Simplified strategy for Immunogenicity cut-point evaluations & some practical considerations

V. Devanarayan, Ph.D.
AbbVie Inc., USA

European Immunogenicity Platform (EIP) Workshop
February 24-26, 2014
Recommendations presented here were influenced by the following collaborations:

Wendell Smith & Ron Bowsher, B²S Consulting

DMPK-BioAnalysis colleagues at AbbVie.

Coauthors of the AAPS white-papers & USP chapters:

• Mire-Sluis et al., 2004, JIM (ADA screening - design elements)
• Koren et al., 2007, JIM (ADA testing strategy)
• Shankar et al., 2008, JPBA (ADA screening – method validation)
• Gupta et al., 2011, JPBA (NAb)
• USP chapter on Immunogenicity screening methods, 2013
• USP chapter on Neutralizing Antibody methods, 2014
Outline

Background *(will skip this as Dr. Shankar has already covered it)*

Screening Cut-Point

Confirmatory Cut-Point

Titer Cut-Point

Some Practical Considerations
ADA Four-tiered Testing Strategy

- **Tier 1:** Identify “reactive” samples
  - Samples with signal above screening cut-point (SCP)

- **Tier 2:** Identify “Ab+” samples by testing reactive samples in the absence and presence of drug
  - Samples with percent inhibition above confirmatory cut-point (CCP)

- **Tier 3:** Determine a sample titer value by serial dilution of Ab+ samples in Tier 2
  - Titer is based on the SCP or a higher “titer cut-point” (TCP). Can be continuous (requires interpolation) or discrete

- **Tier 4:** Evaluate neutralizing effects of antibodies
  - Usually based on cell-based bioassay using Ab+ samples
**Screening CP Evaluation, Shankar et al, 2008**

Data: ~ 50 drug naïve subjects, >= 3 runs (2 analysts)

Investigate Distribution

Non-normal

Transform data (usually log)

Outlier evaluation

Normal

Confirm Distribution

Non-normal

Normal

95th percentile

Mean+1.645*SD or Robust alternative

Validation cut-point (CP.V)

Compare means and variances between runs/instruments/analysts

Means similar

Variance similar

Fixed cut-point (CP.V) per instrument

Fixed cut-point (CP.V) per instrument

Determine CP in each in-study run

Means different

Variance different

Calculate CP.V and CF per instrument

Instrument or Analyst specific floating CP

Determine CP in each in-study run

NC.V = Neg. Control from Validation runs
NC.IS = Neg. Control from In-Study run

Screening cut-point

NC.I*S*(CP.V/NC.V), if log
NC.IS+(CP.V – NC.V), if not
Screening CP Evaluation
- a simpler flow-scheme that works in most cases

Analyze negative control data, identify and exclude outliers → Samples from > 50 drug naïve negative >= 3 runs (2 analysts)

Normalize the data as ratio of signal to neg. control (S/N). All further analysis on log(S/N) scale → Evaluate SCP factor

Identify & exclude analytical & biological outliers, then reevaluate distribution →

If S-W p < 0.005 or |Skewness| > 1*

Nonparametric method: 95th percentile

If |Skewness| < 1*

Parametric method: Mean+1.645xSD (or a robust alternative if S-W p < 0.05)

Alternative transformations may be used if needed.

Evaluate inter-plate/run and Inter-analyst mean & variance differences. Also evaluate other sample factors (e.g., demographic, disease subtype, etc.)

Evaluate relevant sample factors (disease subtype, gender, age, ethnic, …)

Verify suitability of CP for other patient populations, and clinical study samples

Evaluate negative control correlation with subject sera

* Skewness test can be used instead; Brown & Hettmansperger, JASA, 1996
Typical considerations for a cut-point experiment

~ 50 drug naïve ADA negative subjects
  • Preferably from the target disease population if available.
  • Should represent relevant demographic subgroups (gender, race, age, etc.).
  • Multiple disease subtypes can be included, to investigate common or separate cut-points (e.g., n>20 for each cancer type).

Test these subject samples in 6 runs, by 2 analysts (3 runs per analyst)
  • If multiple analysts will test study samples

Each sample tested in duplicate
  • Reportable result: Average of duplicate samples divided by NQC.

Negative QC: 2-3 reportable results/plate, each in duplicate, and located in different parts of the plate.

Low QC and High QC: >=2 reps/plate, each in duplicate

Three plates used per run for testing these samples.

