Immunogenicity cut point setting & outlier evaluation – part 2

V. Devanarayan, GlaxoSmithKline

European Bioanalysis Forum (EBF) Training Day

Practical Aspects of Immunogenicity

March 23-24, 2021, Virtual

Outlier evaluation and criteria

Outliers, Outliers, & Outliers

Hypothetical Lab Colleague

- Why was observation X removed as an outlier when it's in the middle of the data set? Why are similar values left in the analysis?
- These data are distributed normally and all observations are part of the normal variability of the population. So why did you remove so many outliers?

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 = 75th percentile, Q1 = 25th 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.

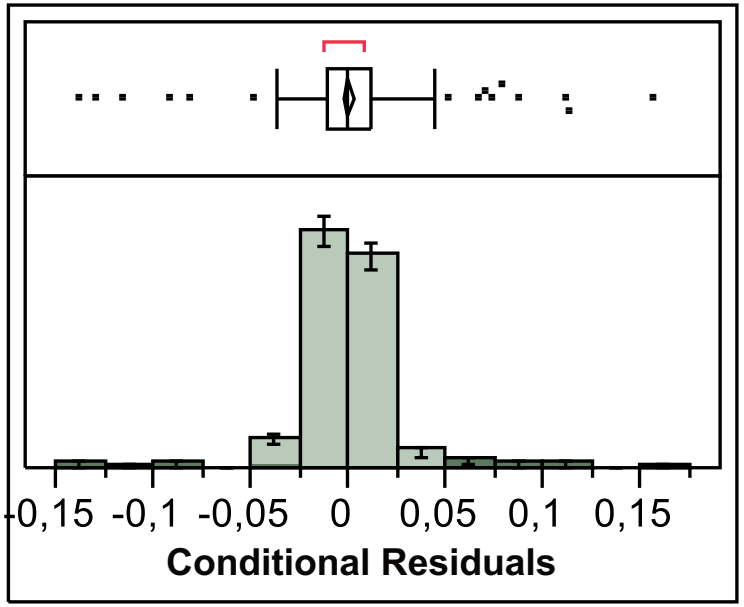
- Normality test may fail due to long tails, but as long as distribution is reasonably symmetric (|skewness| < 1), parametric method can be used.

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

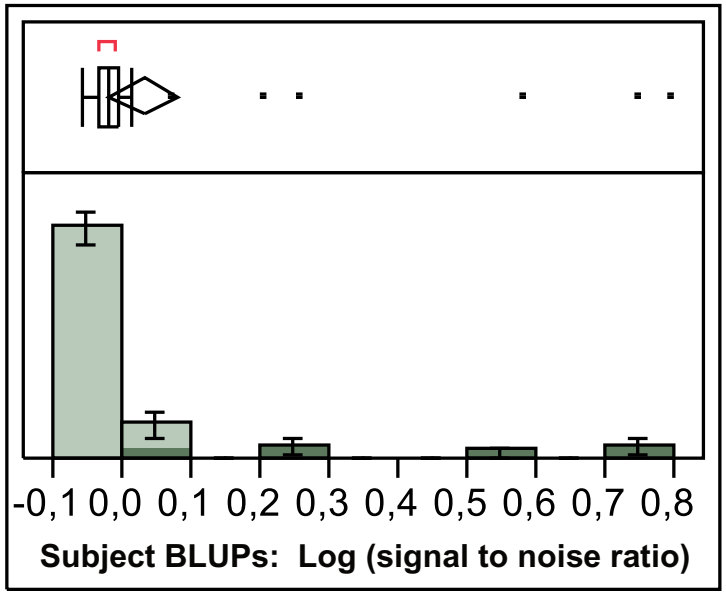
Analytical Outliers

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).

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 in more samples** affect the SCP?

