

# Tutorial

## Practical guidance on immunogenicity cut-point evaluations and a demo of an Excel-based tool

V. Devanarayan, PhD, FAAPS

Eisai Inc. & University of Illinois Chicago

devan@uic.edu

Japan Bioanalysis Forum, Kyoto

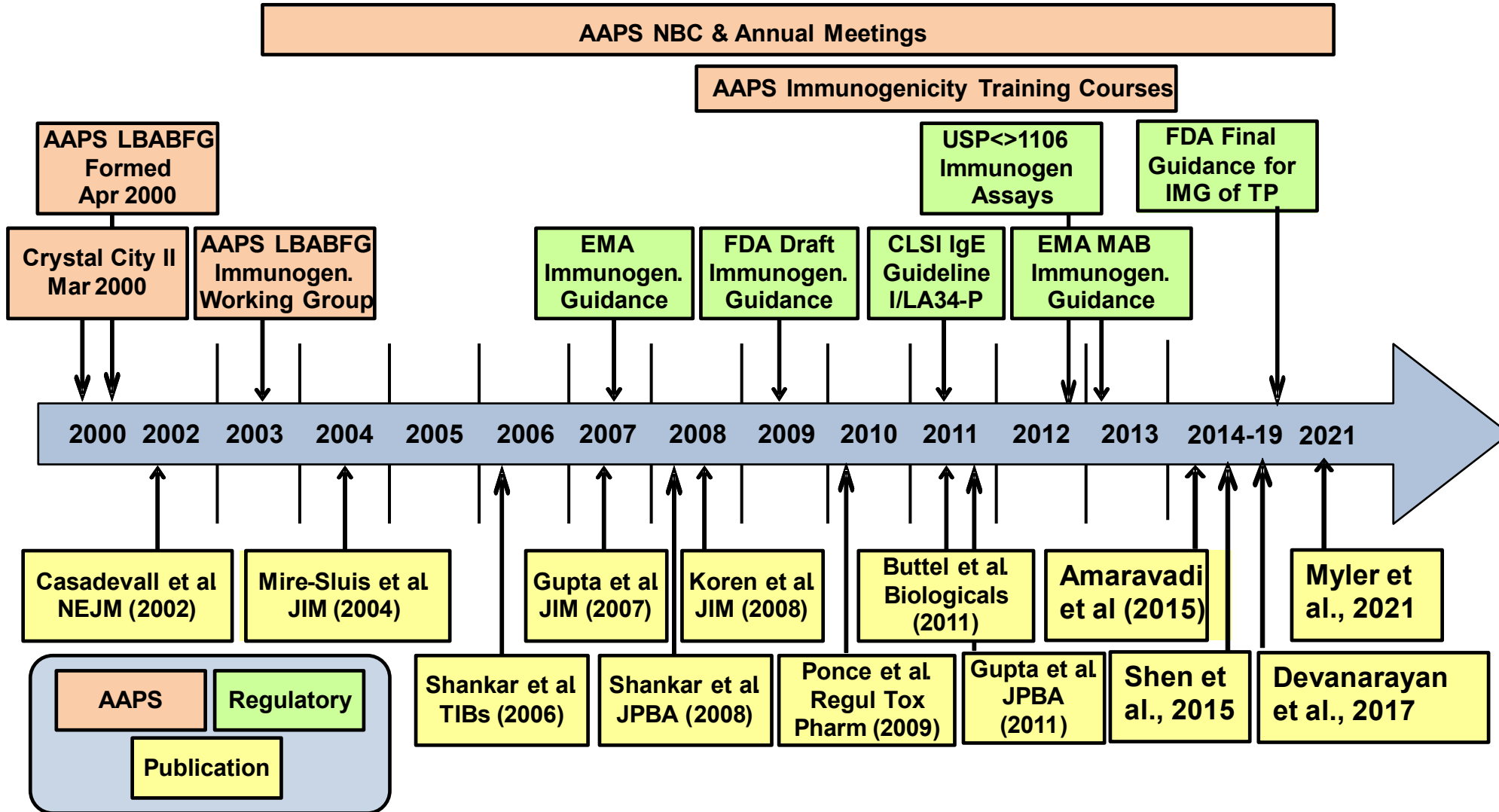
February 6, 2024

# Outline

- Cut point (CP) evaluation process
- CP formula (Point Estimate and Lower Confidence Limit)
- Negative Control (NC) evaluation
- Mean and variance comparisons between assay runs/plates
- Outlier evaluation and criteria
- Analytical & Biological variability
- Titer CP, Titer precision, and Treatment-Boosted ADA
- Insights on “Low” CP and “Low” Signal vs. False Positive Rate
- System suitability criteria
- In-study cut points
- Demo of Excel Tool

*Some topics and slides will be skipped due to limited time. They are provided for future reference*

# Evolution of Immunogenicity Testing



# Tiered Testing Strategy

- **Tier 1: Identify “reactive” samples**
  - Samples with signal above screening cut-point
- **Tier 2: Identify “Ab+” samples by testing reactive samples in the absence and presence of drug**
  - Samples with percent inhibition above confirmatory cut-point
- **Tier 3: Determine a sample titer value by serial dilution of Ab+ samples in Tier 2**
  - Titer is based on the screening cut-point or a higher “titer cut-point”. Can be continuous (requires interpolation) or discrete
- **Tier 4: Evaluate the neutralizing effects of antibodies**
  - Usually based on cell-based bioassay using Ab+ samples

*Cut point calculation process  
(pre-study validation)*

# Cut-point experimental design

## > 50 drug naïve ADA negative subjects

- Target disease population, if available.
- Include relevant demographic subgroups (gender, race, age, etc.).
- Multiple disease subtypes can be included, to investigate common or separate cut-points;  $n > 20$  per disease subtype.

## 6 runs, 2 analysts (3 runs per analyst)

- Each sample tested in duplicate
- Reportable result: Average of duplicate samples.

## Negative QC:

- $\geq 3$  reportable results/plate, each in duplicate, spread across the plate.
- Mean of reportable results used for normalizing subject sera (after excluding outliers).
- *Use median if outliers are not removed.*

# Cut-point experimental design

## Low, Mid and High QC:

- $\geq 2$  reportable results per plate, each in duplicate
- Mid may be replaced by a higher or alternative LPC.

## NC and LPC spiked with excess drug

- For assessing suitability of CCP and specificity for low ADA.

Three plates needed per run for testing these samples.

Include the drug-spiked subject samples in the same plates as their unspiked counterparts, for CCP evaluation.

Precision, CCP, TCP, etc., can be evaluated from this expt.

# Balanced design

Analyst	Assay Run	Assay Plate	Subject groups with serum samples tested over six runs		
			S <sub>1</sub> – S <sub>18</sub>	S <sub>19</sub> – S <sub>36</sub>	S <sub>37</sub> – S <sub>54</sub>
A <sub>1</sub>	R <sub>1</sub>	P <sub>1</sub>	X		
		P <sub>2</sub>		X	
		P <sub>3</sub>			X
	R <sub>2</sub>	P <sub>1</sub>		X	
		P <sub>2</sub>			X
		P <sub>3</sub>	X		
	R <sub>3</sub>	P <sub>1</sub>			X
		P <sub>2</sub>	X		
		P <sub>3</sub>		X	
A <sub>2</sub>	R <sub>4</sub>	P <sub>1</sub>	X		
		P <sub>2</sub>		X	
		P <sub>3</sub>			X
	R <sub>5</sub>	P <sub>1</sub>		X	
		P <sub>2</sub>			X
		P <sub>3</sub>	X		
	R <sub>6</sub>	P <sub>1</sub>			X
		P <sub>2</sub>	X		
		P <sub>3</sub>		X	

## Alternative visual

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

*All samples get tested in every run and every plate, by both analysts.*



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

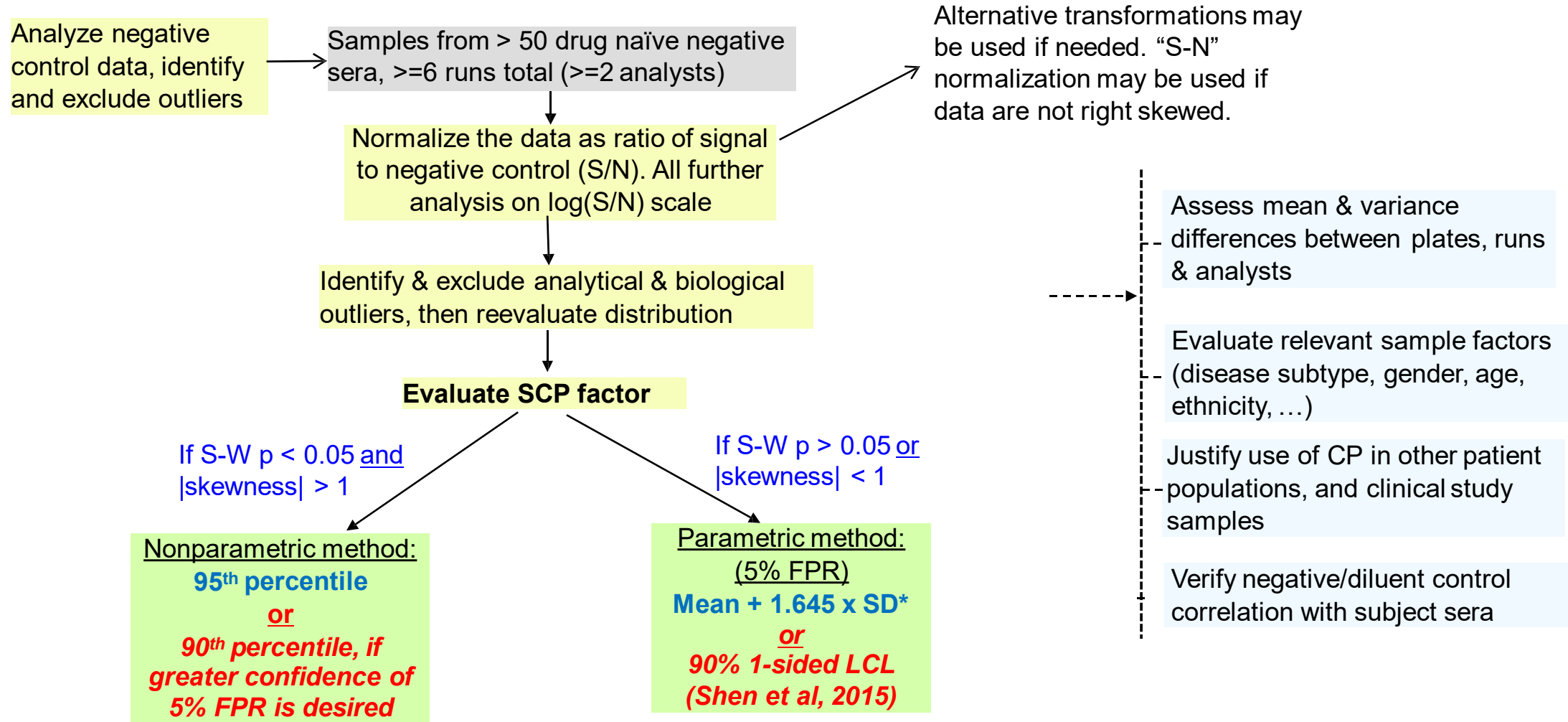
- *NC & LPC with and w/o drug, MPC & HPC*
- *17 subject sera per plate.*

# Rationale for this plate layout & controls

- Controls are *spread across the plate* to account for potential non-uniformity issues.
- *Drug-spiked NC*:
  - Potential need to float the CCP.
  - Confirms suitability of CCP during the in-study phase.
- *3 reportable results* for NC and drug-spiked NC (6 wells):
  - Due to the additional importance (normalization factor).
- *Drug-spiked LPC*:
  - Demonstrates ability to confirm low ADA level.
- *MPC*: demonstrate performance across full dynamic range.

# Screening Cut Point (SCP)

## Pre-study validation



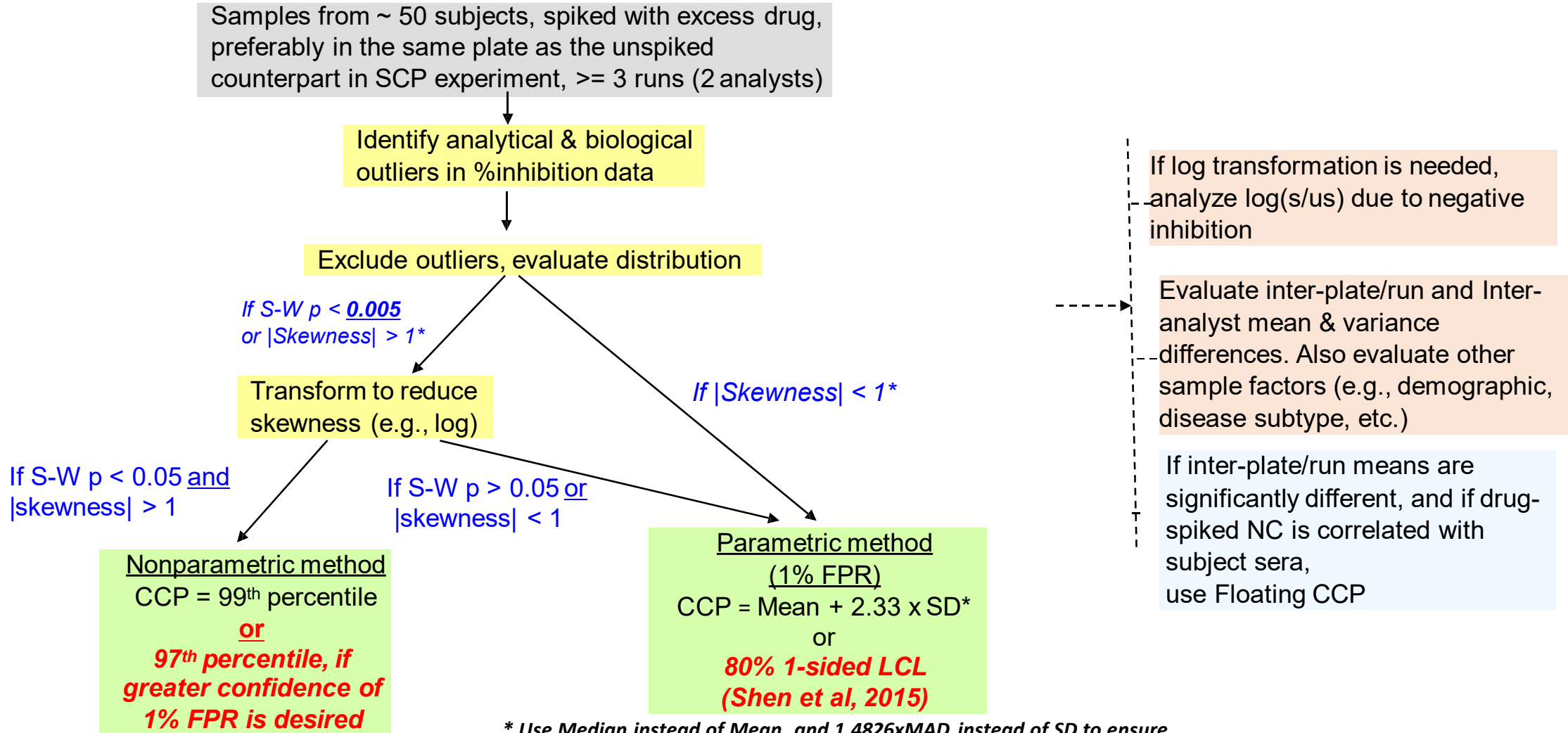
\* Use Median instead of Mean, and 1.4826xMAD instead of SD to ensure robustness when there are borderline outliers.

