

Practical Advice on When and How to Perform In-Study Cut-points

V. Devanarayan, PhD, FAAPS

Eisai Inc.

EIP Training Couse, April 24, 2023

human health care

Outline



- Cut point calculation process (pre-study validation)
- **Outlier** criteria and identification
- In-study cut points
- Overview and demo of **Excel tool** for cut-point calculations
- Other topics, if time permits
 - "Low" cut points and "Low" signal
 - Titer cut point and Titer precision
 - Treatment-boosted ADA



Cut point calculation process (pre-study validation)

Balanced design



Alternative visual

		Analyst 1	L	Analyst 2			
Plate Order	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	
1	Group A	Group B	Group C	Group C	Group B	Group A	
	Plate 1	Plate 1	Plate 1	Plate 1	Plate 1	Plate 1	
2	Group B	Group C	Group A	Group A	Group C	Group B	
	Plate 2	Plate 2	Plate 2	Plate 2	Plate 2	Plate 2	
3	Group C	Group A	Group B	Group B	Group A	Group C	
	Plate 3	Plate 3	Plate 3	Plate 3	Plate 3	Plate 3	

All samples get tested in <u>every run</u> and <u>every plate</u>, by <u>both analysts</u>.





Plate layout for the balanced design

_	1	2	3	4	5	6	7	8	9	10	11	12	
A NC		S2		S6		S10		S13		S17		А	
В	B NC + drug		S2 +	drug	S6 + drug		S10 + drug		S13 + drug		S17 + drug		В
С	C LPC		S	S3 S7		S11		S14		NC		С	
D	D LPC + drug		S3 + drug		S7 + drug		S11 + drug		S14 + drug		NC + drug		D
E MPC		S4		S8		NC		S15		LPC		E	
F	F HPC		S4 +	drug	S8 + drug		NC + drug		S15 + drug		LPC + drug		F
G	G S1		S	5	S 9		S12		S16		MPC		G
Н	H S1 + drug		S5 +	S5 + drug S9 + drug		S12 + drug		S16 + drug		НРС		Н	
	1	2	3	4	5	6	7	8	9	10	11	12	

- NC & LPC with and w/o drug, MPC & HPC
- 17 subject sera per plate.

Screening Cut Point (SCP)

Pre-study validation



robustness when there are borderline outliers.

LCL: Lower confidence limit

Devanarayan et al., 2017

Confirmatory Cut Point (CCP)

Pre-study validation



LCL: Lower confidence limit

Devanarayan et al., 2017

SCP & CCP results from the standard vs. LCL formulae



- CP values from the standard method (point estimate; PE) & lower confidence limit (LCL) formulae are practically similar for most assays.
- As expected, LCL values are slightly lower, and thus the FPRs are slightly higher.
- Differences are more visible for smaller sample sizes (e.g., early-phase in-study samples)
- 7



Outlier criteria and identification

Interpretation of 1.5xIQR and 3xIQR outlier criteria



For normal distribution,

1.5 x IQR is approx. Mean ± 2.67 x SD

- covers ~ 99.2% of the samples
- ~ 3xSD rule that is widely used in other applications.

3 x IQR is approx. Mean \pm <u>4.67</u> x SD

• ~ 99.9997% of the samples

When most scientific applications use 2xSD or 3xSD rule, why apply 4xSD or 5xSD rule for Immunogenicity?



Median

- Median = "middle value" of a distribution.
- Less skewed by high and low outlier values
- More robust than Mean, when outliers are present.
 - > That is, less affected by borderline outliers.
- Useful for skewed distributions
- These characteristics make the Median more appealing for use in computations of ADA cut points

Median Absolute Deviation (MAD)

- A robust measure of the <u>variability</u> of continuous measurements.
- Equals the median of the absolute deviations from the median:
 - MAD = median(|Xi median(x)|)
- SD is replaced in the formula by 1.4826 x MAD
- More resistant to the outliers than SD
- Borderline outliers do not significantly impact the MAD (i.e., no need to debate about 1.5xIQR vs 3xIQR)
- Thus, Median and MAD are safer alternatives to Mean and SD, when borderline outliers are present.
 - Robust to subjective manipulations of outlier criteria.