Outlier criteria

Tukey's outlier box-plots by default is based on the following criteria:

- *High outliers:* ***> Q3 + 1.5 x IQR***
- *Low outliers:* ***< Q1 - 1.5 x IQR***
  - *Q3 = 75th percentile, Q1 = 25th percentile*
  - *IQR = Inter quartile range = Q3 – Q1* criteria:

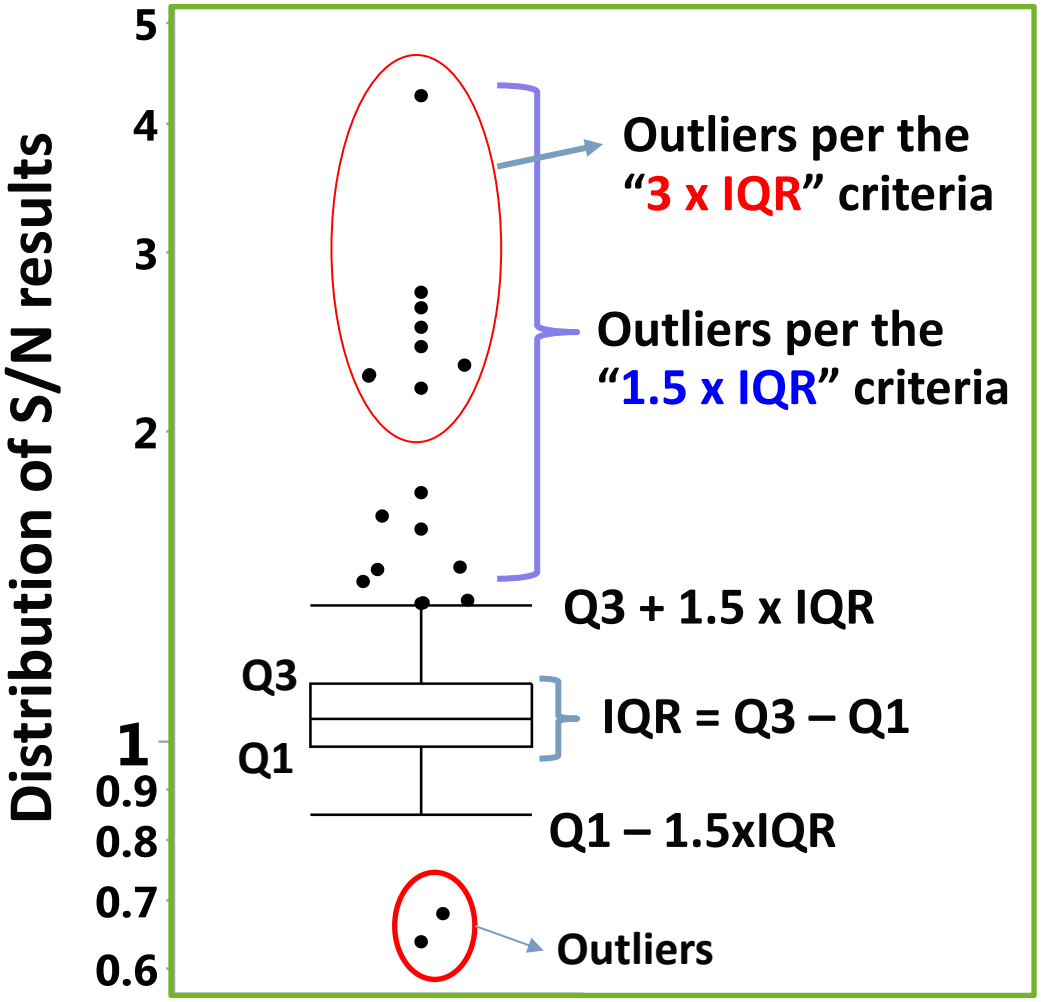
Due to concerns about "low cut points", "too many outliers", etc., this criteria gets subjectively changed to **3xIQR**.

- Several talks at conferences some publications bemoan about "excessive" outlier removal.

Such subjectivity is not necessary if robust approach is used, i.e., Median/MAD instead of Mean/SD.