LCL: Lower confidence limit

Devanarayan et al., 2017

# Confirmatory Cut Point (CCP)

## Pre-study validation



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

LCL: Lower confidence limit

Devanarayan et al., 2017

*The traditional CP formula (point estimate)*

*versus*

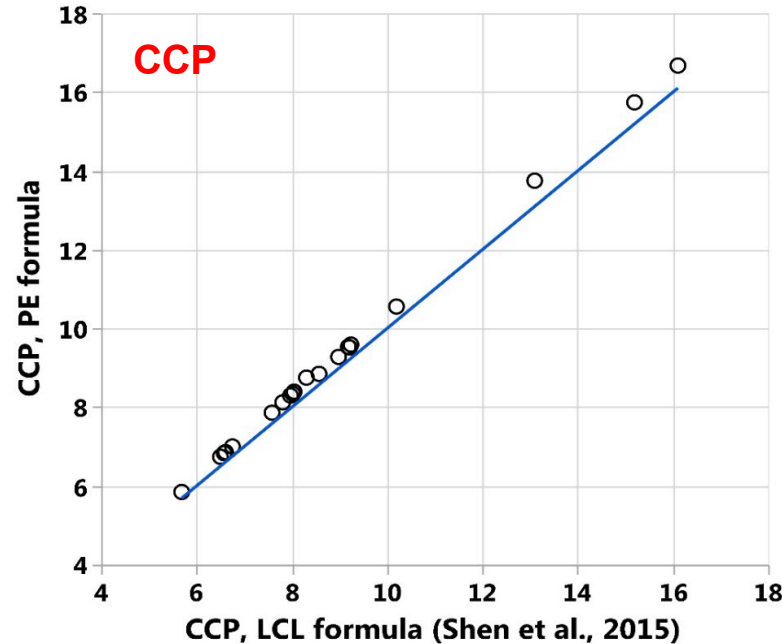
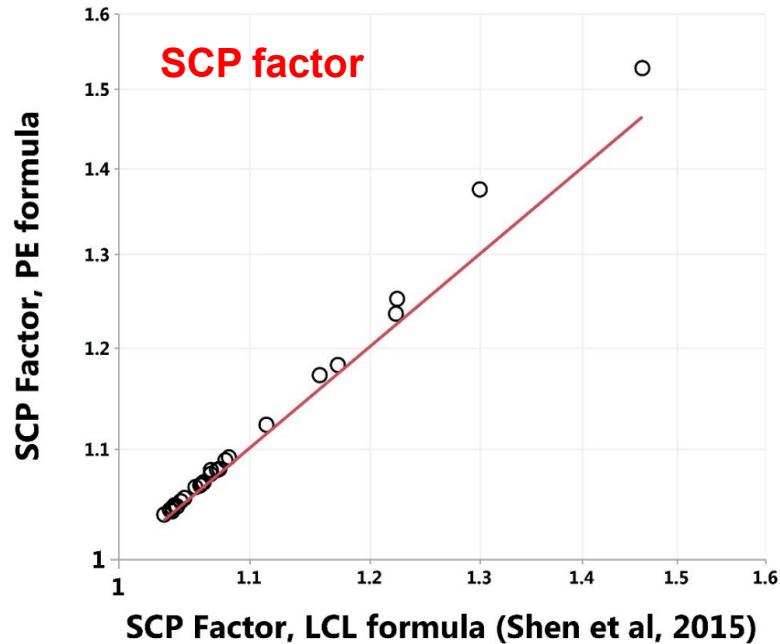
*Shen et al formula (Lower Confidence Limit)*

# Point estimate (PE) vs. Lower confidence limit (LCL)

- Screening CP was defined to yield approximately 5% FPR.
- “Point Estimate” (PE) of 95<sup>th</sup> percentile was proposed in several white papers:
  - **Mean + 1.645 x SD**, or its robust alternative (*Median instead of Mean, 1.4826xMAD instead of SD*).
  - Implies 5% FPR around *half the time* (50% confidence)
- Shen et al (2015) proposed a Lower Confidence Limit (LCL) for the 95<sup>th</sup> percentile (SCP) and 99<sup>th</sup> percentile (CCP).
  - Defined to yield 5% FPR with 90% confidence for SCP.
  - i.e., **90% one-sided LCL for SCP** and 80% LCL for CCP.
- We now explore practical differences with several datasets....

# SCP & CCP results from the standard vs. LCL formulae

Data from ~ 25 assays (mostly ECLs).

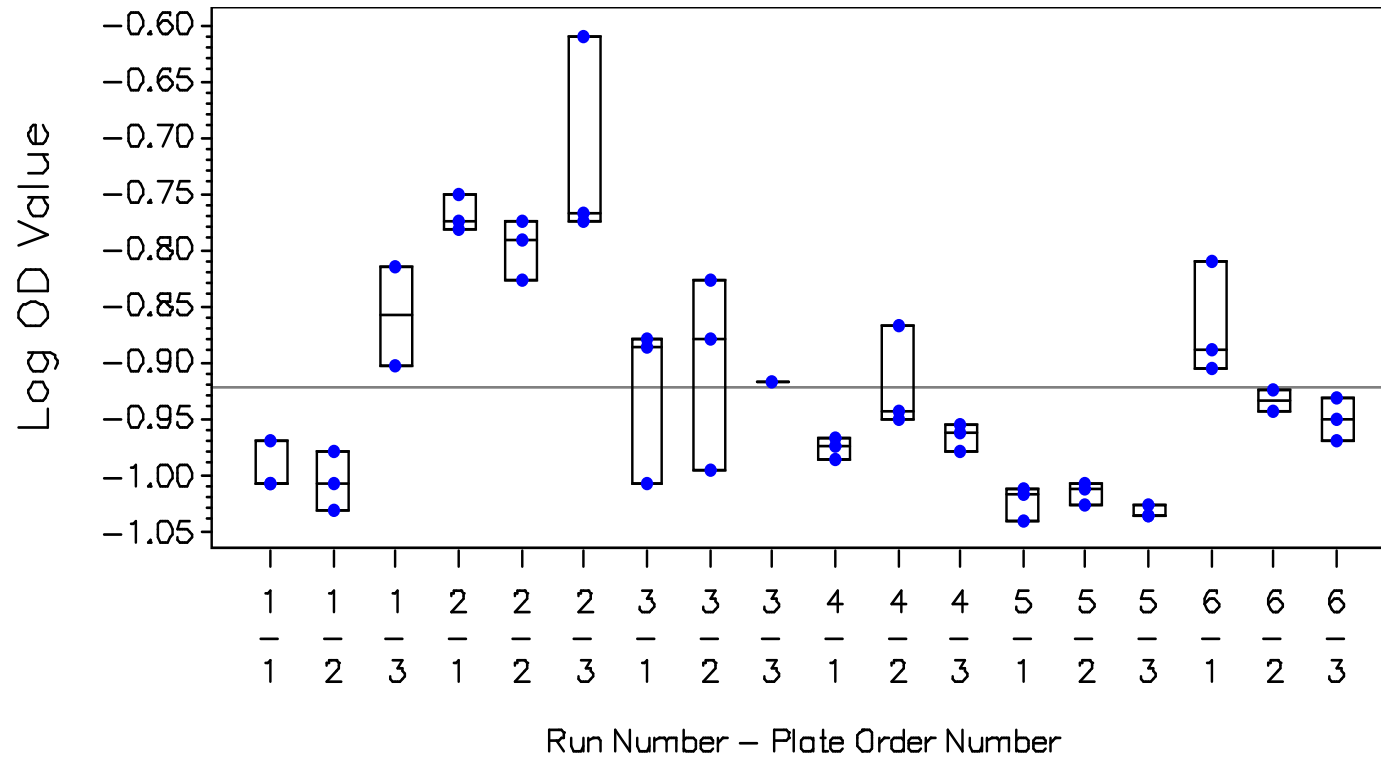


- CP values from the standard method (point estimate; PE) & lower confidence limit (LCL) formulae are practically similar for most assays.
- As expected, LCL values are slightly lower, and thus the FPRs are slightly higher.
- *Differences are more visible for smaller sample sizes (e.g., early-phase in-study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

# Negative Control evaluation



# Natural analytical variation (negative control pool)



- **This illustrates the natural analytical variation across assay plates and runs. This is just the practical reality of these data.**
- **Normalizing the patient samples with negative control helps.**

# Evaluation of Negative Control as a normalization factor

Floating SCP (or “SCP correction 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 the mean of subject sera from each run/plate.
- Evaluate Slope & Correlation (Need Slope close to 1, Correlation > 70%).

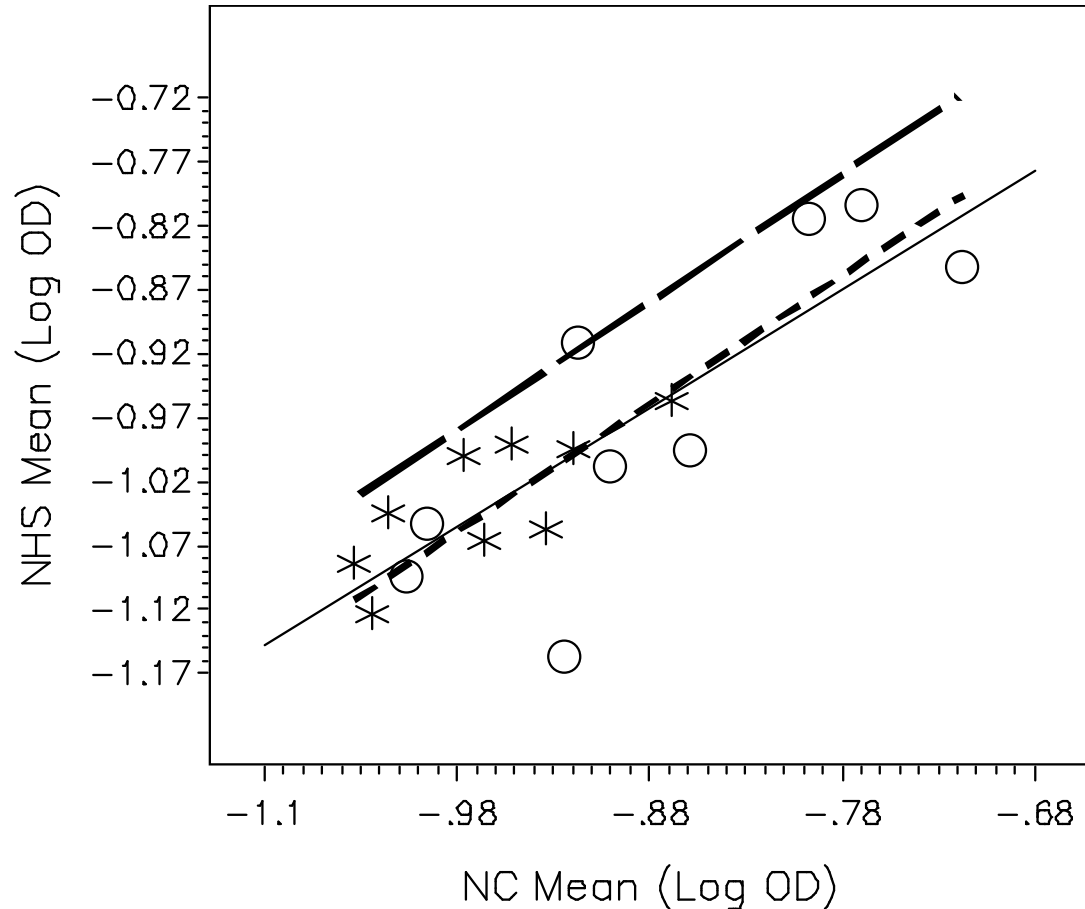
If no correlation, using NC to normalize may not be helpful.

- More likely when analytical variability exceeds biological variability.

## Alternatives:

- **New pool, Diluent/buffer control, other controls (same disease/demographic), ....**
- ***Assess the correlation of these alternative controls with the subject sera.***

# Normalizing patient sample results with negative control



Regression Equation:

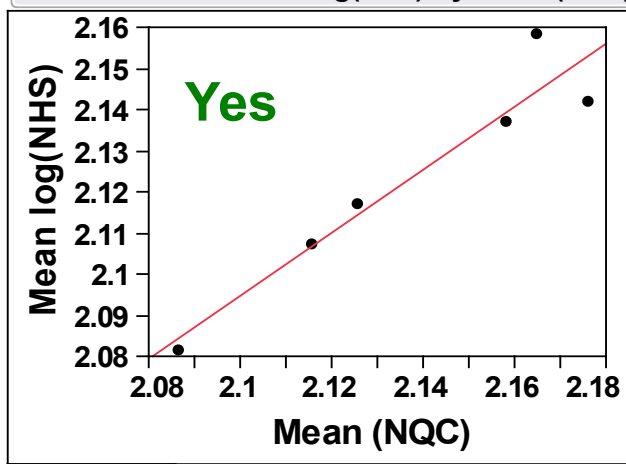
$$\text{NHS\_Mean} = -0.151008 + 0.922165 * \text{NC\_Mean}$$

Legend: Analyst 1=Circle, Analyst 2=Star

- This illustrates how the negative human sera (NHS) drifts in the same direction as the negative QC.
- Hence, normalization with negative QC and the use of a floating cut-point correction factor is necessary in most scenarios.

# Evaluation of NC as a normalization factor (contd.)

## Example-1



— Linear Fit

### Linear Fit

$$\text{Mean log(NHS)} = 0.4838325 + 0.7670203 * \text{Mean (NQC)}$$

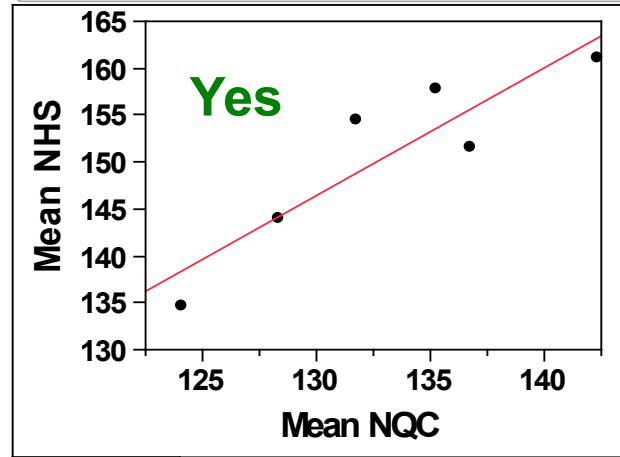
### Summary of Fit

RSquare	0.913008
RSquare Adj	0.89126
Root Mean Square Error	0.009078
Mean of Response	2.123699
Observations (or Sum Wgts)	6

### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	0.4838325	0.25312	1.91	0.1285
Mean (NQC)	0.7670203	0.11838	6.48	0.0029 *

## Example-2



— Linear Fit

### Linear Fit

$$\text{Mean NHS} = -31.16837 + 1.3659528 * \text{Mean NQC}$$

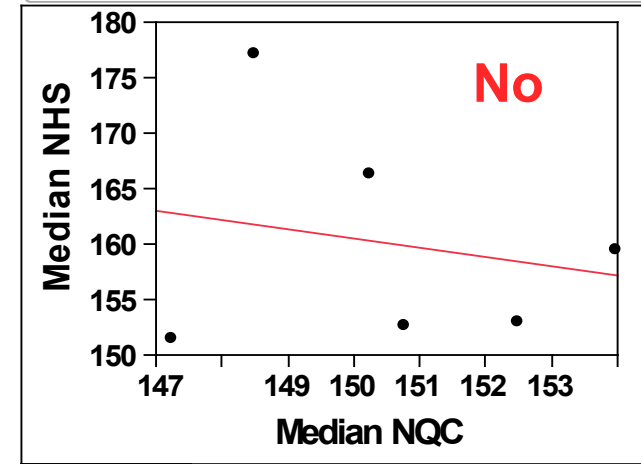
### Summary of Fit

RSquare	0.820393
RSquare Adj	0.775491
Root Mean Square Error	4.62033
Mean of Response	150.6172
Observations (or Sum Wgts)	6

### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	-31.16837	42.57033	-0.73	0.5047
Mean NQC	1.3659528	0.319563	4.27	0.0129 *

## Example-3



— Linear Fit

### Linear Fit

$$\text{Median NHS} = 287.03853 - 0.8435759 * \text{Median NQC}$$

### Summary of Fit

RSquare	0.042934
RSquare Adj	-0.19633
Root Mean Square Error	11.07579
Mean of Response	160.054
Observations (or Sum Wgts)	6

### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	287.03853	299.8051	0.96	0.3926
Median NQC	-0.843576	1.99142	-0.42	0.6936