Include drug-spiked samples as well for confirmatory cut-point. Sensitivity, Precision, Confirmation CP, Titration CP, etc., can all be evaluated from this expt.
Balanced design (Shankar et al., 2008)

<table>
<thead>
<tr>
<th>Analyst</th>
<th>Assay Run</th>
<th>Assay Plate</th>
<th>Validation Serum Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R₁</td>
<td>P₁</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₂</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₃</td>
<td></td>
</tr>
<tr>
<td>A₁</td>
<td>R₂</td>
<td>P₁</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₂</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₃</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R₃</td>
<td>P₁</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₂</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₃</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>R₄</td>
<td>P₁</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₂</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₃</td>
<td>X</td>
</tr>
<tr>
<td>A₂</td>
<td>R₅</td>
<td>P₁</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₂</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₃</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>R₆</td>
<td>P₁</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₂</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P₃</td>
<td></td>
</tr>
</tbody>
</table>

All samples get tested in every run and every plate, by both analysts.
**Statistical modeling approach for outlier evaluation**

1. Fit a mixed-effects model on the normalized response.
   - **Random effects**: Subjects nested within Subject Groups, Run number nested within Analyst, and Plate ID.
   - **Fixed effects**: Subject Groups, Analyst, Plate testing order and the interaction of Analyst and Plate testing order (+ gender, disease types, demographics, ..., as appropriate).

2. Obtain **conditional** residuals from this model.
   - Difference between the observed and predicted values that includes random subject effect (*reflects only measurement error*).
   - Readily available from statistical programs such as JMP.

3. Use the “outlier box-plot” criteria to identify outliers from the conditional residuals ➔ These are **Analytical outliers**.

4. Iterate steps 1-3 until all analytical outliers are removed.
**Statistical modeling approach for outlier evaluation (contd.)**

5. Refit the model without these analytical outliers, and then obtain Best Linear Unbiased Predictor (BLUP) for each subject.

6. Use the “outlier box-plot” criteria to identify & exclude outliers from subject BLUPs. → These are **Biological outliers**.

7. Iterate steps 5-6 until all biological outliers are removed.

8. Refit the model without these outliers. Obtain **marginal** residuals.
   - Difference between the observed & mean predicted values; *reflects both subject random effect & measurement error.*

9. Assess normality & symmetry of these marginal residuals
   - Shapiro-Wilk test & Skewness Test (Brown & Hettmansperger, JASA, 1996)

**Outlier box-plot criteria:** Samples \( > Q3 + 1.5*(Q3-Q1) \) or \( < Q1 - 1.5*(Q3-Q1) \)

\( Q3 = 75^{th} \) percentile, \( Q1 = 25^{th} \) percentile
Illustration of outlier evaluation with statistical modeling approach

Analytical Outliers
Conditional Residuals from mixed-effects model

<table>
<thead>
<tr>
<th>-0.15</th>
<th>-0.1</th>
<th>-0.05</th>
<th>0</th>
<th>0.05</th>
<th>0.1</th>
<th>0.15</th>
</tr>
</thead>
</table>

| Graph of Conditional Residuals |

Biological Outliers
Subject BLUPs from mixed-effects model after removing analytical outliers

<table>
<thead>
<tr>
<th>-0.1</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
</tr>
</thead>
</table>

| Graph of Subject BLUPs: Log (signal to noise ratio) |

- Mixed effects model (as described earlier) is fit on the ratio of individual samples to the negative control. Conditional Residuals are evaluated.

- Analytical outliers are first identified and excluded. After iteratively excluding all the analytical outliers, the biological outliers are evaluated.
Simpler alternative for outlier evaluation that may be used if the statistical modeling approach is not feasible

- Identify & exclude analytical outliers (AO) from each assay run/plate separately. Iterate until no more AO.
- Then identify & exclude biological outliers (BO) by evaluating the distribution of subject averaged data. Iterate until no more BO.
- Then verify distribution of subject averaged data. Use the flow-scheme to decide on the appropriate cut-point factor calculation.

• This usually yields similar cut-points as those from the statistical modeling method described in previous slides (esp., if robust approaches are used).
SD evaluation for screening cut-point

SD should incorporate all variance components (total variation) expected during sample testing. Can be calculated easily using Excel.

After excluding all the outliers, calculate SD of all the data from the validation experiment.

• Suppose 50 subjects were tested in 6 runs (300 samples).
• If there were 20 outlier samples, then calculate SD of the remaining 280 samples directly from Excel using STDEV function.

In most cases, this is quite similar to the more rigorous calculation of variance components from random-effects ANOVA.

• Random-effects ANOVA helps understand the relative contribution of different variance components.

Use MAD instead of SD and Median instead of Mean, if tails are long after outlier exclusion (see flow-scheme).
Confirmatory cut-point (CCP) – Evolution

Arbitrary thresholds (e.g., 50%) were widely used in the past (< 2006).


Alternative approaches were discussed for several months by a focus group of industry and FDA scientists.

- E.g., using individual mock low-positive samples to define a lower 99.9% limit as the CCP.

*Conclusions from this discussion (Smith et al, 2011):*

- Shankar et al (2008) method can be the default, as it works well in most cases.
- Alternative approaches may be tried in some special cases.
Confirmatory Cut Point (CCP) Evaluation
- a simple flow-scheme that works in most cases

Samples from ~ 50 subjects, spiked with excess drug, preferably in the same plate as the unspiked counterpart in SCP experiment, >= 3 runs (2 analysts)

- Identify analytical & biological outliers in %inhibition data
- Exclude outliers, evaluate distribution

- If S-W p < 0.005 or |Skewness| > 1*
  - Transform to reduce skewness (e.g., log)
  - Nonparametric method: CCP = 99.9th percentile

- If |Skewness| < 1*
  - Parametric method: CCP = Mean+3.09xSD (or robust alternative, if S-W p < 0.05)

- Evaluate inter-plate/run and Inter-analyst mean & variance differences. Also evaluate other sample factors (e.g., demographic, disease subtype, etc.)