- **Results from 1.5xIQR vs 3xIQR are usually similar if Median/MAD is used.**

Interpretation of 1.5xIQR and 3xIQR outlier criteria



For ~ normal distribution,

1.5 x IQR criteria is equivalent to **Mean ± 2.67 x SD**

- ~ covers 99.2% of the samples
- Similar to 3xSD criteria widely used in other applications.

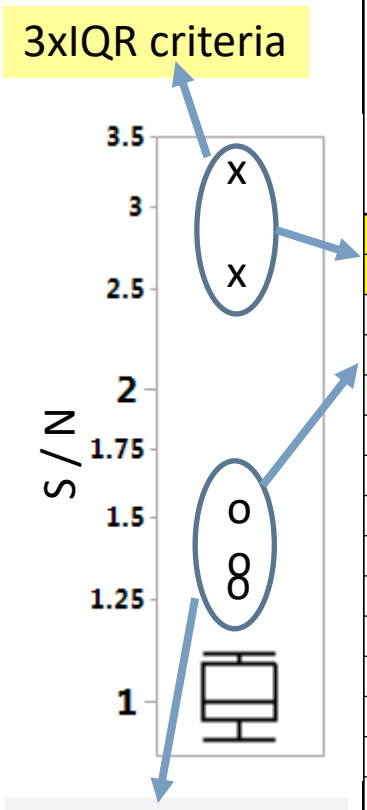
3 x IQR criteria is equivalent to **Mean ± 4.67 x SD**

- ~ 99.9997% of the samples

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

Illustration: Robustness to outliers

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



S/N	log(S/N)	Absolute Deviation: | log(S/N) - Median |		
		All Data	w/o 2 outliers	w/o 5 outliers
3.267	0.514	0.497		
2.682	0.428	0.412		
1.574	0.197	0.180	0.193	
1.325	0.122	0.106	0.118	
1.278	0.106	0.090	0.102	
0.919	-0.037	0.053	0.041	0.035
1.112	0.046	0.029	0.042	0.047
1.086	0.036	0.019	0.031	0.037
1.088	0.037	0.020	0.032	0.038
0.999	0.000	0.017	0.005	0.001
1.022	0.009	0.007	0.005	0.011
1.057	0.024	0.007	0.020	0.025
0.988	-0.005	0.022	0.010	0.004
0.997	-0.001	0.018	0.006	0.000
0.919	-0.037	0.053	0.041	0.035
0.983	-0.007	0.024	0.012	0.006
1.088	0.036	0.020	0.032	0.038
0.952	-0.022	0.038	0.026	0.020
0.961	-0.017	0.034	0.022	0.016
0.977	-0.010	0.027	0.015	0.009