# Mean and Variance differences between assay plates/runs

# Mean/Variance differences in SCP evaluations

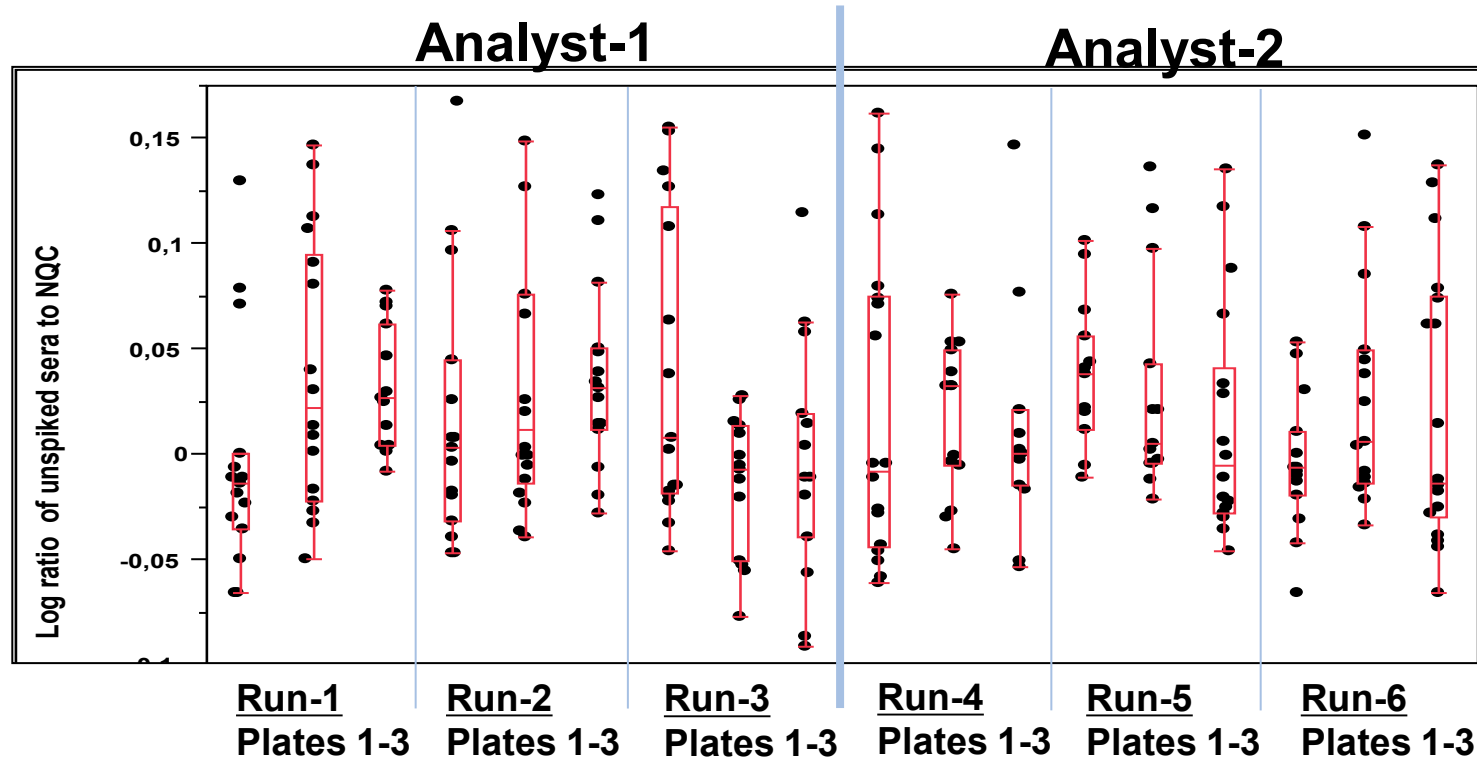
## If the assay run means and variances are similar in validation,

- Fixed cut-point is adequate. However, the Floating CP (SCP factor) is safer as it would account for potential drift during sample analysis.

## If the assay run variances are significantly different,

- Dynamic cut-point may be relevant, but requires significant resources to allow adequate sample size for CP evaluation in each run.
- **Re-optimize the assay to stabilize the variance (2019 FDA guidance)**
  - May not resolve the issue and may significantly delay the clinical program.
- **Alternative solutions:**
  - Pool the variances across all assay runs and estimate Floating CP (SCP factor) (Ref: USP chapter <1106>, Devanarayan et al., 2017). Although not optimal, it should be adequate if it yields around 5% FPR in pre-dose study samples.
  - Conservative approach: Drop the assay runs with the highest variability until the variances are no longer significantly different. Then pool the variance across the remaining assay runs to estimate the Floating CP (SCP factor).

# Mean/Variability differences in SCP evaluations (contd.)



- Inter-batch variability is significantly different (also after excluding additional outliers).
- *Floating* method for SCP can be used (i.e., SCP factor), either by pooling all the variability (Ref: USP chapter <1106>), or by dropping the runs with highest variances until no longer significantly different (Amaravadi et al., 2015).

# Mean/Variability differences in CCP evaluations

**If the assay run means and variances are similar in validation,**

- Fixed Cut Point is justified.

**If the assay run means/variances are significantly different,**

- Fixed CCP is still the standard choice.
- Evaluate the correlation of drug-spiked subject sera vs. drug-spiked NC across plates/runs.
- *If this correlation is significant, drug-spiked NC may be used as a normalization factor for defining Floating CCP (Devanarayan et al, 2017)*
  - *Calculate “Normalized % inhibition”:* divide the signal of the original and drug-spiked subject sample by the signal of the corresponding original and drug-spiked NC sample.
  - *This should help reduce the differences in means and variances between plates/runs.*



# *Outlier criteria and identification*

# Analytical vs. Biological Outliers

- ▶ **Biological Outlier** (*i.e.*, inter-subject) – An individual SUBJECT whose measured values consistently deviate from the overall **mean of all subjects**.
  - Generally greater impact on resultant CP values
  - Often Biological outliers can display appreciable %INH (*i.e.*, pre-existing ADA?)
- ▶ **Analytical Outlier** (*i.e.*, intra-subject) – An OBSERVED result for a test sample that deviates from the **mean response value for a specific subject and/or other ANOVA model factor(s)**.
  - May not be apparent based on visual inspection of observed responses
- ▶ See Devanarayan et al (2017) for details.

# Statistical modeling approach for outlier evaluation

## 1. Fit a mixed-effects model on the normalized response (S/N).

- Random effects: Subjects, Run # nested within Analyst, and Plate ID.
- Fixed effects: Analyst, Plate testing order, interaction of Analyst and Plate testing order + gender, disease types, etc., 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 on these *conditional residuals* to identify the outliers. → Analytical outliers

*Outlier box-plot criteria:  $Samples > Q3 + 1.5*(Q3-Q1)$  or  $< Q1 - 1.5*(Q3-Q1)$   
 $Q3 = 75^{th}$  percentile,  $Q1 = 25^{th}$  percentile*

## 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. Apply the “outlier box-plot” criteria on these subject BLUPs to identify the outlier subjects → **Biological outliers**
7. Refit the model without all the analytical & biological outliers, for analyzing the assay characteristics (analyst effect, plate/run differences, variability differences, etc.)

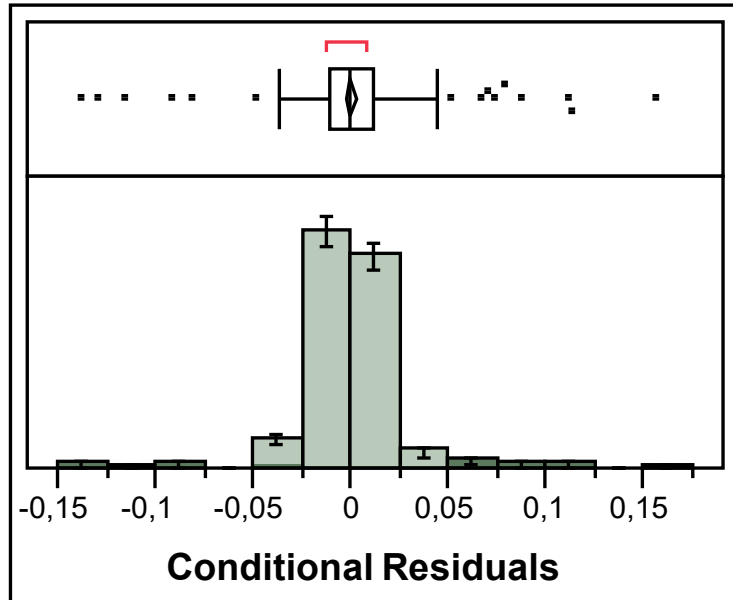
Some borderline outliers may remain.

- Normality test may fail due to long tails, but as long as distribution is reasonably symmetric ( $|\text{skewness}| < 1$ ), parametric method can be used.

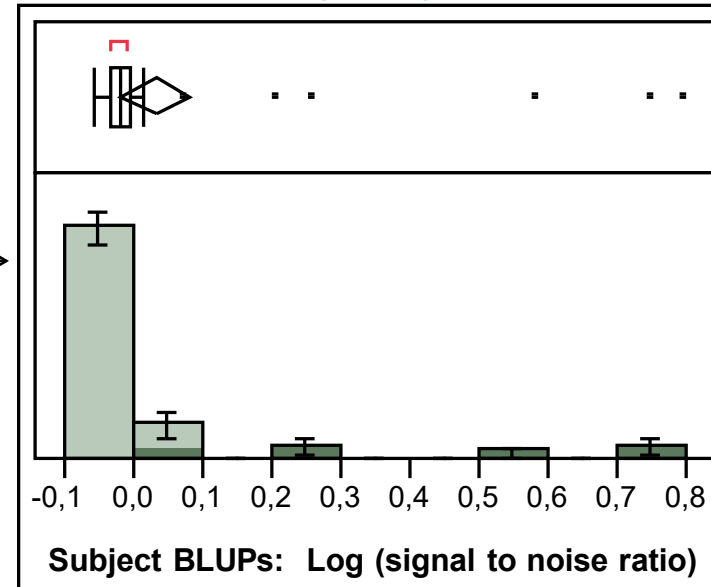
***Use of Median & MAD (“robust parametric”) instead of Mean/SD in the SCP & CCP calculation*** will alleviate this issue.

# Illustration of outlier evaluation with statistical modeling approach

**Analytical Outliers**  
Conditional Residuals from mixed-effects model



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



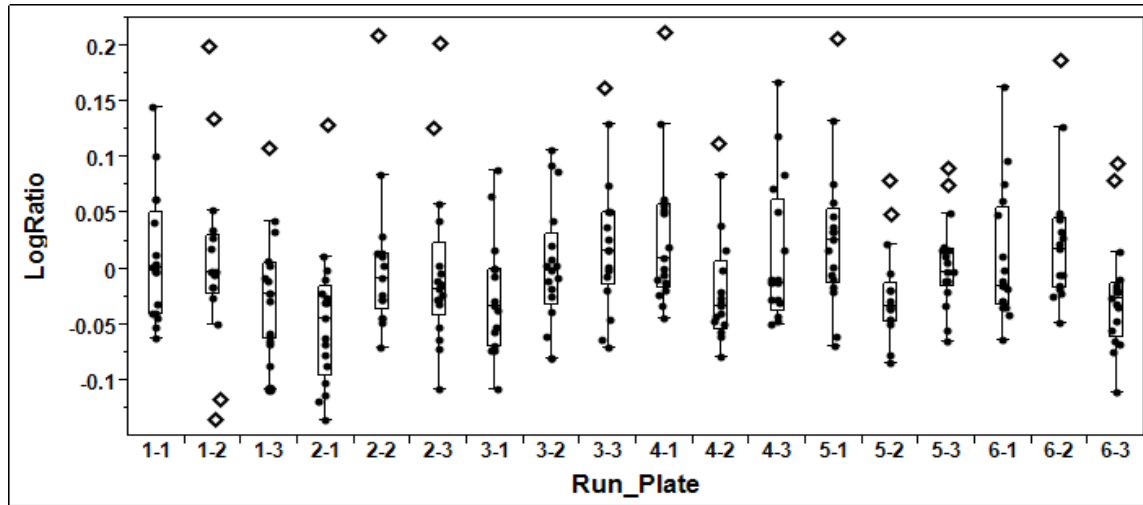
Mixed effects model is fit on  $\log(S/N)$ . Conditional Residuals are evaluated to identify analytical outliers.

After excluding analytical outliers, model is refit to the remaining data to identify the biological outliers.

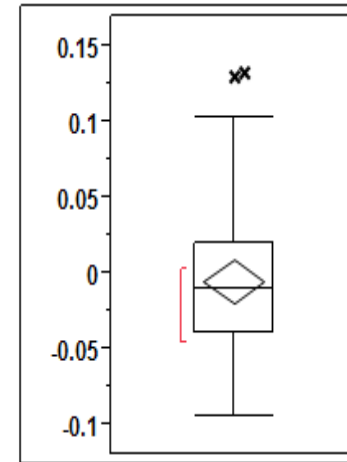
*This method and a simpler alternative are described in Devanarayan et al (2017).*

# Simpler alternative for outlier evaluation that may be used if the statistical modeling approach is not feasible

Plate-level Outliers



Subject-level Outliers



Subject averaged data

- Identify & exclude *plate-level outliers* (PO) from each assay run/plate separately. Iterate until there are no more PO.
- Then identify & exclude *subject-level outliers* (SO) by evaluating the distribution of subject averaged data.
- Then verify the 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).*

*NB: Iterations are not critical if robust methods are used.*

# What criteria to use for Outliers?

## Hypothetical Lab Colleague



- ▶ My screening CP factor is **too low**. You removed too many outliers!
- ▶ I am concerned that I will have **too many positive samples** in Tier 1.
- ▶ Can you **re-examine** how you removed the outliers?
- ▶ Can you **relax** the outlier criteria?
- ▶ How will **leaving more samples** affect the SCP?

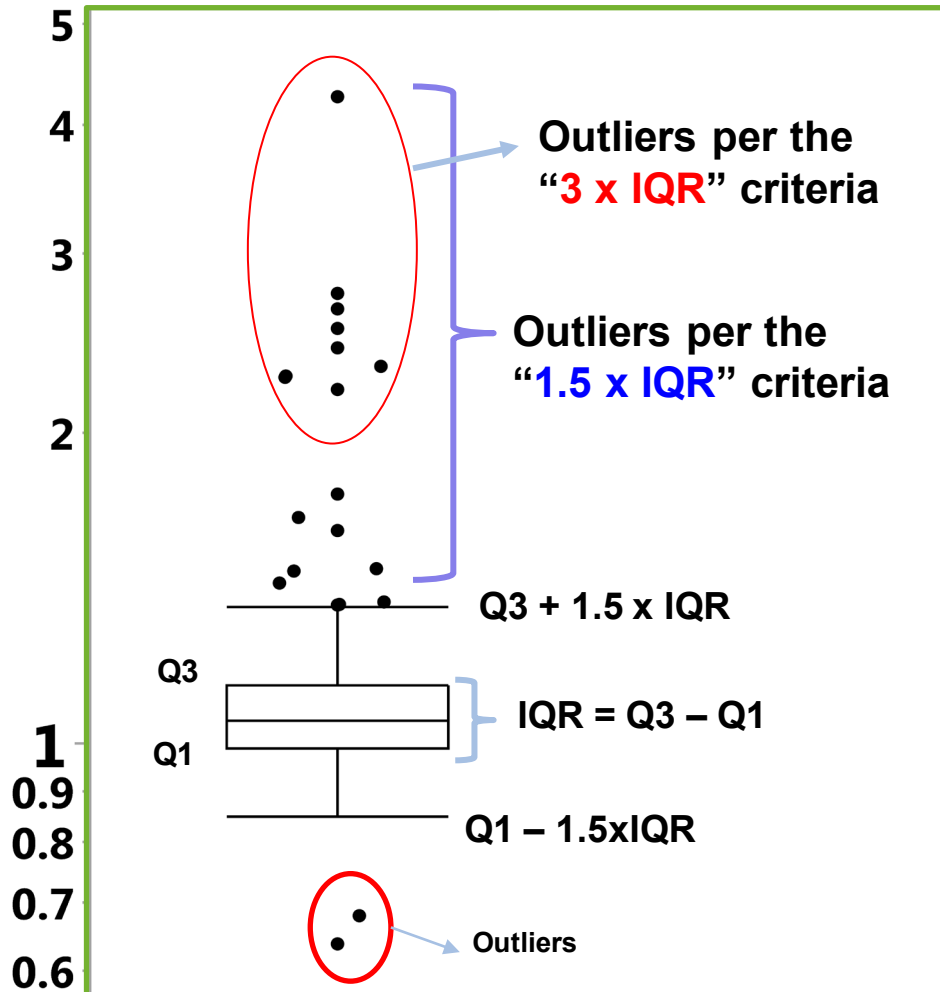
# Outlier criteria

- Tukey's outlier box plots by default are based on the following criteria:
- *High outliers:*  $> Q3 + 1.5 \times IQR$
- *Low outliers:*  $< Q1 - 1.5 \times IQR$ 
  - $Q3 = 75^{th}$  percentile,  $Q1 = 25^{th}$  percentile
  - $IQR = \text{Inter quartile range} = Q3 - Q1$  criteria:
- Due to concerns about “low cut points”, “too many outliers”, etc., these criteria gets subjectively changed to  $3 \times IQR$ .
- Several talks at conferences and some publications bemoan about “excessive” outlier removal.
- Such subjectivity is not necessary if the robust approach is used, i.e., Median/MAD instead of Mean/SD.
- **Results from  $1.5 \times IQR$  vs  $3 \times IQR$  are usually similar if Median/MAD is used.**