- Verify suitability of CP for other patient populations, and clinical study samples

- Evaluate correlation between mean log(S/N) of unspiked sera vs. mean % inhibition for each subject

- If log transformation is needed, analyze log(s/us) due to negative inhibition

* Skewness test can be used instead; Brown & Hettmansperger, JASA, 1996
**Titration cut-point (TCP)**

Screening cut-point may fall on the lower plateau of the positive control dilution curve.

- This will result in highly noisy/variable titers.

In such cases, use a higher cut-point for evaluating titers (Titration CP)

- Using the *same data from the screening cut-point experiment*, calculate $TCP = \text{Mean} + 3.09\times\text{SD}$ or $\text{Mean} + 6\times\text{SD}$

- $3.09$ corresponds to $\sim 99.9^{th}$ percentile.

\[ \text{Titer} = \text{MRD for confirmed positives that fall between SCP & TCP}. \]

Other methods based on only negative control may be considered.

- NB: Purpose of Titer CP is fairly simple, so extensive research and additional calculations don’t add much value.
Some Practical Considerations
Validity of Negative Control for SCP factor evaluation

Floating SCP or SCP factor assumes that NC drifts in the same direction as individual subject samples.

• i.e., assumes that NC is correlated with subject sera.

This can be formally justified using validation data.

• Plot the NC mean versus mean of subject sera from each run/plate.
  — Evaluate Slope & Rank Corr. (Need Slope ~ 1, Correlation > 70%).

If this assumption fails, using NC for Floating-CP may not be helpful.

• More likely when analytical variability exceeds biological variability.

Alternatives: New pool, Subject-specific cut-point, other controls (same disease/demographic), ....

Need 3 reportable results of NC, located in different parts of the plate.
Testing the validity of Negative Control (contd.)

**Example-1**

![Graph showing linear fit and yes/no decision](image1)

**Example-2**

![Graph showing linear fit and yes/no decision](image2)

**Example-3**

![Graph showing linear fit and yes/no decision](image3)
Can the same CP or CF be used for a different disease/target population?

Are the variances similar between the two groups? (use Levene’s test)

- **Yes**
  - **Yes**
    - Same CP/CF can be used.
  - **No**
    - Need to derive a new CP/CF

- **No**
  - **Yes**
    - Need new Negative Control (NC) pool.
    - Use floating CP, with CF from previous population.
    - Apply it to the *new NC pool* during in-study.
In-study justification of validation cut-points
Why & When?

Why?
• SCP factor during validation is set to yield \(\sim 5\%\) false positives. But this is an estimate. As with any estimate, there is variability.
• Based on our Monte-Carlo simulations assuming a typical dataset from a balanced design of Shankar et al (2008), *false positive rate* can vary between 3 to 14\% during the in-study sample testing phase.

When?
• Therefore as a general guideline, \(\geq 15\%\) *false positives in clinical baseline samples should trigger follow-up evaluations*.
• Also necessary if the validation and in-study populations are different (e.g., healthy vs. disease, differences in disease & demog.)
In-study justification of validation cut-points (contd.)

Why and When? (example)

- SCP factor from pre-study validation is 1.15.
- Variability of in-study clinical baseline data is significantly higher, p<0.0001 (Levene’s test).
- Results in > 30% false positives in clinical baseline samples.
- Follow-up evaluation using in-study clinical baseline data is needed.

How to justify pre-study validation cut-points with in-study data, and what are the corrective actions?
In-study justification of validation cut-points (contd.)

How?

Compare the distribution (mean and variance) of validation data versus in-study clinical baseline data.

If the variances are different (Levene’s test), need to derive a new cut-point using the in-study clinical baseline data.

- However, if clinical baseline data are small (n < 25 subjects from < 2 runs), redo the validation experiment with subjects similar to clinical study population.

If only the means are different, same validation cut-point can be used after redefining the negative control based on disease population (for floating CP).

Apply similar criteria to decide whether to reevaluate the Confirmatory CP using the drug spiked baseline samples.
Summary

• As SCP is usually “Floating”, normalize by NQC, evaluate SCP factor.
  – Assumes NQC drifts with subject sera. This should be verified.
• Multiple populations can be tested in the same balanced design.
• Outlier evaluations are critical (use mixed-effects ANOVA or simpler alternatives).
• SD evaluation should include relevant variance components (inter-analyst, inter-run, intra-run, inter-subject, etc.)
• Titration CP can be defined at higher limits using Screening CP data.
• Cut-points from validation should be justified with in-study clinical baseline data.
• Also, should justify use in other sample types (disease, demog, etc.).