0.497 =| 0.514 – median(log(S/N)) |

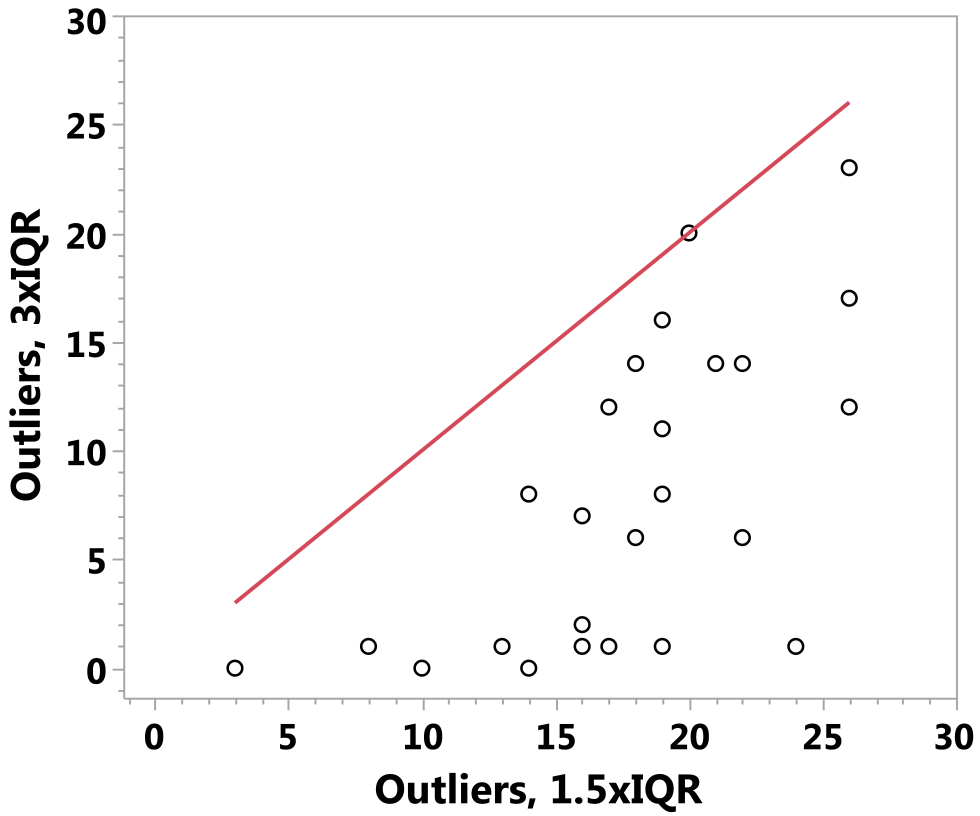
MAD = Median of all these absolute deviations

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

MAD = 1.4826 x MAD*

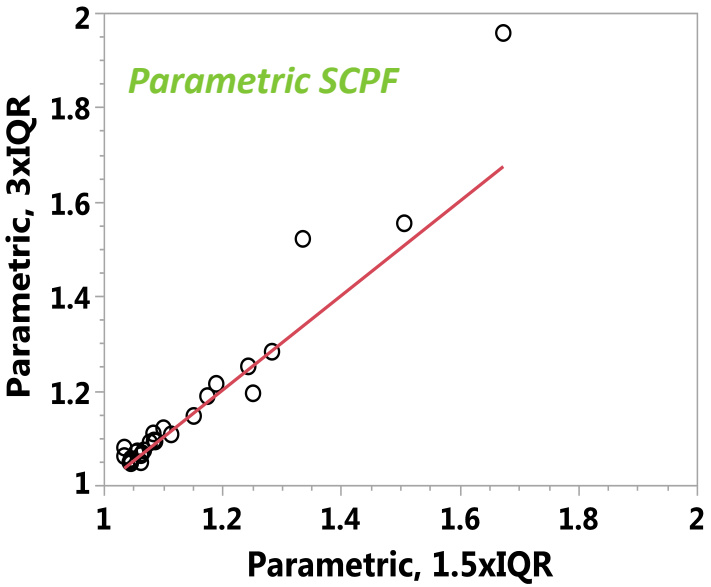
"Robust parametric" is more resistant to borderline outliers.

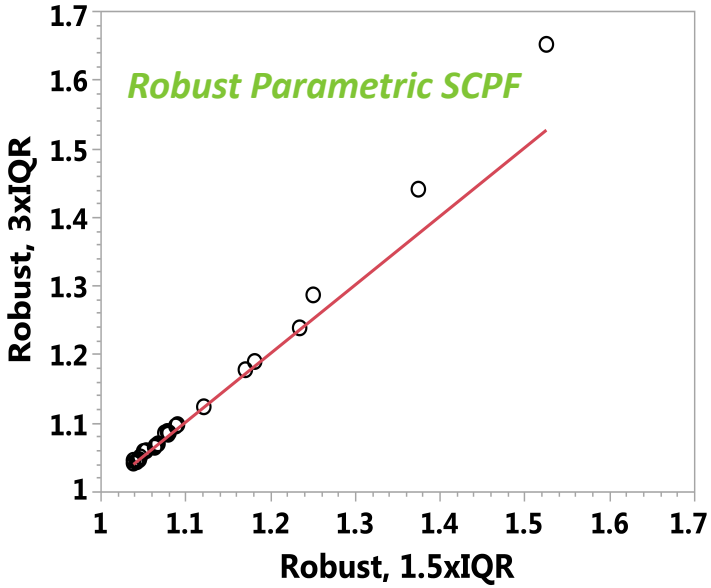
1.5xIQR vs. 3xIQR criteria for 25 assays