# Interpretation of 1.5xIQR and 3xIQR outlier criteria

Distribution of S/N results



For normal distribution,

**1.5 x IQR** is approx. Mean  $\pm 2.67 \times SD$

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

**3 x IQR** is approx. Mean  $\pm 4.67 \times SD$

- $\sim 99.9997\%$  of the samples

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

# Median

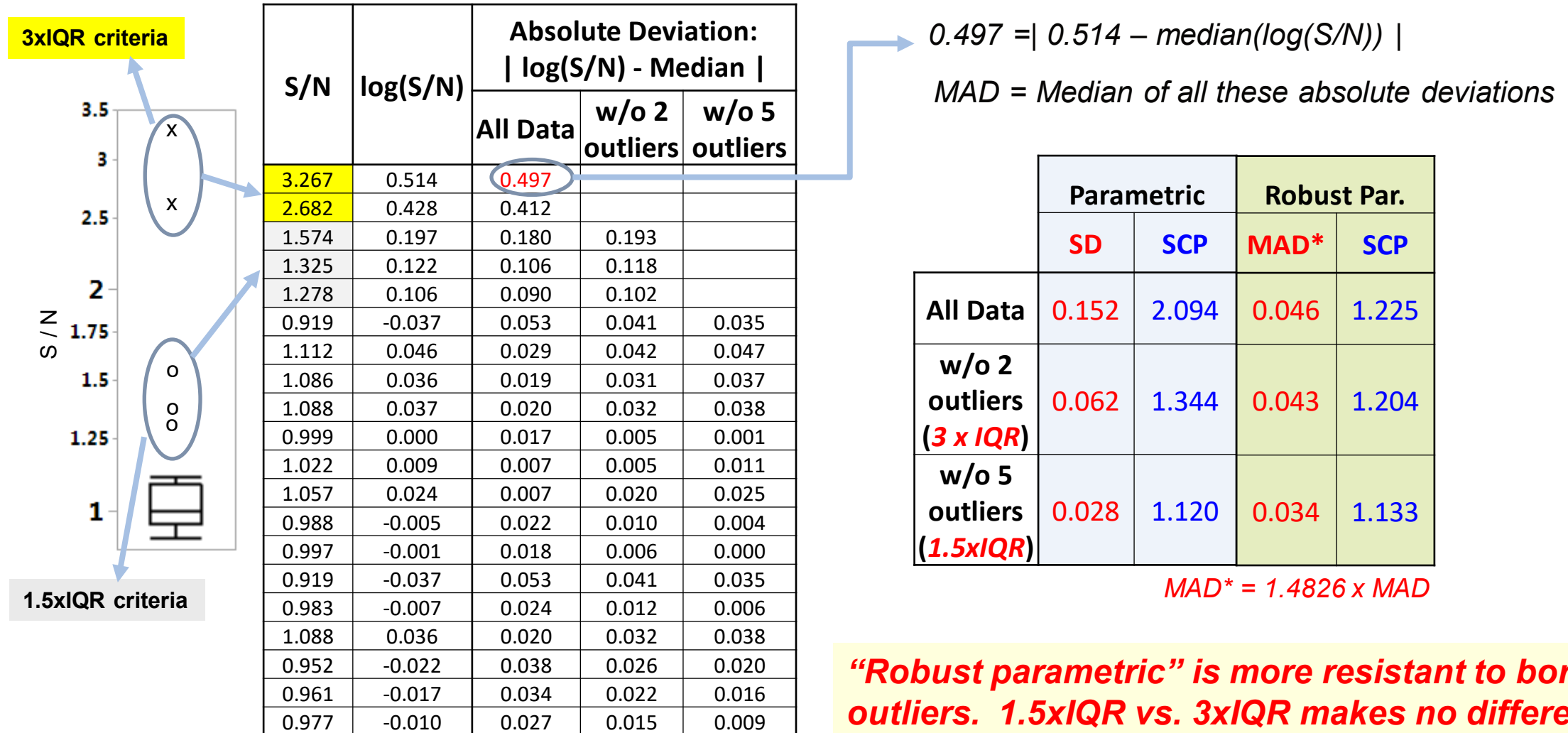
- Median = “middle value” of a distribution.
- Less skewed by high and low outlier values
- More robust than Mean, when outliers are present.
  - That is, less affected by borderline outliers.
- Useful for skewed distributions
- These characteristics make the Median more appealing for use in computations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

# Median Absolute Deviation (MAD)

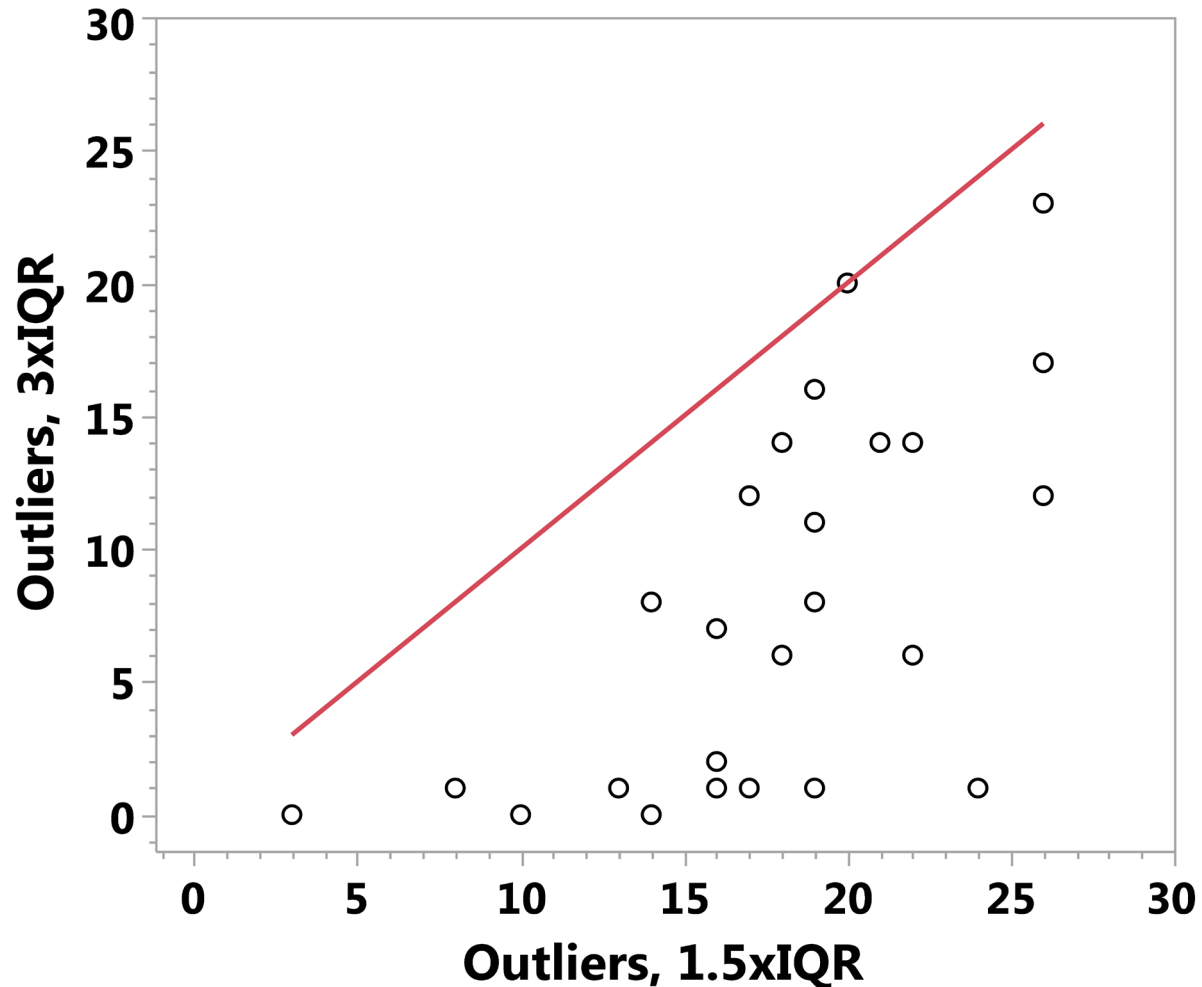
- A robust measure of the variability of continuous measurements.
- Equals the median of the absolute deviations from the median:
  - $MAD = \text{median}(|X_i - \text{median}(x)|)$
- *SD is replaced in the formula by  $1.4826 \times MAD$*
- More resistant to the outliers than SD
- Borderline outliers do not significantly impact the MAD (i.e., no need to debate about  $1.5 \times IQR$  vs  $3 \times IQR$ )
- 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.

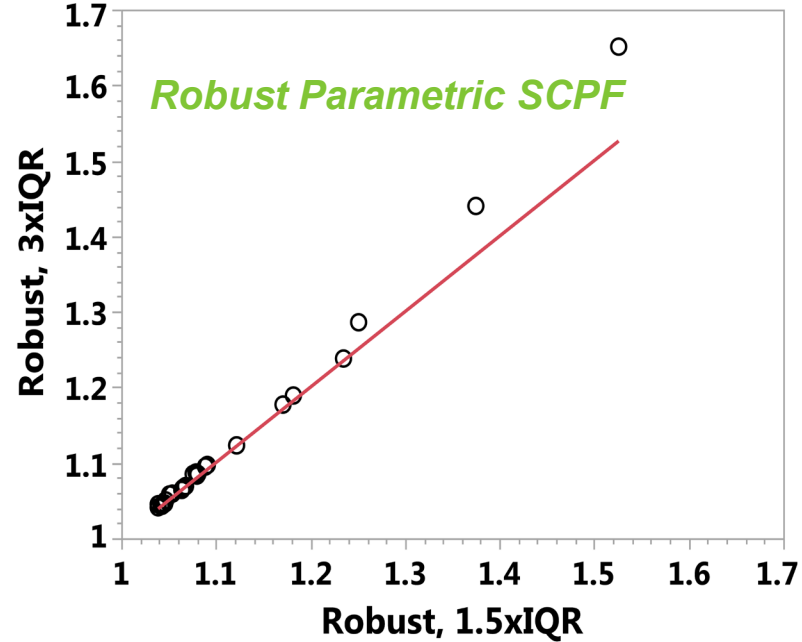
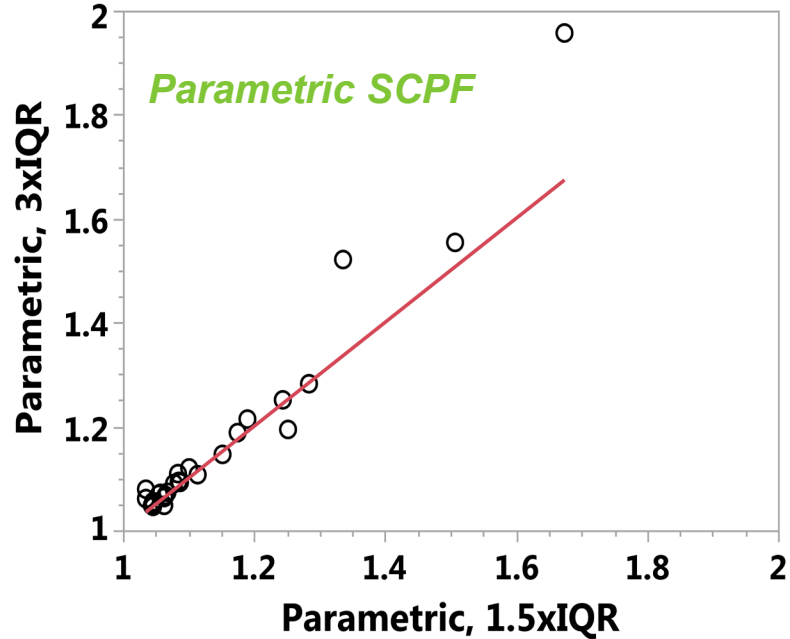


# 1.5xIQR vs. 3xIQR criteria: data from 25 assays



- 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.

# 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  $\sim 2.7xSD$ , it is widely used in the statistics literature, results in a similar CP when using the robust method, and is, therefore, a good default.

# Summary points about Outlier criteria/methods

- Mean/SD are highly sensitive to borderline “outliers”.
  - **SCP & CCP change significantly for different outlier criteria/methods.**
- Median/MAD are robust to borderline “outliers” and criteria.
  - **SCP & CCP are practically similar, regardless of the criteria used.**
- 1.5xIQR criteria is ~ popular 3xSD rule.
  - Widely used diagnostic in scientific literature, software, etc. (Tukey).
- 3xIQR criteria may exclude real outliers (e.g., pre-existing Abs).
- Iterative process is overkill. Additional outliers often not meaningful.
- CP decision flow scheme in our 2017 paper recommends:
  - **1.5xIQR criteria, without iteration.**
  - **Use of Median/MAD for skewed/non-normal distribution, and is a good default.**
- *Robust or Go-bust 😊*

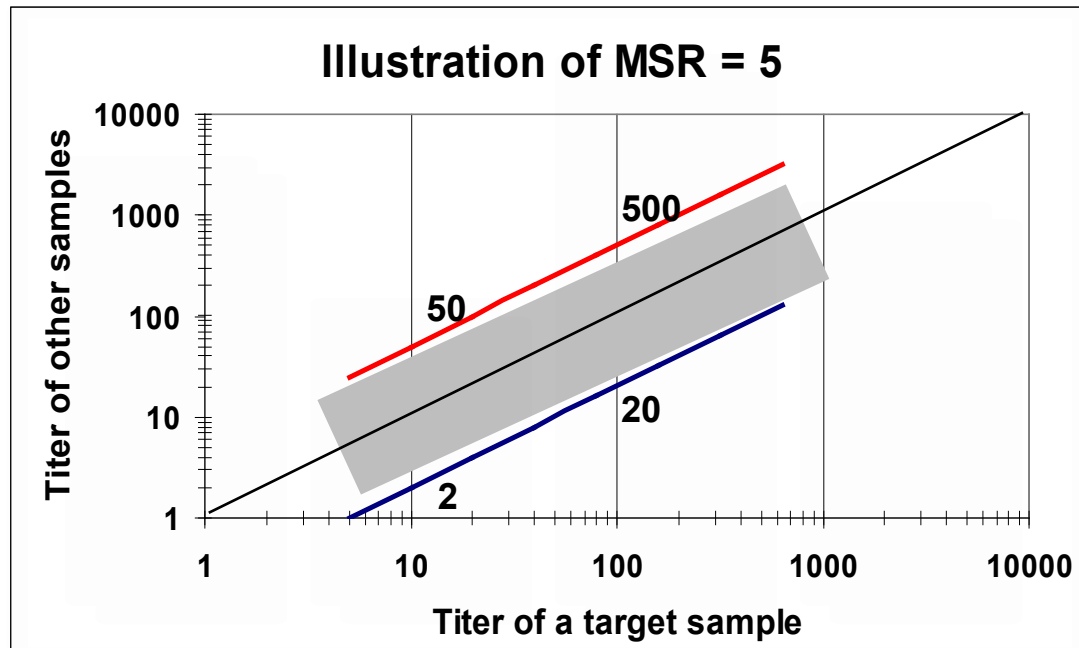
# Titer CP, Titer Precision & Treatment-boosted ADA



# Precision of Titers (MSR)

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

- Useful for defining *Titer Precision*, and for setting 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 treatment-boosted ADA.

## Criteria:

**MSR < 3 is generally considered desirable.**

# Evaluation of Titer MSR

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

- 2-fold serial dilutions of HPC pools (or MPC),  $\geq 3$  runs,  $\geq 2$  analysts
- Compute titer by interpolating from each dilution curve
- $\geq 6$  titer values (3 runs x 2 analysts)
- Evaluate SD of  $\log(\text{titer})$  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 difference of two results.
- $t_{0.05, df}$  is the two-sided t-distribution threshold for 5% error rate
- $n = \#$  of titer results
- Anti-log ( $10^{\wedge}$ ) of the difference of  $\log(\text{titers}) = \text{Ratio of Titters}$ .
- Hence this is the Minimum Significant Ratio of two titer results (T-MSR).