Illustration: Robustness to outliers

For the sake of illustration, we use 20 S/N values from the Screening CP experiment.



→ 0.497 =| 0.514 – median(log(S/N)) |

MAD = Median of all these absolute deviations

	Parar	netric	Robust Par.			
	SD	SCP	MAD*	SCP		
All Data	0.152	2.094	0.046	1.225		
w/o 2 outliers (<u>3 x IQR</u>)	0.062	1.344	0.043	1.204		
w/o 5 outliers (<u>1.5xIQR</u>)	0.028	1.120	0.034	1.133		
MAD* = 1.4826 x MAD						

"Robust parametric" is more resistant to borderline outliers. 1.5xIQR vs. 3xIQR makes no difference

12



1.5xIQR vs. 3xIQR criteria: data from 25 assays



- As expected, the 1.5xIQR identifies several more outliers than the 3.IQR criteria.
- Parametric (Mean/SD) method is skewed by outliers not caught using 3xIQR criteria.

13



1.5xIQR vs. 3xIQR criteria: data from 25 assays



- If "robust parametric" is used, CP values from 1.5xIQR and 3xIQR are practically similar.
- 3xIQR criteria is not the cure for "low" cut points!
- 1.5xIQR is ~ 2.7xSD, it is widely used in the statistics literature, results in a similar CP when using the robust method, and is, therefore, a good default.



In-study cut points



In-study cut points; when?

In-study FPR deviating considerably from 5% implies that the pre-study validation samples are not representative of the study samples

- Differences in demographics, disease, or clinical characteristics.
- Changes in reagent quality or other analytical factors

From practical experience and different publications, in-study FPR ranging from 2% to 11% is generally considered reasonable, after excluding samples with preexisting Ab. *Myler et al. (2021), Devanarayan et al. (2017), Amaravadi et al. (2015)*

Different simulation approaches have yielded different ranges. This criterion is mostly *driven by practical considerations*.



In-study cut points; when?

In-study FPR < 2% usually raises concerns, requiring CP re-evaluation.

• For high-risk products, in-study FPR closer to 4-5% may be expected.

In-study FPR > 11% raises concerns mostly for the sponsor, but it depends on the nature of the assay and clinical study.

- For small studies (e.g., phase-I), this might not be a concern.
- For larger studies (e.g., phase III), this is a major concern as it leads to unnecessary additional lab work, reagent wastage, reporting delays, etc.
- This may also be a concern for assays with "borderline sensitivity".
 - Lower CP overestimates sensitivity.

In-study cut points; diagnostics

Visual and statistical assessment may provide some insights

- Box plots of the distribution of S/N values between pre-study and in-study samples
- Comparison of means and variances of the log(S/N) data between pre-study and in-study samples, via ANOVA, after excluding outliers.
- If multiple disease populations were tested (e.g., Oncology), a comparison of the distribution across these populations will also be useful.





In-study cut points; design and data requirements

Use all the available pre-dose subject sera

• Ideally, > 50 subjects for phase-II studies and > 100 subjects for phase III studies

Balanced design and multiple testing of each subject sera are not needed.

- Assay characteristics have been established during pre-study validation.
- The main/only objective is to calculate the *in-study cut point*.
- For this, we need an estimate of the total variability and background signal.

One test/reportable result per subject will suffice.

Need to spread the subject sera *across* >= 3 *plates/runs and* >= 2 *analysts*.

Total variability from these data will reflect all the relevant sources of variability

• Analyst, inter-run, intra-run, etc.

In-study cut points; design and data requirements (contd.)

If only a subset of pre-dose subject sera are used, for e.g., in phase III studies, not all pre-dose subject sera may be tested at the same time,

- Stratify with respect to key patient demographic and clinical characteristics, to ensure the study population is well represented.
- Confirm the in-study cut-point results with additional samples later, if possible.
 - Verify / Confirm the FPR of the next batch of pre-dose samples.
 - Re-evaluate / Update the in-study cut point with the additional samples, if needed.

For rare disease or pediatric studies where samples are limited:

- Use SCPF as the starting point until more pre-dose samples are accrued.
- Alternatively, skip the screening phase, and tests all samples in the confirmatory assay.