- The number of outliers identified via 1.5xIQR and 3xIQR criteria are shown here.
- 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 method is used, CP results between 1.5xIQR vs. 3xIQR are mostly similar.**
- ***3xIQR criteria is not the cure for "low" cut points!***
- **1.5xIQR is ~ 2.7xSD, widely used in statistics literature, thus a good default.**

Low cut points, Low signal, and False Positive Rate

"Low" Cut Points

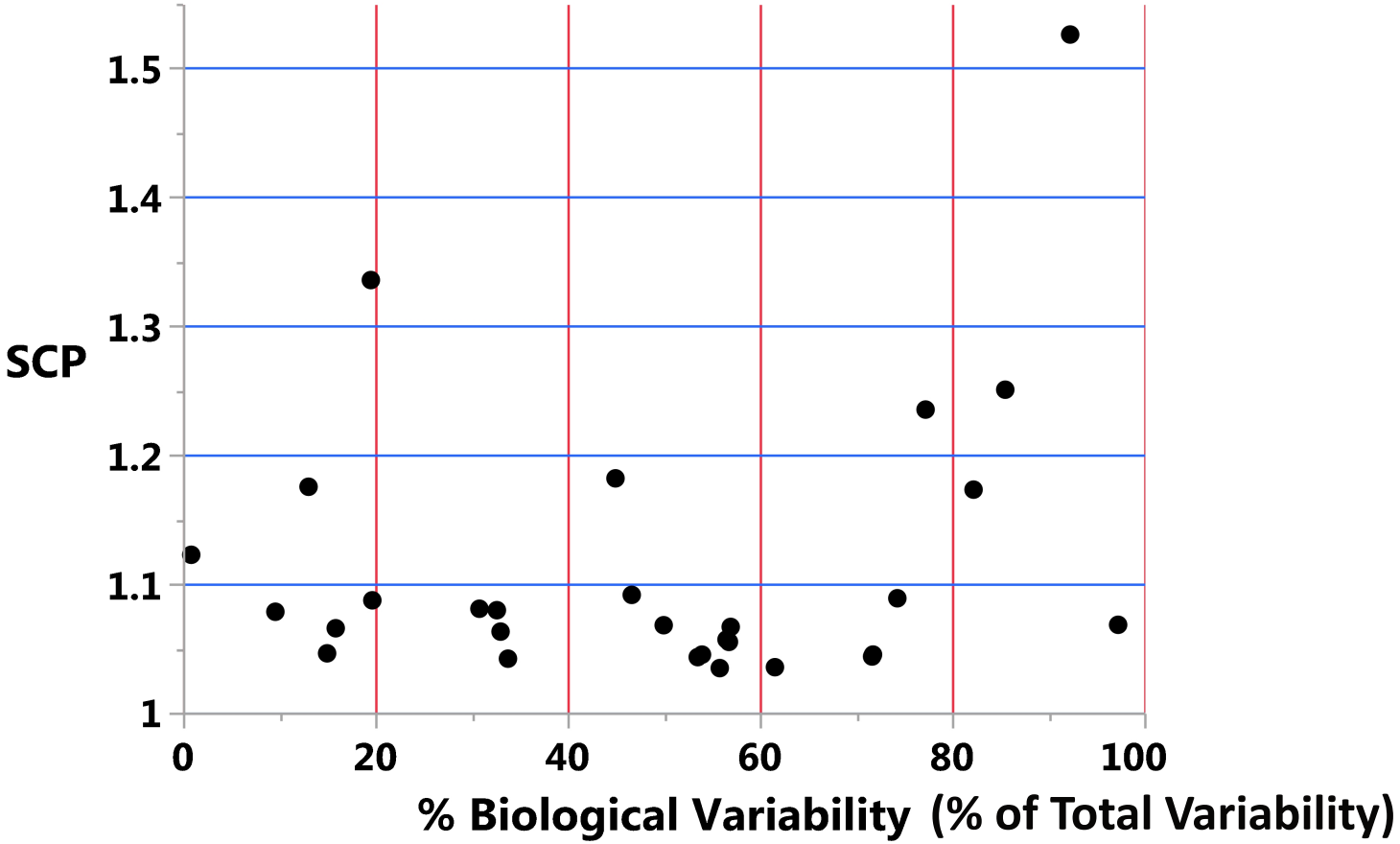
SCPF can be quite low, sometimes < 1.1 and often < 1.2