# Where to set the Titer cut point (TCP)?

- **If SCP > lower plateau of PC dilution curve, TCP = SCP**
  - More likely to happen when SCP factor is high enough, (e.g., > 1.2).
- **If not, samples may not dilute down to the SCP.**
  - Titers may not be measurable and will be highly noisy/variable.
  - Progressively raise TCP threshold to 99%, 99.9% and 6xSD, until Titers are measurable & precise (use MSR)
  - ***TCP at 99% or 99.9% upper limit works well most of the time.***
  - Alternative methods based on NC data may be used when appropriate.
- ***Titer = MRD for confirmed positives that fall between SCP & TCP.***

# 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 can be the criteria for treatment-boosted ADA. For 3-fold serial dilution, a 9-fold difference can be the criteria, etc.
- *This ignores assay and biological variability. This may lead to under/over-reporting ADA incidence. Even if titers are diluted 2-fold, differences of 2 to 3-fold between pre- vs. post-dose titers may be statistically significant for many assays.*

## 2. MSR of titers or S/N

- MSR evaluation proposed for differentiating titer results can be used as a criterion for defining treatment-boosted ADA (Ref: USP chapter <1106.1>).
- This approach is relevant only when titers are interpolated. For endpoint (discrete) titers, the MSR of S/N values may be used as the criteria.

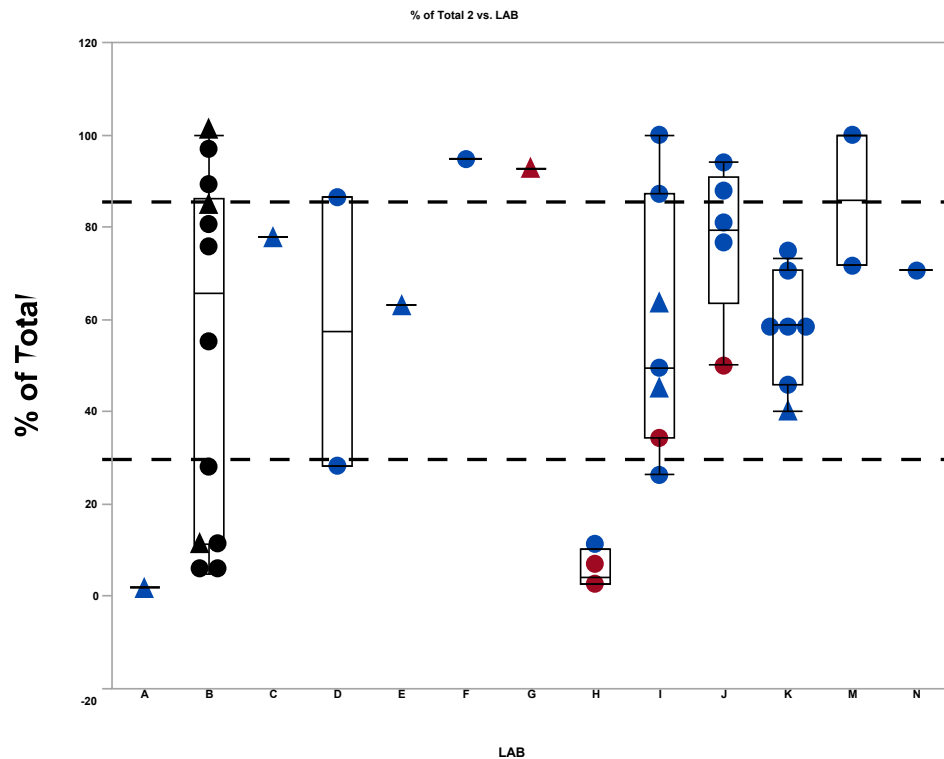
# Analytical & Biological Variability

# Analytical & Biological variability

- SCP is designed to yield an overall FPR of 5% in a study.
- As the study samples are tested across several assay plates and runs, and by multiple analysts, all sources of variation should be included in CP calculation.
- Total Variability = Analytical Variability + Biological Variability
- Analytical includes various sources of assay variability.
  - inter-run, intra-run, inter-analyst, etc.
- Biological variability is the inter-subject component.
- Need at least 6 assay runs and > 50 subject sera for CP evaluations.
  - 4 runs are not adequate if the analytical variability is usually a bigger component of the total variation than the biological variability.
  - Relative contributions of analytical vs biological variation varies greatly across assays/studies.

# Screening Assay: Analytical Variability (% of Total)

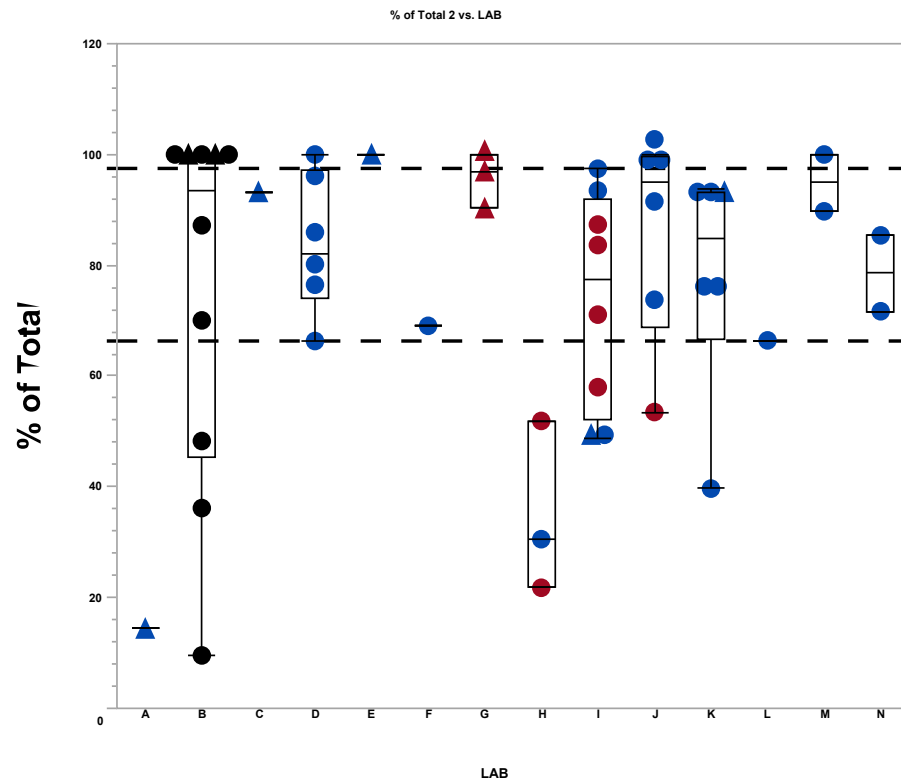
- % of total variability that is Analytical from 43 assays is plotted below.
  - **Median = 64%** (IQR = 34.8 to 86.5 %)
- % of total variability that is Biological:
  - **Median = 36%** (IQR = 13.5 to 65.7%)



*~ 64% of the total observed variability is Analytical (after outlier removal)*

# Confirmatory Assay: Analytical Variability (% of Total)

- % of total variability that is Analytical from 51 assays is plotted below.
  - **Median = 86%** (IQR = 66.3 to 97.4%)
- % of total variability that is Biological:
  - **Median = 14%** (IQR = 2.6 to 33.7%)



*~ 86% of the total observed variability is Analytical (after outlier removal)*

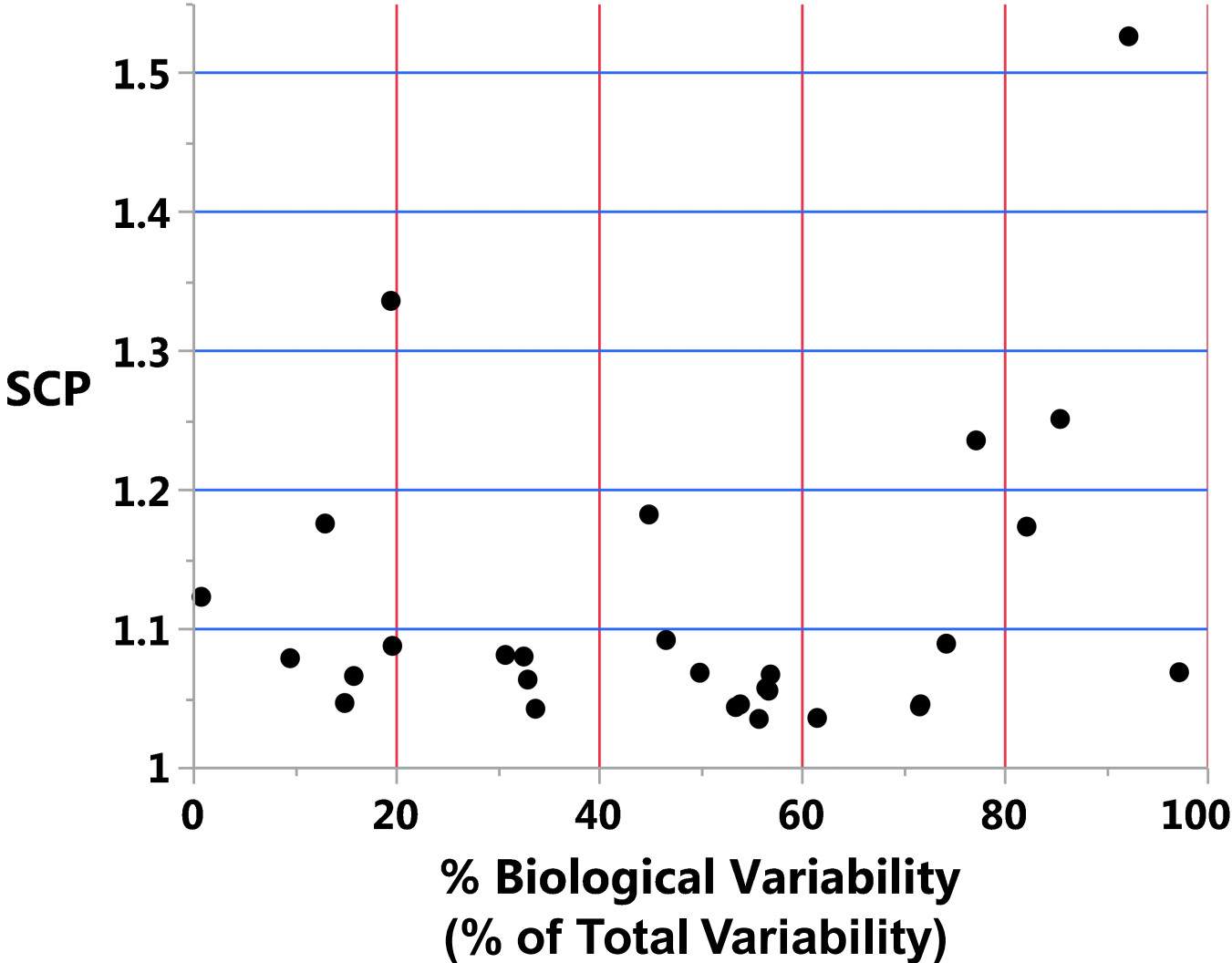


*Low cut points, Low signal, FPR & clinical relevance*

# “Low” Cut Points

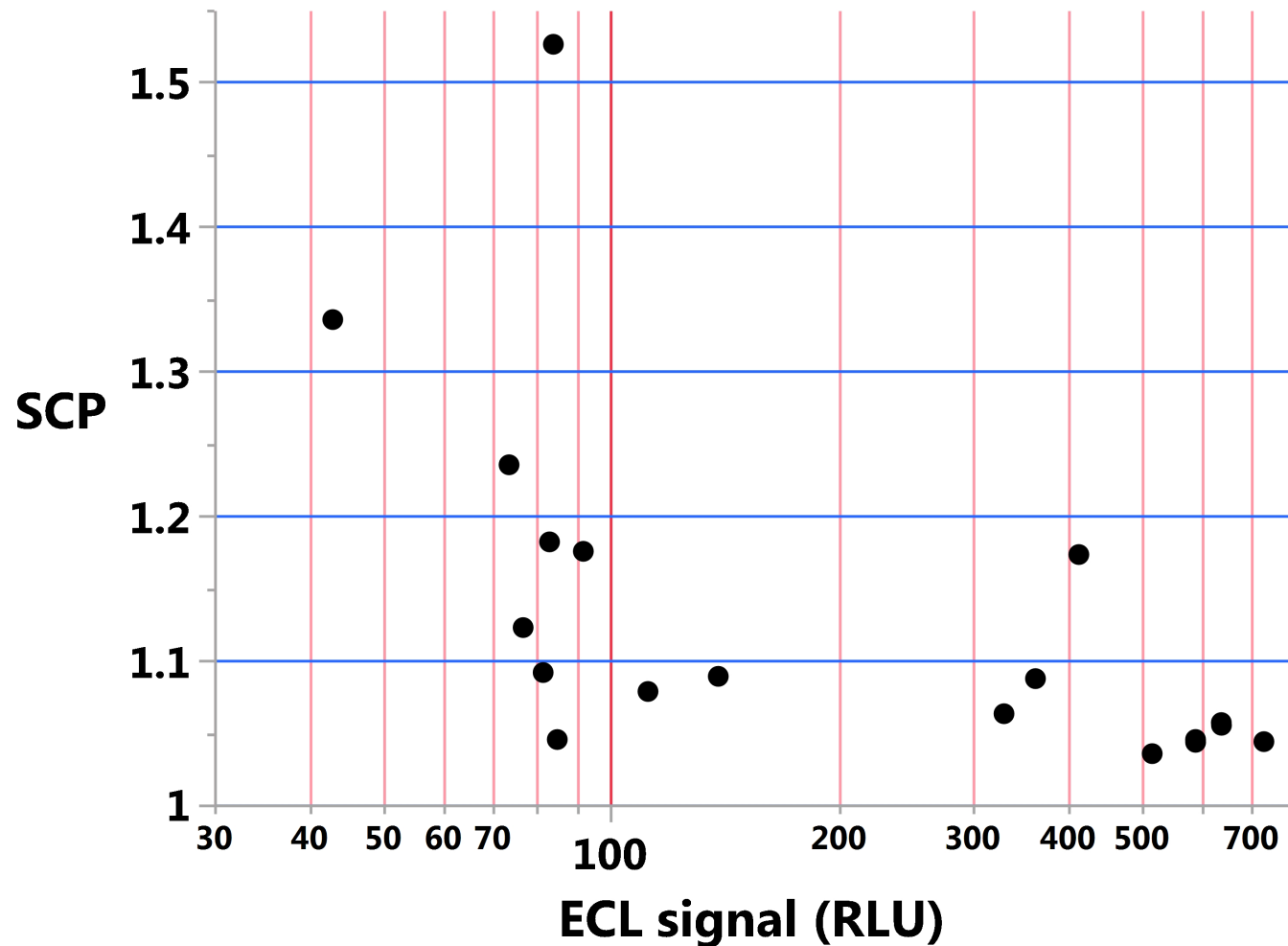
- SCPF can be quite low, sometimes  $< 1.1$  and often  $< 1.2$
- FAQs/concerns:
  - Is this due to *relatively low biological variability*?
  - Is this due to *low assay signal* (e.g., RLU) values?
  - Will this lead to *high in-study FPR*?
    - Will it require *re-evaluation of in-study cut points*?
  - Excluding *too many outliers*? Try different outlier criteria?
    - *This was already addressed in earlier part of this presentation!*
  - Will this *dilute the overall 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