In-study Screening Cut Point; analysis

Apply the same considerations from the pre-study SCP calculation process

- Assess distribution, identify and exclude outliers, etc.
 - Due to not using a balanced design, and with only one result per subject serum, it is not possible to estimate variances separately for the analytical, biological, and other operational components.
 - Total variability captures all the relevant components.



* Use Median instead of Mean, and 1.4826xMAD instead of SD to ensure robustness when there are borderline outliers.

In-study Confirmatory Cut Point; design

As the CCP is set at 1% FPR, a practically acceptable range for the in-study FPR of say 0.5% to 3% can be derived/defined. Sample size limitations and outliers may pose practical challenges (due to such a low percentage)

Therefore, if in-study SCP is needed, in-study CCP should be considered.

Design:

- Spike all the in-study pre-dose samples that were used for the in-study SCP calculation with excess study drug. A subset may be considered, for large clinical studies such as phase-III.
- Test the unspiked and spiked counterpart samples together on the same plate (similar to prestudy validation).
- As noted for the in-study SCP evaluation, do not need a balanced design; one test result per subject serum will suffice for this in-study CCP calculation.



In-study Confirmatory Cut Point; analysis



* Use Median instead of Mean, and 1.4826xMAD instead of SD to ensure robustness when there are borderline outliers.





- Although distributions look very similar, means & variances are significantly different, mostly due to large "n"
- FPR = 3.8%
- Don't need to use in-study CP









- Variances are <u>not</u> significantly different.
- But means are significantly different.
- FPR = 0%
- Need to use in-study CP



- Variances are significantly different
- **FPR = 30%**, after excluding pre-existing Ab
- Need in-study CP



Overview & demo of the Immunogenicity Cut Point Analysis Tool (I-CAT)



Some Key Points about I-CAT

- Provides preliminary estimates of screening, confirmatory, and titer cut points based on data from typical Immunogenicity cut point experiments (pre-study validation & in-study cut points).
- Due to limitations with Excel, the stats methods implemented for outlier identification are not based on the rigorous mixed-effects model-based approach recommended by Devanarayan et al (2017).
- It is based on a simplified alternative proposed in that same paper that is more amenable for use in Excel and related applications.
 - Although the outliers may not exactly overlap, the cut point results are usually very similar to the more rigorous method, especially when the "robust parametric" method is used.
- This tool is meant for only exploratory/informal use during the method development and prevalidation phase or for preliminary analysis of validation data or in-study pre-dose sample results, prior to the more formal analysis via a validated and specialized statistical program such as SAS or R.

File management

30



- Save the XLT file in the "Custom Office Templates" folder under "My Documents". In addition, save this XLT file in a project folder.
- Use only the XLS spreadsheet copy of this tool when analyzing new dataset.
 - Do not analyze data directly in the original XLT file. Keep the data input space empty in the XLT file.
- To open the XLS copy, click on File -> New in Excel. This file will be listed under the "Personal" tab. Click on it to open the XLS spreadsheet copy.
 - Alternatively, double-clicking on the file name of the original XLT file in windows explorer will also open the XLS copy of this tool, but this tends to take a longer time to open.
- Do not reuse a previously used copy for the analysis of a new dataset. Open a new XLS copy of the original XLT file for the analysis of each new dataset.



Data analysis instructions

- Enter the data in "Input.Data" worksheet.
 - This is the only worksheet that requires user input. All cut point evaluations will be automatically carried out in the next few worksheets.
 - Data for pre-study validation cut points should ideally come from a balanced experimental design, as recommended by Shankar et al (2008) and Devanarayan et al (2017).
 - This tool can handle unbalanced design as well, however, note that certain factors may be confounded.
 - In-Study screening, confirmatory, and titer cut points can also be calculated using this tool. Balanced design is not required. One test result per subject will suffice, but we recommend spreading the subject sera across >=3 runs and >=2 analysts so that all sources of variability are captured in the cut point calculations.
 - The data table can accommodate up to 1000 rows.



