  cards;
  1 0.768 1 0.601 1 0.887
  2 0.460 2 0.398 2 0.519
; run;
proc mixed method=rem1 covtest cl;
  class day;
  model y= / solution cl;
  random day;
run;
```
**Example - SAS output**

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Z Value</th>
<th>Pr Z</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>day</td>
<td>0.0388</td>
<td>0.06077</td>
<td>0.64</td>
<td>0.2612</td>
<td>0.05</td>
<td>0.007062</td>
<td>175.52</td>
</tr>
<tr>
<td>Residual</td>
<td>0.01215</td>
<td>0.008592</td>
<td>1.41</td>
<td>0.0786</td>
<td>0.05</td>
<td>0.004362</td>
<td>0.1003</td>
</tr>
</tbody>
</table>

**Solution for Fixed Effects**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.6055</td>
<td>0.1465</td>
<td>1</td>
<td>4.13</td>
<td>0.1511</td>
<td>0.05</td>
<td>-1.2560</td>
<td>2.4670</td>
</tr>
</tbody>
</table>

**Results that make biological sense**

\[
\text{Accuracy} = 100 \cdot \frac{0.6055}{0.7} = 86.4\%
\]

\[
\text{Repeatability} = 100 \cdot \frac{\sqrt{0.01215}}{0.6055} = 18.2\%
\]

\[
\text{Reproducibility} = 100 \cdot \frac{\sqrt{0.03887 + 0.01215}}{0.6055} = 37.3\%
\]
### Results that make statistical sense

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy $\mu$</td>
<td>0.6055</td>
<td>-1.2560 - 2.4670</td>
</tr>
<tr>
<td>Repeatability $\sigma$</td>
<td>0.0122</td>
<td>0.004362 - 0.1003</td>
</tr>
</tbody>
</table>
| Between Days precision $\sqrt{\sigma^2 + \sigma^2}$ | 0.2259 | ????

### DOE to measure intermediate precisions

- **4 Days, 2 Analysts**
- **2 Days, 4 Analysts**
**Reporting intermediate precisions**

Accuracy = \(100 \cdot \frac{\mu}{\mu_0}\)

Between Day Precision = \(100 \cdot \frac{\sqrt{\sigma^2_{\text{Day}} + \sigma^2}}{\mu}\)

Between Analyst Precision = \(100 \cdot \frac{\sqrt{\sigma^2_{\text{Analyst}} + \sigma^2}}{\mu}\)

Repeatability = \(100 \cdot \frac{\sigma}{\mu}\)

CI for \(\mu, \sigma\) = \(\sqrt{\sigma^2_{\text{Day}} + \sigma^2}, \sqrt{\sigma^2_{\text{Analyst}} + \sigma^2}\)

---

**Example 2**

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Z Value</th>
<th>Pr &gt;</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyst</td>
<td>0.000548</td>
<td>0.000796</td>
<td>0.69</td>
<td>0.2455</td>
<td>0.05</td>
<td>0.000106</td>
<td>0.8027</td>
</tr>
<tr>
<td>Day</td>
<td>0.002250</td>
<td>0.002132</td>
<td>1.21</td>
<td>0.1139</td>
<td>0.05</td>
<td>0.000821</td>
<td>0.03765</td>
</tr>
<tr>
<td>Residual</td>
<td>0.000123</td>
<td>0.000057</td>
<td>2.398</td>
<td>0.0010</td>
<td>0.05</td>
<td>0.000102</td>
<td>0.000377</td>
</tr>
</tbody>
</table>

Solution for Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.5621</td>
<td>0.03045</td>
<td>1</td>
<td>18.46</td>
<td>0.0344</td>
<td>0.05</td>
<td>0.1753</td>
<td>0.9490</td>
</tr>
</tbody>
</table>
Relative Standard Deviation

- Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ iid. Define

\[ CV = \frac{\sigma}{\mu} \quad \text{and} \quad RSD = \frac{s}{\bar{X}} \]

- McKay derived the approximate distribution of RSD in 1932:

\[ f_b(t) = \frac{n}{2} \frac{\Gamma(n/2)}{\Gamma(n)} \int_0^\infty x^{n-1} e^{-x/2} \left[ \frac{t^2}{2} \frac{1}{\sqrt{\pi}} \Gamma \left( \frac{n-1}{2} \right) \right] \]

- This can be used to obtain CI for CV, but would one extend that to Mixed Models?

Jackknife

- Idea: systematically re-computing the statistic estimate leaving out one or more observations at a time from the sample set

- From this new set of replicates of the statistic, an estimate for the bias and an estimate for the variance of the statistic can be calculated

- If we delete one observation at a time we get n subsamples

- Then we calculate estimate CV out of the n subsamples, and obtain an estimate for it variation

- This estimate can be used to obtain a CI
Fieller’s theorem application

- Same trick as in Fieller’s theorem – look at
  \[ U = s - CV \cdot \bar{X} \]

- Then
  \[ V(U) = V(s) + CV^2 \cdot \frac{\sigma^2}{n} = \frac{\sigma^2}{2(n-1)} + CV^2 \cdot \frac{\sigma^2}{n} \]

- The obtained CI is
  \[ (0.100 \times (\bar{x}^2 - \frac{s^2}{n})\bar{x}\bar{s} + \sqrt{\bar{x}^2 s^2 - (\bar{x}^2 - \frac{s^2}{n})s^2 (s^2 - \frac{s^2}{n})/n(s^2 - \frac{s^2}{n})/2(n-1))) \]

Delta method

- Let \( T_n \) be a MLE of a (multidimensional) parameter \( \theta \).
- It is known that \( T_n \) is asymptotically Normally distributed:
  \[ \sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \Sigma) \]

- Consider a function \( h(\theta) \). We can expand its according to Taylor:
  \[ h(T_n) \approx h(\theta) + \nabla h(\theta) \cdot (T_n - \theta) \]
**Delta method**

\[
V[h(T_n)] \approx V[h(\theta) + \nabla h(\theta)'(T_n - \theta)] = \\
= V[h(\theta) + \nabla h(\theta)'T_n - \nabla h(\theta)'\theta)] = \\
= V[\nabla h(\theta)'T_n] = \nabla h(\theta)'V[T_n] \cdot \nabla h(\theta) = \\
= \frac{1}{n} \nabla h(\theta)'\Sigma \cdot \nabla h(\theta)
\]

**Application of Delta method**

- In our framework:
  \[
  \theta = (\mu, \sigma^2)
  \]
  \[
  T_n = (\bar{X}_n, S_n^2)
  \text{ where } S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X}_n)^2
  \]
  \[
  h(x, y) = \sqrt{y} / x
  \]

- This leads to the following CI:
  \[
  \left(0.100 \times \left[ \frac{s_n}{\bar{x}} + z_{0.95} \sqrt{\frac{s_n^4}{x^4n} + \frac{s_n^2}{2x^2n}} \right] \right)
  \]
Parametric bootstrap

- Estimate model parameters
- Simulate N new datasets based on estimated parameters
- Estimate parameter under interests for each of the simulated datasets to get a sample of N simulated estimates
- Use 2.5% and 97.5% sample quartiles as a CI

Note: for RSD, we use the 95% quartile as an upper confidence limit, since the lower confidence limit is zero.
Which method should we use?

- We should consider
  - Distributional assumptions – are they correct? Are they needed?
  - Robustness
  - Ease of computation
  - “back calculation” – Can we calculate sample sizes?