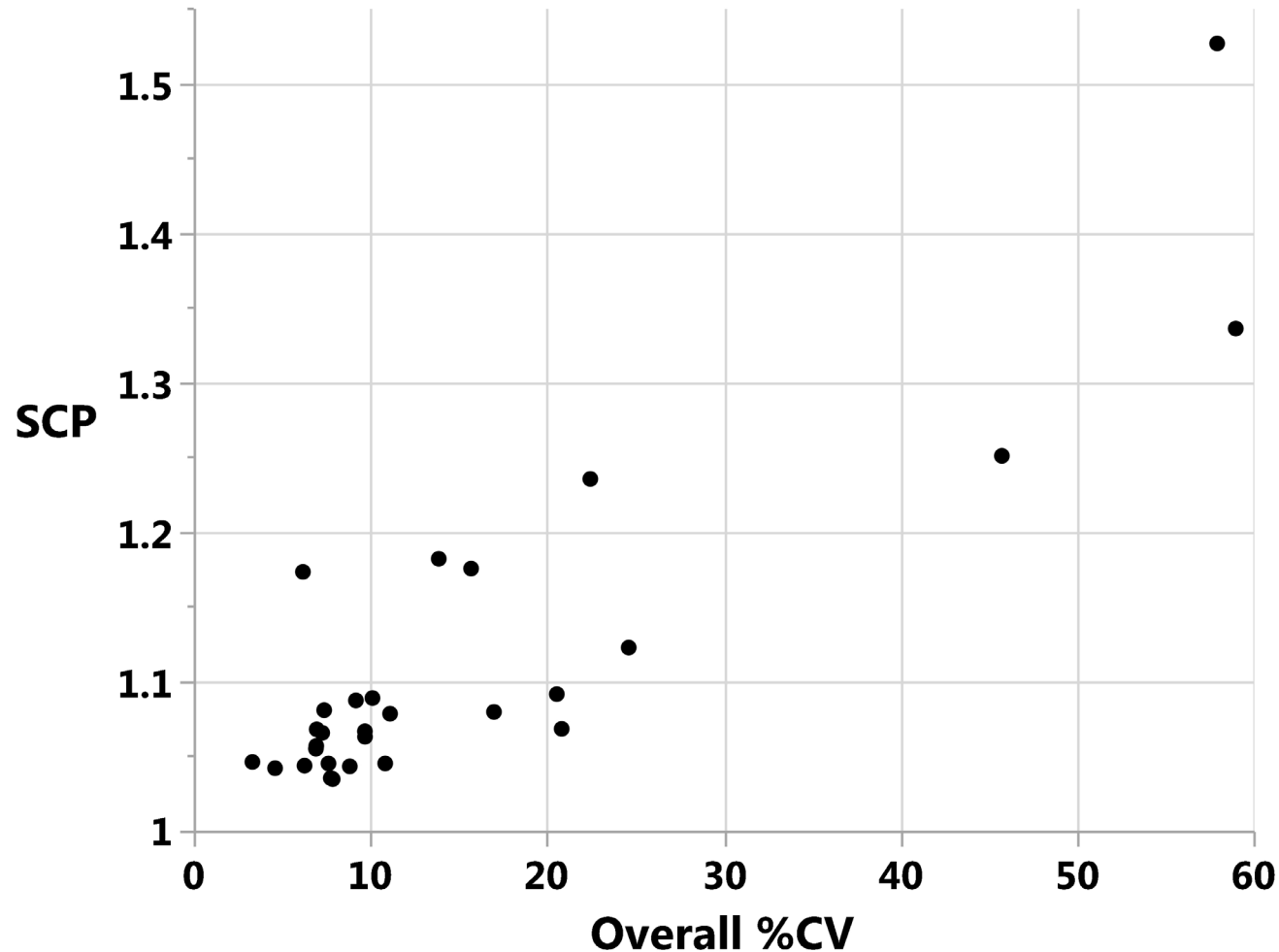
*Assays with high biological variability can also have low SCP factors.*

# SCP factor vs. Assay signal



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

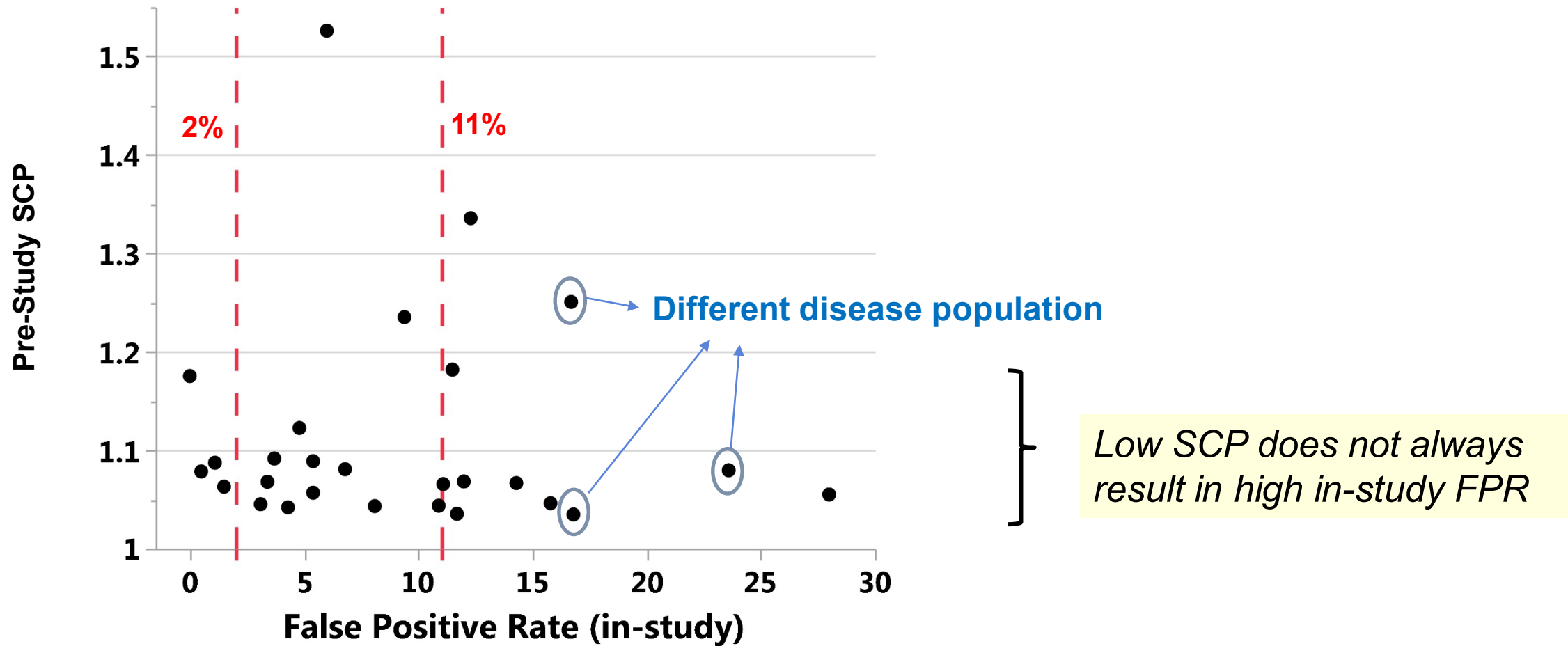
# SCP factor vs. Total Variability (biological + analytical)



- *As expected, assays with higher total variability have higher SCP factors.*
- *Some departure from correlation is due to the differences between S and N*

# Low cut points and in-study FPR

Data from ~ 25 assays



# Impact of low cut points on clinical relevance ??

Lower cut points → higher incidence of low-titer ADAs.

Most of these ADAs may not be clinical relevant, but

*This may not dilute the clinical relevance of higher titer ADA positives.*

*On the contrary, this may strengthen the evidence around clinical relevance of the high titer ADA results.*

*Examples in the next few slides.*

# Example 1

*Impact of increased incidence of low-titer ADAs...*

	Titer < 100	Titer > 100
CI no	8	10
CI yes	2	20
Total	10	30

*p-value ~ 0.01*

*Suppose increasing the assay sensitivity led to more low-titer ADAs:*

	Titer < 100	Titer > 100
CI no	80	10
CI yes	20	20
Total	100	30

*p-value ~ 0.000001*

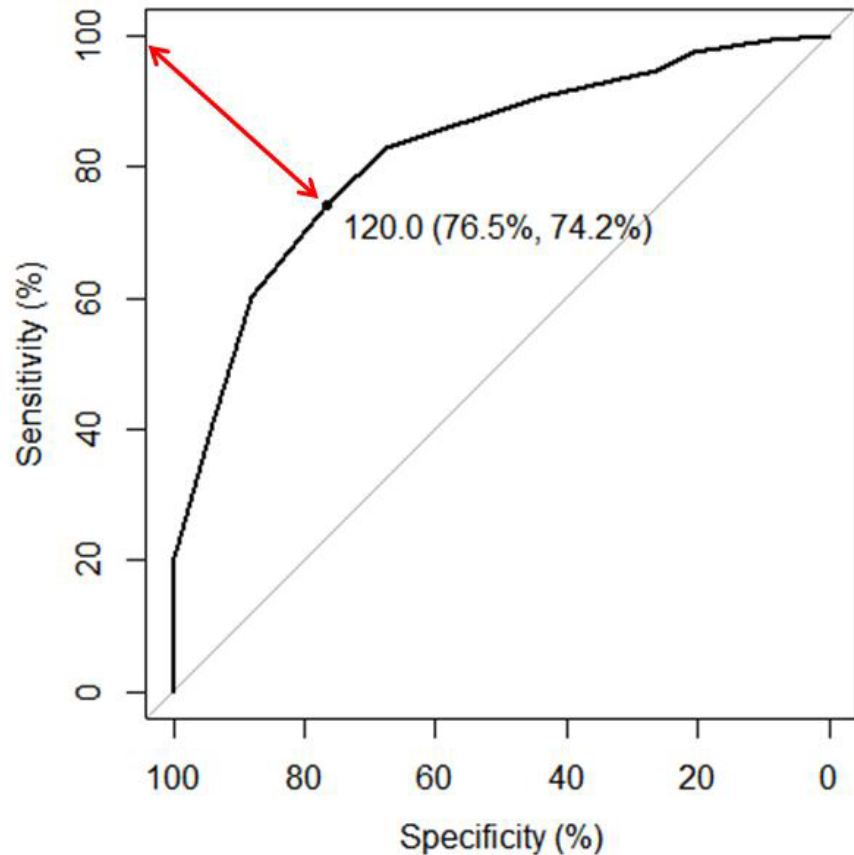
**Increased assay sensitivity (or lower cut point) doesn't dilute the clinical relevance of higher titer ADAs! *On the contrary, it strengthens the evidence.***



## Example 2

### *ADA Titer vs. Clinical efficacy*

*ROC curve demonstrates association of ADA Titer with Clinical Efficacy*



**Titer threshold at 120 provides ~ 76% Specificity & ~74% Sensitivity.**

**i.e., 74% of patients with favorable efficacy have Titer < 120, and 76% of patients with poor efficacy have Titer > 120.**

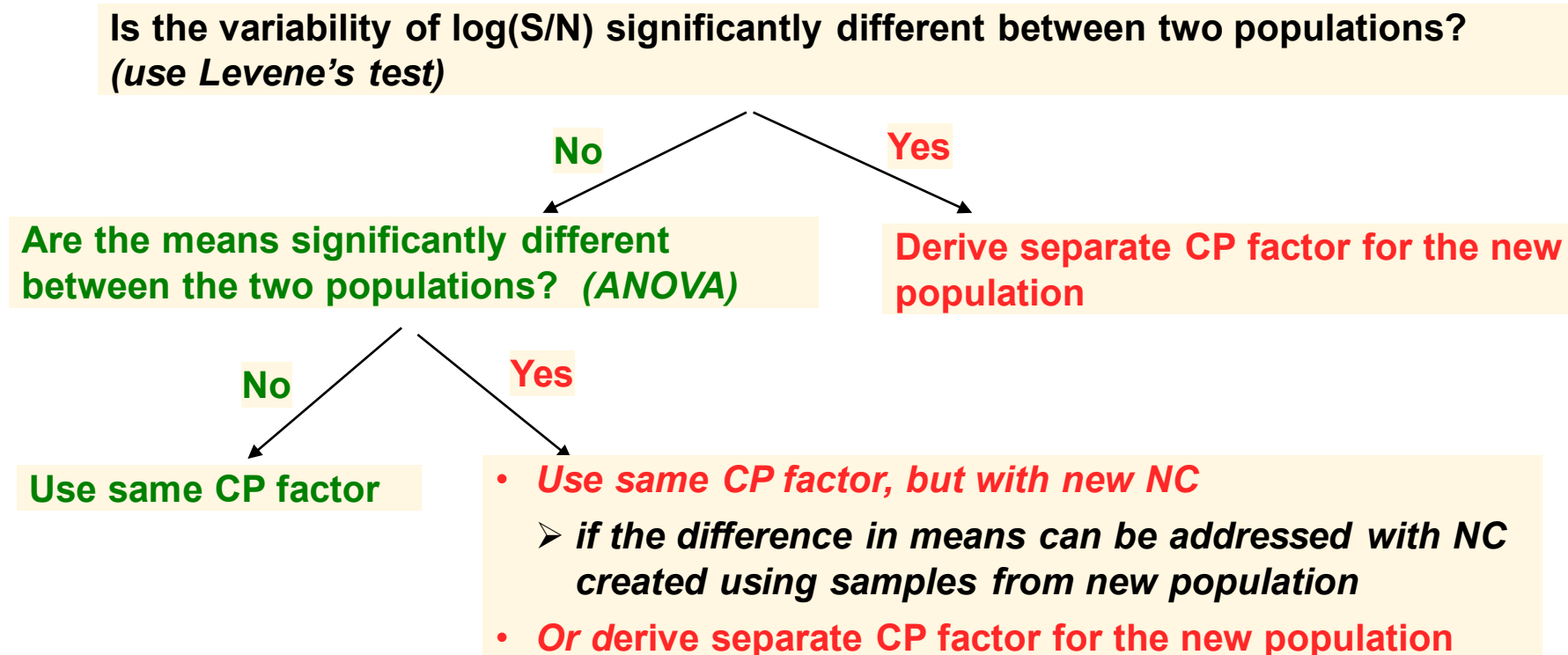
***Increased assay sensitivity, with higher incidence of low-titer ADAs does not change this conclusion or the optimal Titer threshold.***

***That is, it does not dilute this clinical impact.***

## *Extensions to other populations*

# Extensions of CP factor to other populations

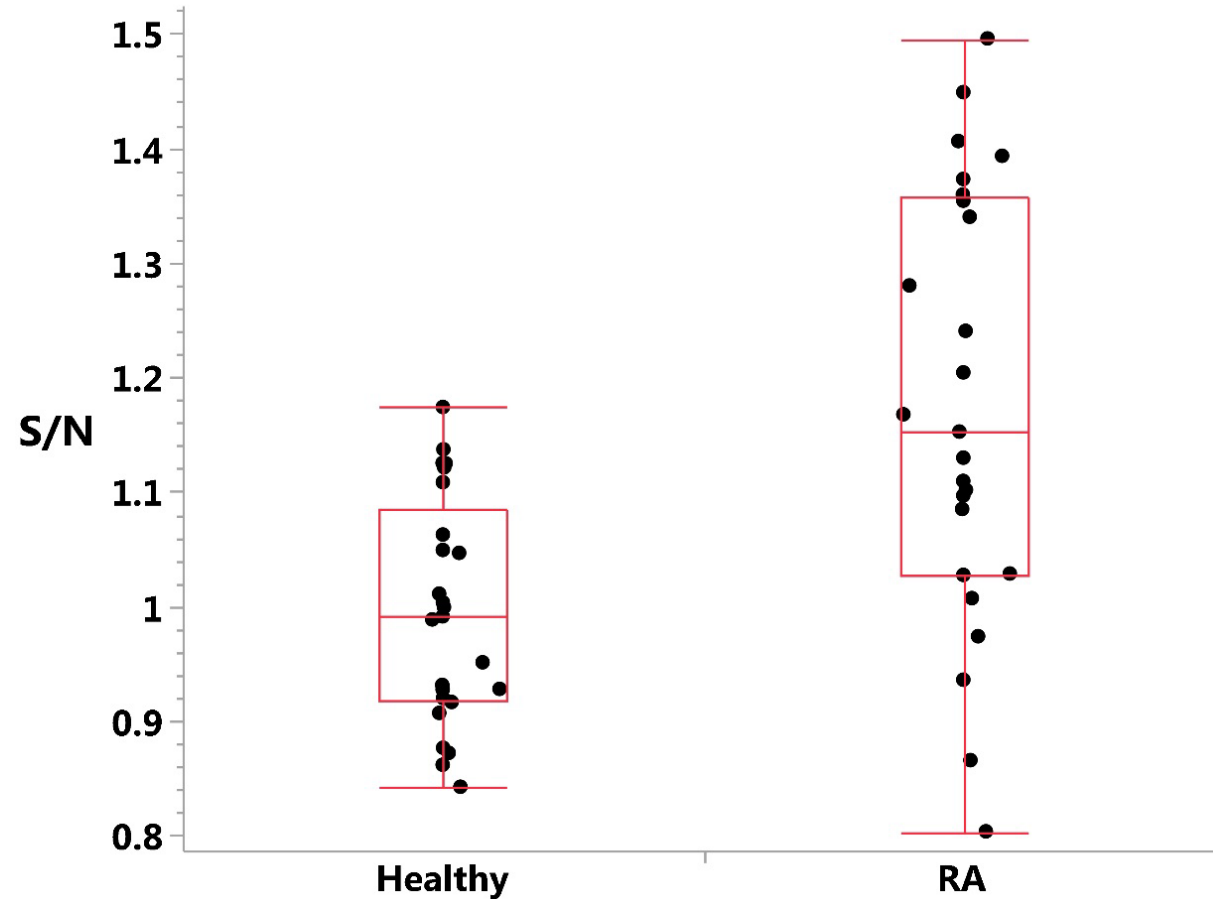
- For extending the CP to other populations (different disease, demographic, etc.), this decision tree may be useful.
- Need > 20 drug naïve individual sera from each population; assess outliers, distributions, etc.