Data analysis instructions

- Proceed to the "SCP-analysis" worksheet to view intermediate analysis results for screening cut point (SCP) evaluation.
 - Do not make any edits to this worksheet as they may corrupt the calculations.
 - Analytical outlier samples are indicated in column N (1: outlier, 0: non-outlier).
 - These analytical outlier samples are first excluded before assessing the biological outliers.
 - Biological outlier subjects are indicated in column O (1: outlier, 0: non-outlier), with the corresponding unique subject ID in column L.
 - Samples that are either analytical or biological outliers are indicated in column P (1: outlier, 0: non-outlier).
 - All outlier samples are then excluded prior to the cut point calculations.
 - Use the "Filter" option in columns N, O, and P to list samples that are analytical or biological outliers.

Data analysis instructions (contd.)

- Various graphs of log(S/N) values are provided in "SCP-graphs" sheet.
- Distribution/histogram, boxplots of trends across plates/runs/analysts, and correlation plot of subject sera vs. negative control, averaged by plates/runs.
- Screening and titer cut point factor (SCPF and TCPF) calculations are summarized in the worksheet "SCP-summary".
 - Do not make any edits to this worksheet as they may corrupt the calculations.
 - Skewness and Excess-Kurtosis of log(S/N) distribution are calculated.
 - SCPF and TCPF values from different methods are provided in the table.
 - CPF results from the Parametric, Robust Parametric and Nonparametric methods are reported using the Point Estimate (PE) formula and the 90% Lower Confidence Limit (LCL) formulae (Shen et al, 2015).



Data analysis instructions (contd.)

- Following criteria can be applied for choosing between the Parametric vs. Robust Parametric vs. Nonparametric method for reporting the PE and LCL-based cut points. These criteria were derived based on the totality of insights on skewness and excess kurtosis from different publications.
 - If skewness is between -2 and 2 and excess-kurtosis is less than 4, use the <u>robust parametric</u> <u>method</u>. The relatively greater level of skewness and heaviness in tails (kurtosis) is handled well by the outlier-resistant measures (Median and MAD) used in the robust parametric method. The <u>parametric method</u> may be used when the skewness is between -0.5 and 0.5 and excess-kurtosis is less than 2, but due to the lack of the S-W normality test in this analysis tool, the robust parametric method is a safer choice for this scenario as well.
 - If skewness is not between -2 and 2, and if excess-kurtosis is < 2, use the <u>nonparametric method</u>. The nonparametric method is highly sensitive to borderline outliers (manifested when excess-kurtosis is greater than 2) for estimating the 95th or 99th percentile.
 - If skewness is not between -2 and 2, and if excess-kurtosis is > 2, use <u>alternative methods</u> (other data transformations and/or models).



Data analysis instructions (contd.)

- Fixed and Floating Confirmatory cut point (CCP) evaluations are provided in the next few worksheets.
- While the Fixed CCP is currently the most widely used approach, the Floating CCP approach may be useful when the % inhibition values are significantly different between the assay plates/runs (Devanarayan et al., 2017).
 - Do not make any edits to the worksheets as they may corrupt the calculations.
 - The outlier results, graphs, and cut point calculations are presented in the same manner as with the SCP analysis.
 - Recommendations provided for the choice of methods for SCP apply for the CCP as well.
 - Further details are provided below the summary table in the SCP and CCP summary worksheets.



Outlier identification process

- Uses the simplified alternative method proposed in Devanarayan et al (2017) that is amenable for programs such as Excel (bottom right of page 1490)
- First evaluate the differences in the S/N result of each observation of a subject sample from the median S/N value of the corresponding subject (subject-level residuals).
- Outliers can then be identified using Tukey's outlier box plot criteria on all the residuals across all subjects.
- The sample observations that meet this outlier criterion are the analytical outliers and should be excluded.
- After excluding all the analytical outliers, the median of the reportable results from each subject is recalculated and Tukey's outlier box plot criteria is applied to these medians to identify and exclude any apparent biological outliers.
- Options for 1.5xIQR and 3xIQR criteria are provided.