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 previous section!*

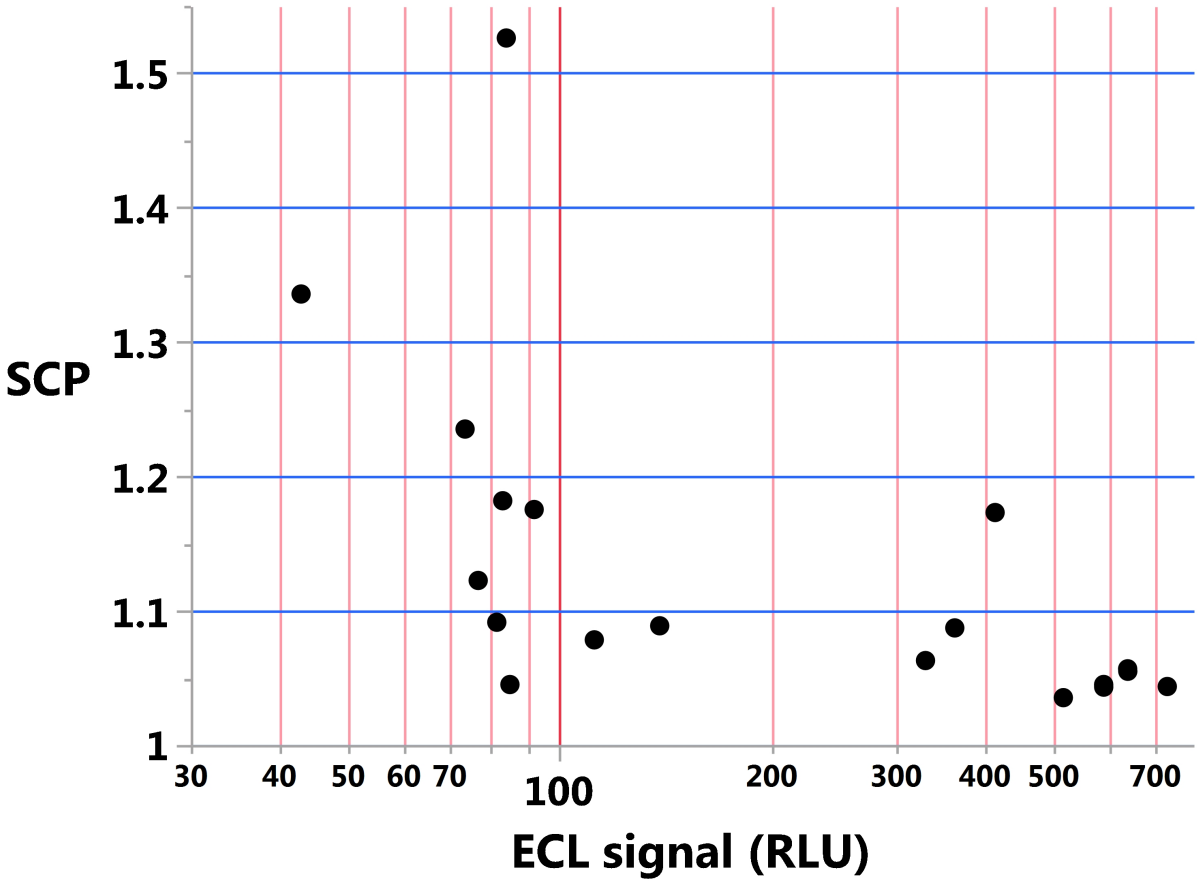
These questions will be addressed via retrospective evaluation of 25-30 assays; most of these assays have SCPF < 1.2