Devanarayan et al., 2017

# Extensions of CP factor to other populations

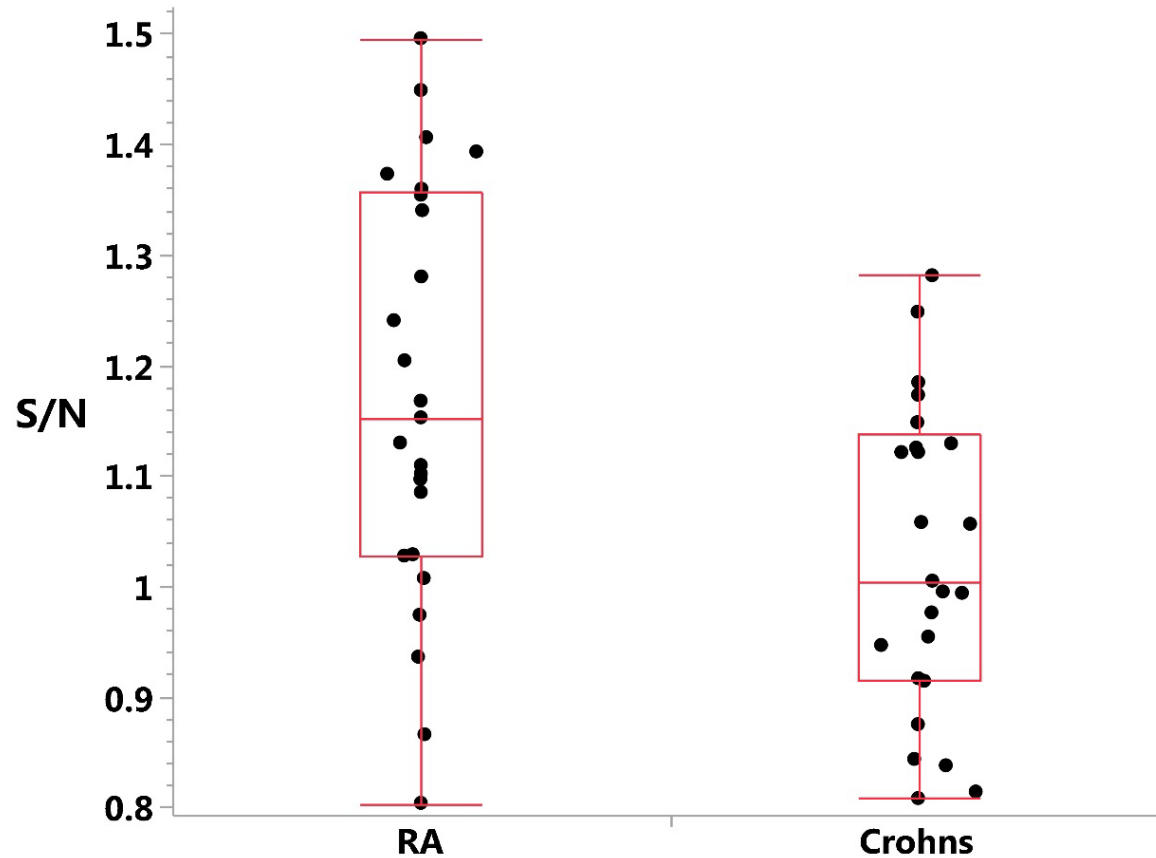
## Case study 1



- *Means and Variances are significantly different*
- *Need to derive a new CP factor for RA population*

# Extensions of CP factor to other populations

## Case study 2

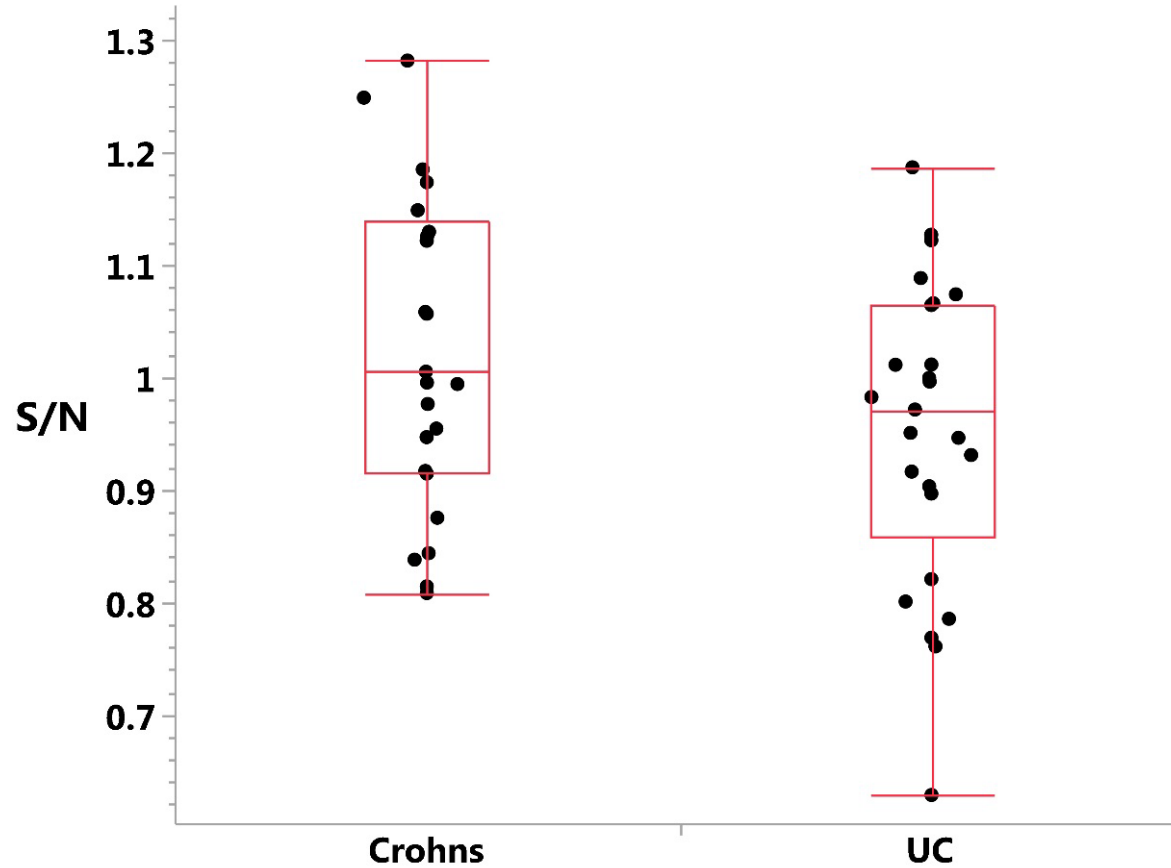


***Variances are not significantly different***  
***But the means are significantly different***  
***( $p=0.016$ )***

- If this difference is addressed with a new NC pool, use the same CP factor.***
- Or derive a new CP factor for Crohn's***

# Extensions of CP factor to other populations

## Case study 3



- *Means and Variances are not significantly different*
- *The same CP factor can be used for Ulcerative Colitis (UC) population.*

# System Suitability Criteria

# System Suitability Criteria (in-study plate acceptance)


## Screening Assay

- Data from all NC, LPC, and HPC samples tested during pre-study validation can be used to evaluate the following in-study plate acceptance criteria:
  - 99% upper limit of NC
  - 99% lower limit of LPC/NC ratio
  - 99% lower limit of HPC/NC ratio
  - $HPC/NC > LPC/NC > SCP$  factor
- If pre-study validation data are inadequate (say, < 20 plates) or if there is a change in assay reagents, additional in-study data can be used.
- Formulae provided in Shankar et al, 2008.
  - Alternative formulae such as beta-expectation tolerance interval (e.g., Mee, 1988) can be considered.



# System Suitability Criteria (in-study plate acceptance)

## Confirmatory Assay

- Data from all drug-spiked NC and LPC samples from pre-study validation can be used to evaluate the following in-study plate acceptance criteria:
- 99% upper limit of % inhibition of drug spiked NC.
- 99% lower limit of % inhibition of drug spiked LPC.
- %Inhibition of HPC and LPC > CCP > %Inhibition of NC  *only if the NC matrix is similar to the subject matrix.*
- Additional data from the in-study phase can be used if needed, esp. if pre-study validation data are inadequate or if there is a change in assay reagent.

# System Suitability Criteria (in-study plate acceptance)

## Titration Assays

- *Titer of HPC should be within the Minimum Significant Ratio (MSR) determined during pre-study validation. (USP <1106>)*
- Suppose we have the following from pre-study validation:
  - *Titer of HPC = 1000*
  - *MSR of Titters = 2.5*
- Then the *Titer of HPC in every in-study run (plate) should be within 400 to 2500* ( $1000 / 2.5 = 400$ ,  $1000 \times 2.5 = 2500$ )
- Alternatively, 1-2 dilution fold criteria can be applied, but note that it is very subjective and doesn't take into account of the Titer variability.

## *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 pre-existing 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.

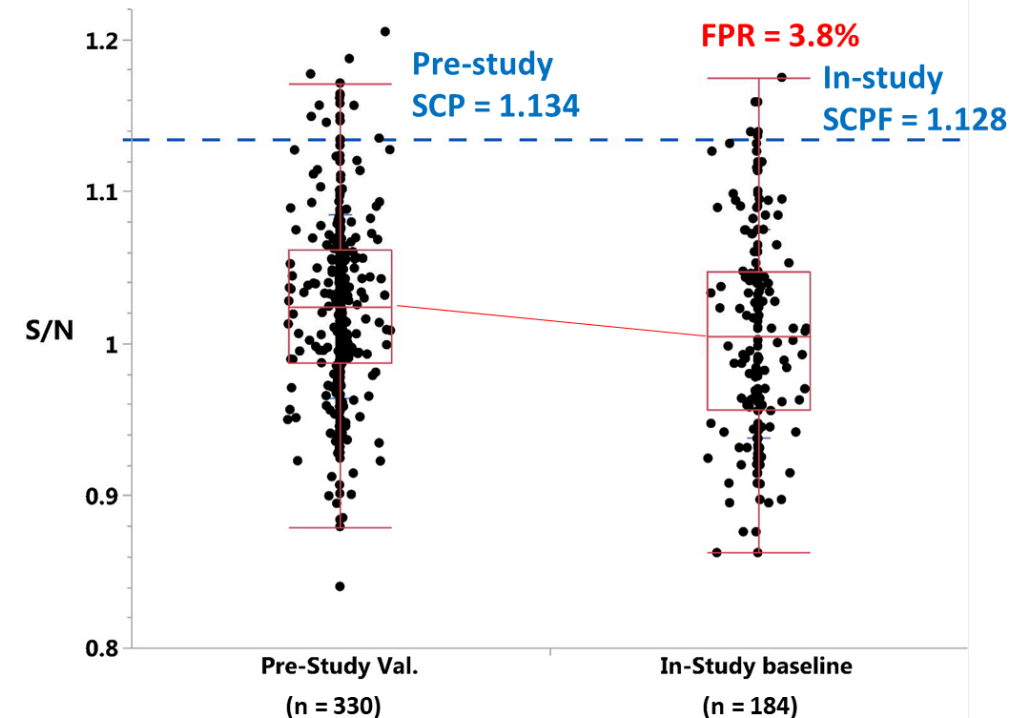
In-study FPR > 11% raises concerns mostly for the sponsor, but it depends on the nature of the assay and clinical study.

- 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.

Primary 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  $\geq 3$  plates/runs and  $\geq 2$  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.

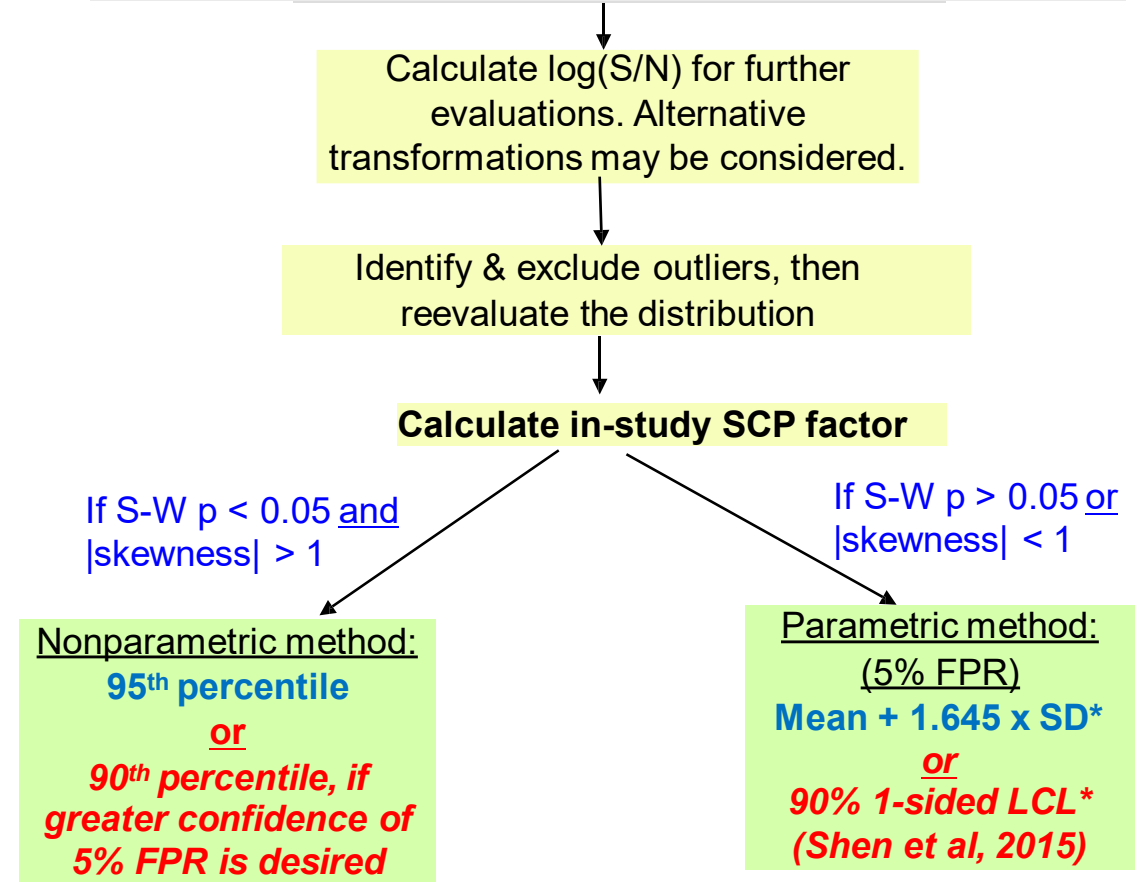


# In-study Screening Cut Point; analysis

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.

> 50 pre-dose subject sera, after excluding pre-existing Ab.  
One test result per subject serum will suffice but spread the samples across  $\geq 3$  runs and  $\geq 2$  analysts



\* 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 pre-study validation).
- As noted for the in-study SCP evaluation, balanced design is not necessary; one test result per subject serum will suffice for this in-study CCP calculation.

# In-study Confirmatory Cut Point; analysis

Spike the same pre-dose subject sera used for in-study SCP evaluation with excess study drug. Test the unspiked and spiked counterpart samples together on the same plate. Spread the subject sera across  $\geq 3$  runs and  $\geq 2$  analysts. Calculate % inhibition.

Identify and exclude outliers. Evaluate distribution. Consider transformation if needed.