Demo of the tool

- https://www.ncbi.nlm.nih.gov/books/NBK92434/
 - <u>https://www.ncbi.nlm.nih.gov/books/NBK92434/#immunometh.lmmunogenicity</u>
 <u>Cut Point Anal</u>



Additional topics, if time permits



Low cut points, Low signal, FPR & clinical relevance

"Low" Cut Points



- FAQs/concerns:
- Is this due to relatively low biological variability?
- Is this due to *low assay signal* (e.g., RLU) values?
- Will this lead to *high in-study FPR*?
 - Will it require re-evaluation of in-study cut points?
- Excluding *too many outliers*? Try different outlier criteria?
 - This was already addressed in earlier part of this presentation!
- Will this *dilute the <u>overall</u> clinical relevance* of the ADA results?
- These questions will be addressed via retrospective evaluation of 25-30 assays; most of these assays have SCPF < 1.2

SCP factor vs. Biological variability



SCP factor vs. Assay signal



- Assays with high RLUs can also have low SCP factors.
- Low RLU (<100) does not always imply low SCP.

SCP factor vs. Total Variability (biological + analytical)



• As expected, assays with higher total variability have higher SCP factors.

Eisai

• Some departure from correlation is due to the differences between S and N

43

Low cut points and in-study FPR

Data from ~ 25 assays





Titer precision and Titer cut point

Precision of Titers (MSR)

Minimum Significant Ratio (MSR) Ref: USP chapter <1106>

- Useful for defining Titer Precision, and for setting a threshold for Titer CP
- Criteria for Treatment-boosted ADA



Titer of a sample (x-axis) is not significantly different from samples falling in the grey area.

If MSR of titers = 5, and if pre-dose titer = 10, post-dose titer should be > 50 to be treatmentboosted ADA.

Criteria:

MSR < 3 for most assays from our experience and is considered reasonable.

Evaluation of Titer MSR

Use the data from the sensitivity experiment (pre-study validation)

- 2-fold serial dilutions of HPC pools (or MPC), >= 3 runs, >= 2 analysts
- Compute titer by interpolating from each dilution curve
- >= 6 titer values (3 runs x 2 analysts)
- Evaluate the <u>SD of log(titer)</u> results, for use in the formula below.

 $MSR = 10^{t_{0.05,n_1} * \sqrt{2} * SD}$

- Derived from 95% one-sided upper confidence limit of the <u>difference</u> of two results.
- $t_{0.05,df}$ is the two-sided t-distribution threshold for 5% error rate
- n = # of titer results
- Anti-log (10[^]) of the difference of log(titers) = Ratio of Titers.
- Hence this is the Minimum Significant <u>Ratio</u> of two titer results (T-MSR).



Where to set the Titer cut point (TCP)?

TCP at 99% or 99.9% upper limit works well most of the time, but sometimes 95% works. *Need to apply an objective approach*!

• Avoid defaulting to 99.9% without justification.

If SCP falls above the lower plateau of PC dilution curve, TCP = SCP

• More likely to happen when SCP factor is high enough, (e.g., > 1.2).

When that's not the case, samples may not dilute down to SCP.

- Titers may not be measurable and will be highly noisy/variable.
- Progressively raise TCP threshold from 95% to 99%, 99.9% and 6xSD.
 - Evaluate the MSR of the Titers evaluated at each of these thresholds.
- TCP is set at the lowest upper limit (>=95%) where MSR < 3.

Alternative methods based on NC data may be used when appropriate.

Titer = MRD for confirmed positives that fall between SCP & TCP.



Treatment-boosted ADA

Criteria for identifying treatment-boosted ADA



- 1. Dilution-dependent criteria (adapted from clinical serology)
 - If the titers are determined via 2-fold serial dilutions, a 4-fold difference between pre-dose vs. post-dose titers is suggested as a criterion for treatment-boosted ADA. For 3-fold serial dilution, a 9-fold difference is suggested, etc. This is apparently common in other applications.
 - <u>Ignores assay & biological variability</u> & doesn't control error rates. This may lead to underreporting ADA incidence. For e.g., even if titers are diluted 2-fold, differences of 2 to 3-fold between pre-dose vs. post-dose titers may be statistically significant for many assays.

2. MSR of titers:

- MSR evaluation proposed for differentiating titer results can be used as a criterion for defining treatment-boosted ADA (<u>Ref:</u> USP chapter <1106.1>).
- Procedure for evaluation is explained earlier in this presentation.