SCP factor vs. Biological variability



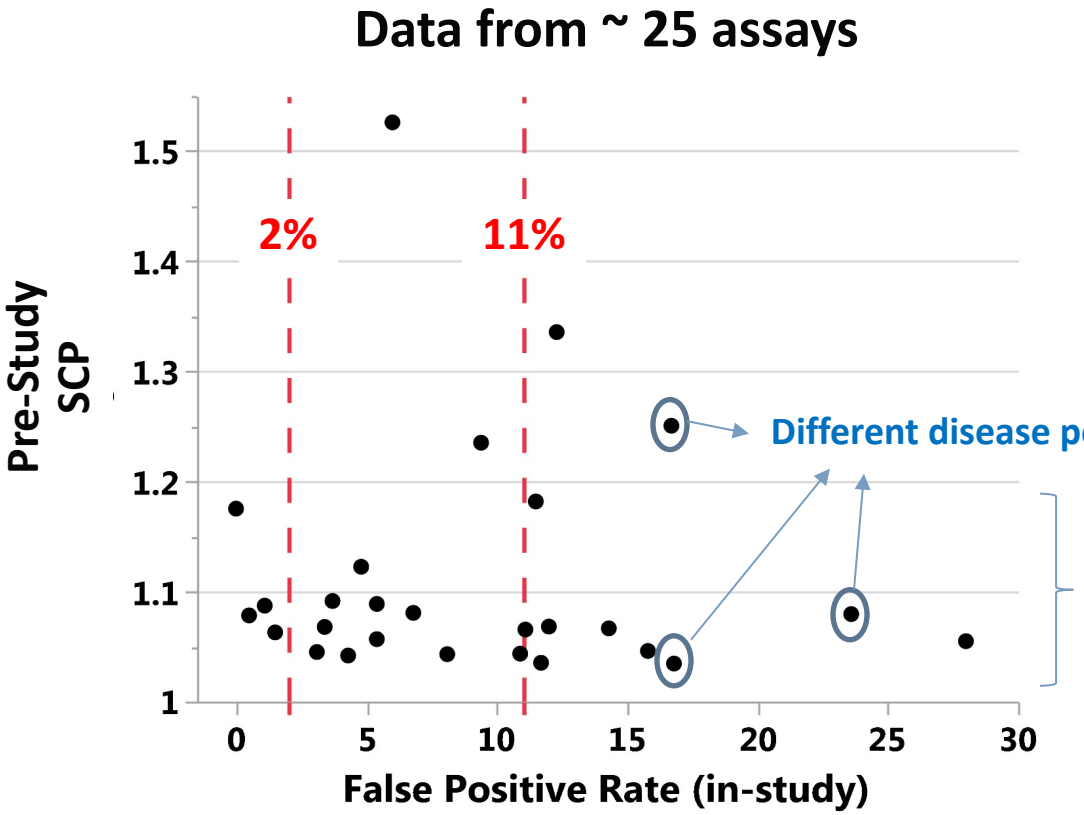
Assays with high biological variability can also have Low Cut Points!

SCP factor vs. Assay signal



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

Low cut points & in-study FPR



Low SCP does not always result in high in-study FPR

Thank you for your interest & attention!

Questions ?