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

Transform to reduce skewness (e.g., log)

If  $|Skewness| < 1^*$

If S-W  $p < 0.05$  and  $|skewness| > 1$

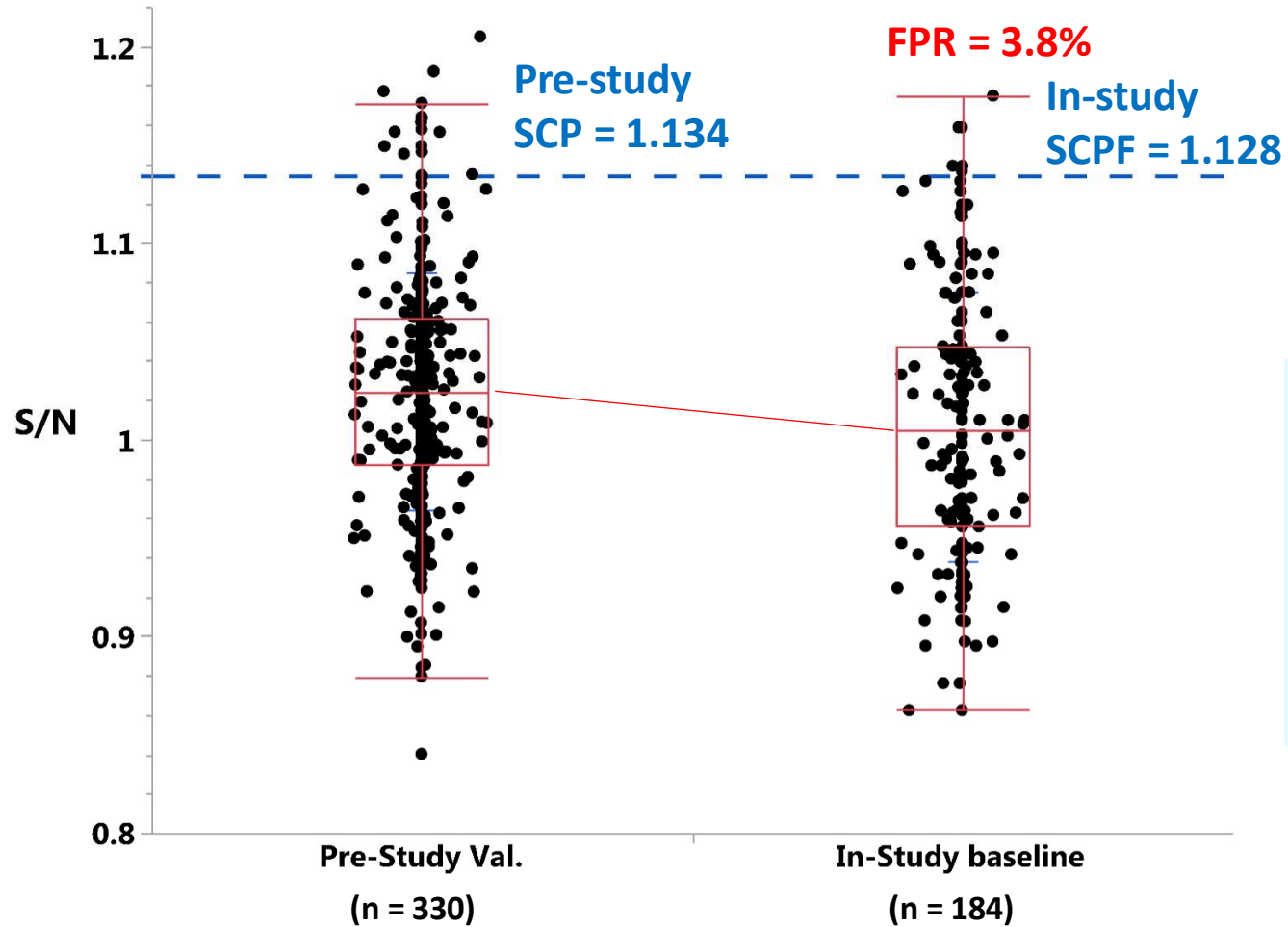
If S-W  $p > 0.05$  or  $|skewness| < 1$

Nonparametric method  
CCP = 99<sup>th</sup> percentile  
**or**  
**97<sup>th</sup> percentile, if greater confidence of 1% FPR is desired**

Parametric method  
(1% FPR)  
CCP = Mean + 2.33 x SD\*  
or  
**80% 1-sided LCL\* (Shen et al, 2015)**

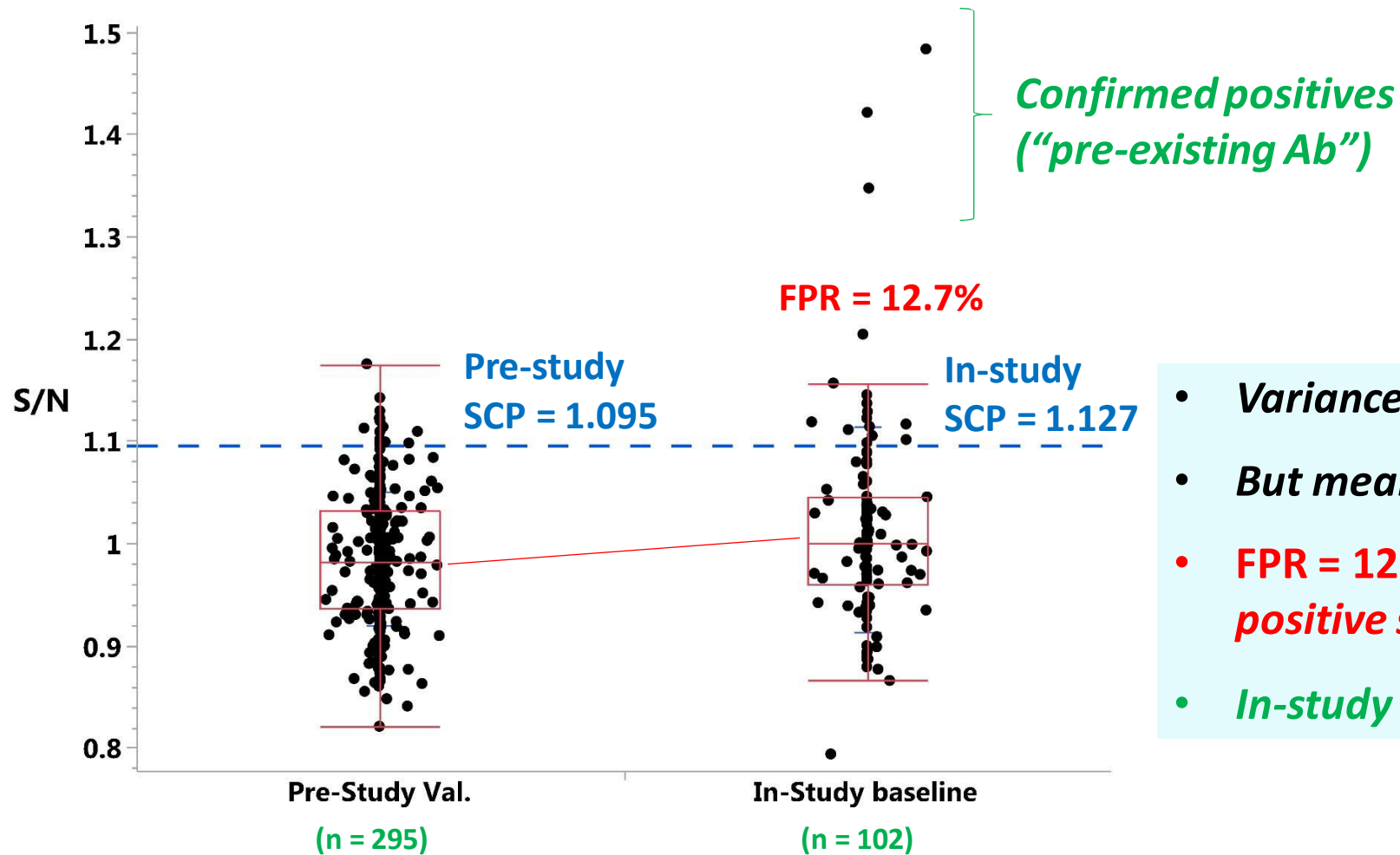
\* Use Median instead of Mean, and 1.4826xMAD instead of SD to ensure robustness when there are borderline outliers.

# Case-study 1



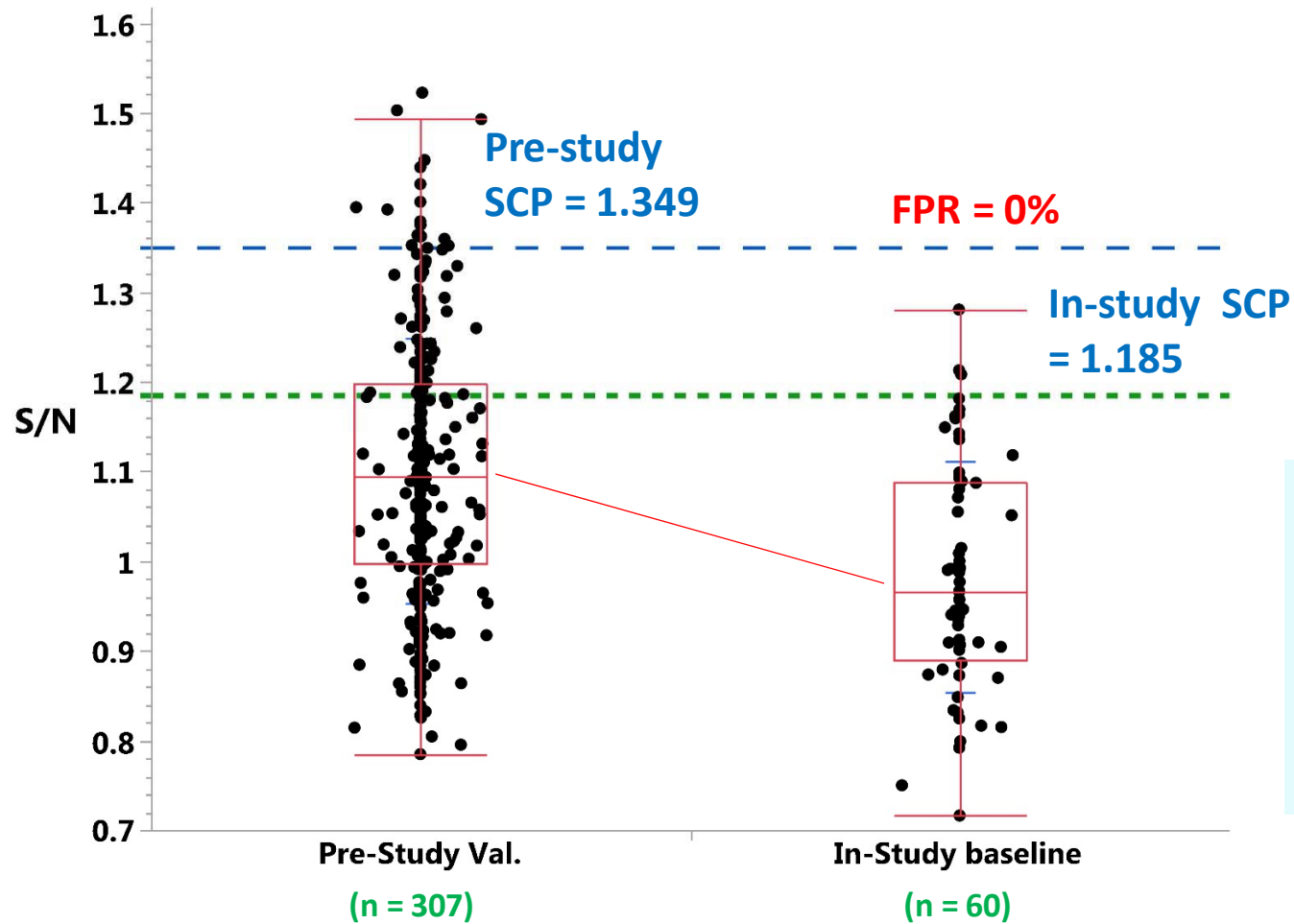
- *Although distributions look very similar, means & variances are significantly different, mostly due to large “n”*
- **FPR = 3.8%**
- *Don't need to use in-study CP*

# Case-study 2



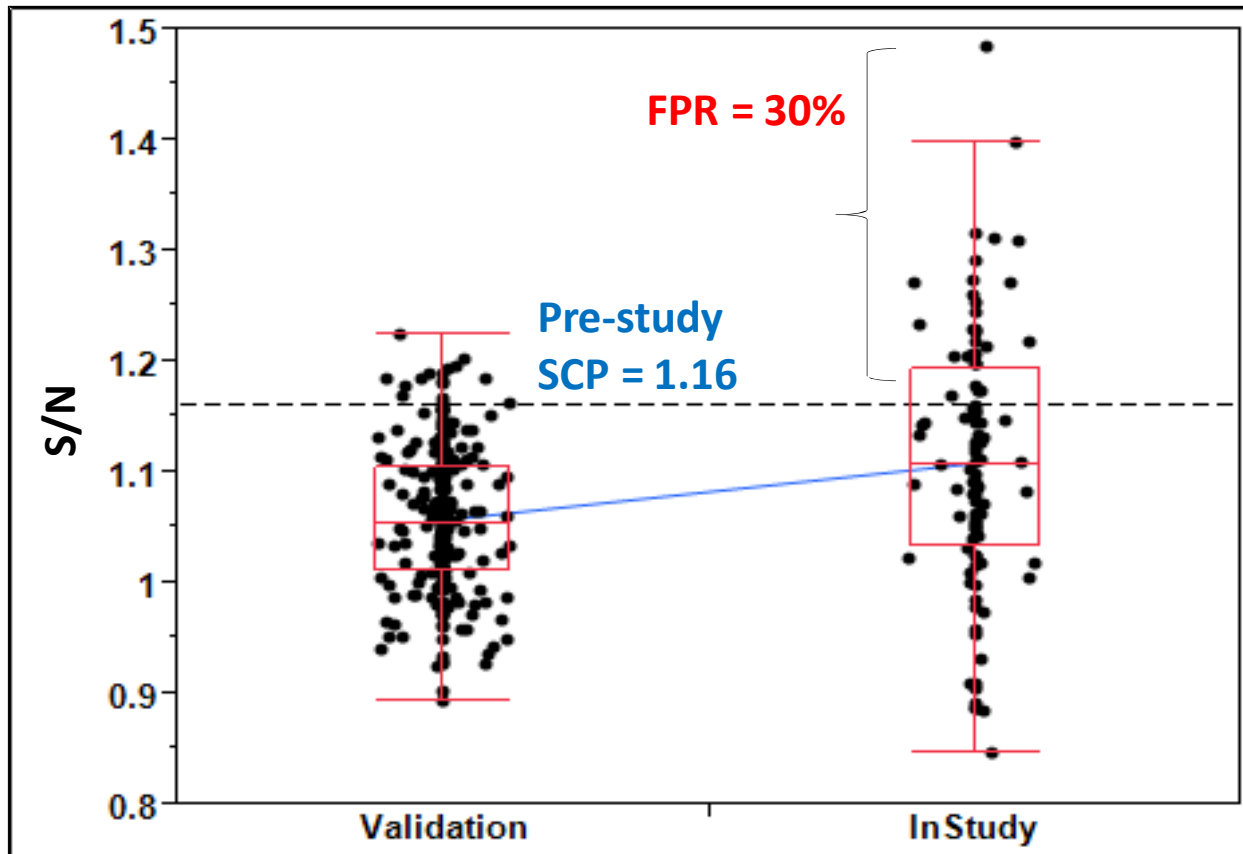
- *Variations are not significantly different.*
- *But means are significantly different.*
- *FPR = 12.7% (after excluding 3 confirmed positive samples)*
- *In-study CP may be used*

# Case-study 3



- *Variations are not significantly different.*
- *But means are significantly different.*
- **FPR = 0%**
- *Need to use in-study CP*

# Case-study 4



- *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)*

- <https://www.ncbi.nlm.nih.gov/books/NBK92434/>
  - [https://www.ncbi.nlm.nih.gov/books/NBK92434/#immunometh.Immunogenicity\\_Cut\\_Point\\_Anal](https://www.ncbi.nlm.nih.gov/books/NBK92434/#immunometh.Immunogenicity_Cut_Point_Anal)



# 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 pre-validation 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

- This tool is originally in the form of a "template file", i.e., .XLT.
- 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  $\geq 3$  runs and  $\geq 2$  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 [robust parametric method](#). 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 [parametric method](#) 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 [nonparametric method](#). 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 [alternative methods](#) (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/>
- [https://www.ncbi.nlm.nih.gov/books/NBK92434/#immunometh.Immunogenicity\\_Cut\\_Point\\_Anal](https://www.ncbi.nlm.nih.gov/books/NBK92434/#immunometh.Immunogenicity_Cut_Point_Anal)