Practical advice & insights on Immunogenicity cut points and some assay validation parameters

V. Devanarayan, PhD, FAAPS

Charles River Laboratories

Training Course

European Immunogenicity Platform (EIP) Symposium

Lisbon, Portugal, Feb 25, 2019

Acknowledgments

Ron Bowsher, Ph.D., B2S Life Sciences

Wendell Smith, Ph.D., B2S Life Sciences

Members of AAPS & EBF/EIP Immunogenicity working groups

Outline

1. Cut point (CP) evaluation process
2. Outlier criteria and misconceptions
3. Upcoming Excel-based tool for CP evaluations (*informal; non-GxP*)
4. Insights on “Low” CP, “Low” Signal, FPR & Clinical Relevance
5. Extensions to other populations
6. In-study CP
7. Titer precision and Titer CP
8. Criteria for Treatment-boosted ADA
9. System suitability criteria
10. Sensitivity & Low Positive Control

Cut point evaluation process

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:

- >= 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:

- >= 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		
			$S_1 – S_{18}$	$S_{19} – S_{36}$	$S_{37} – S_{54}$
A_1	R_1	P_1	X		
		P_2		X	
		P_3			X
	R_2	P_1		X	
		P_2			X
		P_3	X		
	R_3	P_1			X
		P_2	X		
		P_3		X	
A_2	R_4	P_1	X		
		P_2		X	
		P_3			X
	R_5	P_1		X	
		P_2			X
		P_3	X		
	R_6	P_1			X
		P_2	X		
		P_3		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 CP Experimental 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	

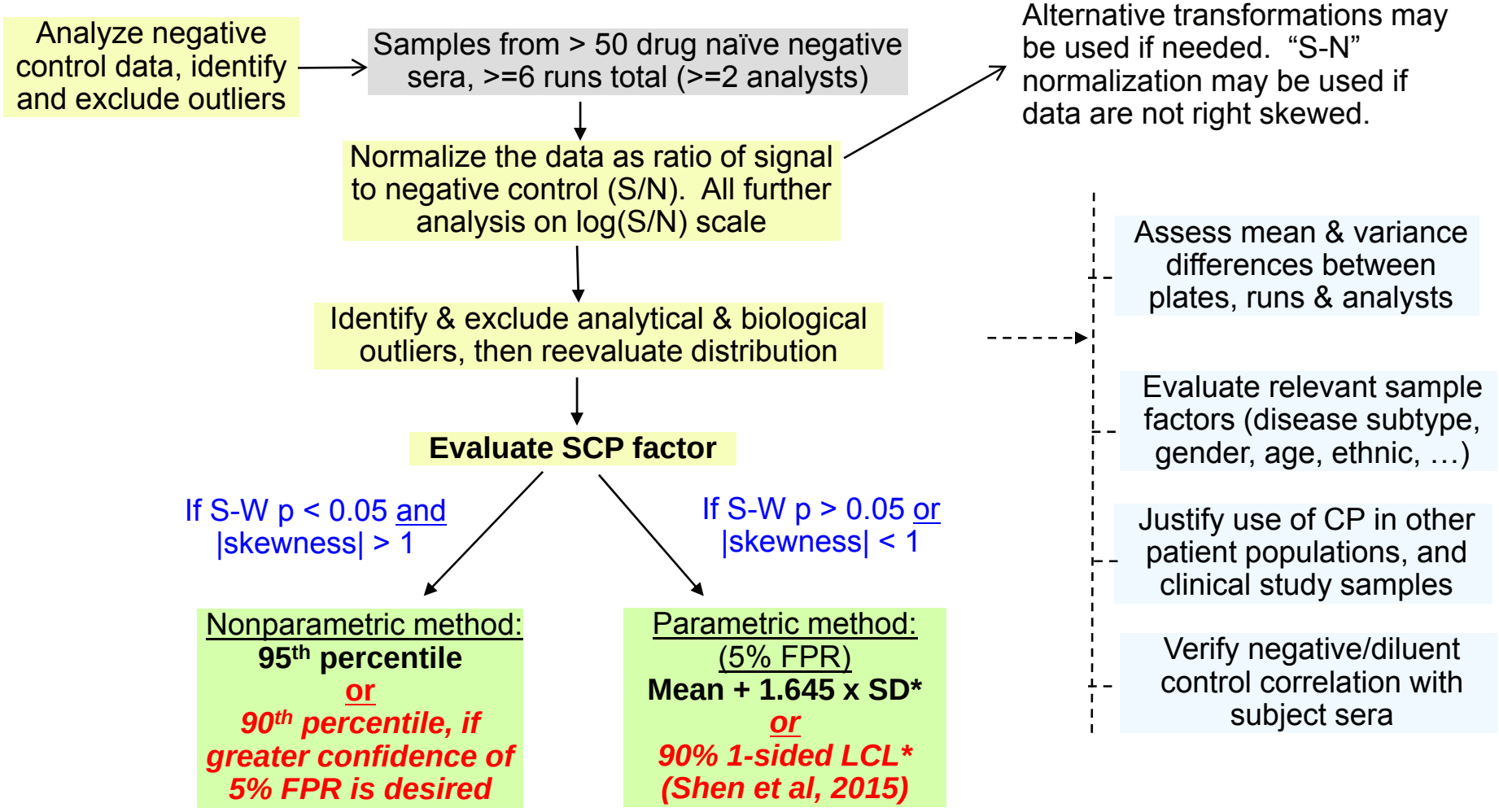
- *NC & LPC with and w/o drug, MPC & HPC*
- *17 subject sera per plate.*

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 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 CP Evaluation (SCP)

Simplified flow-scheme that should work in most cases

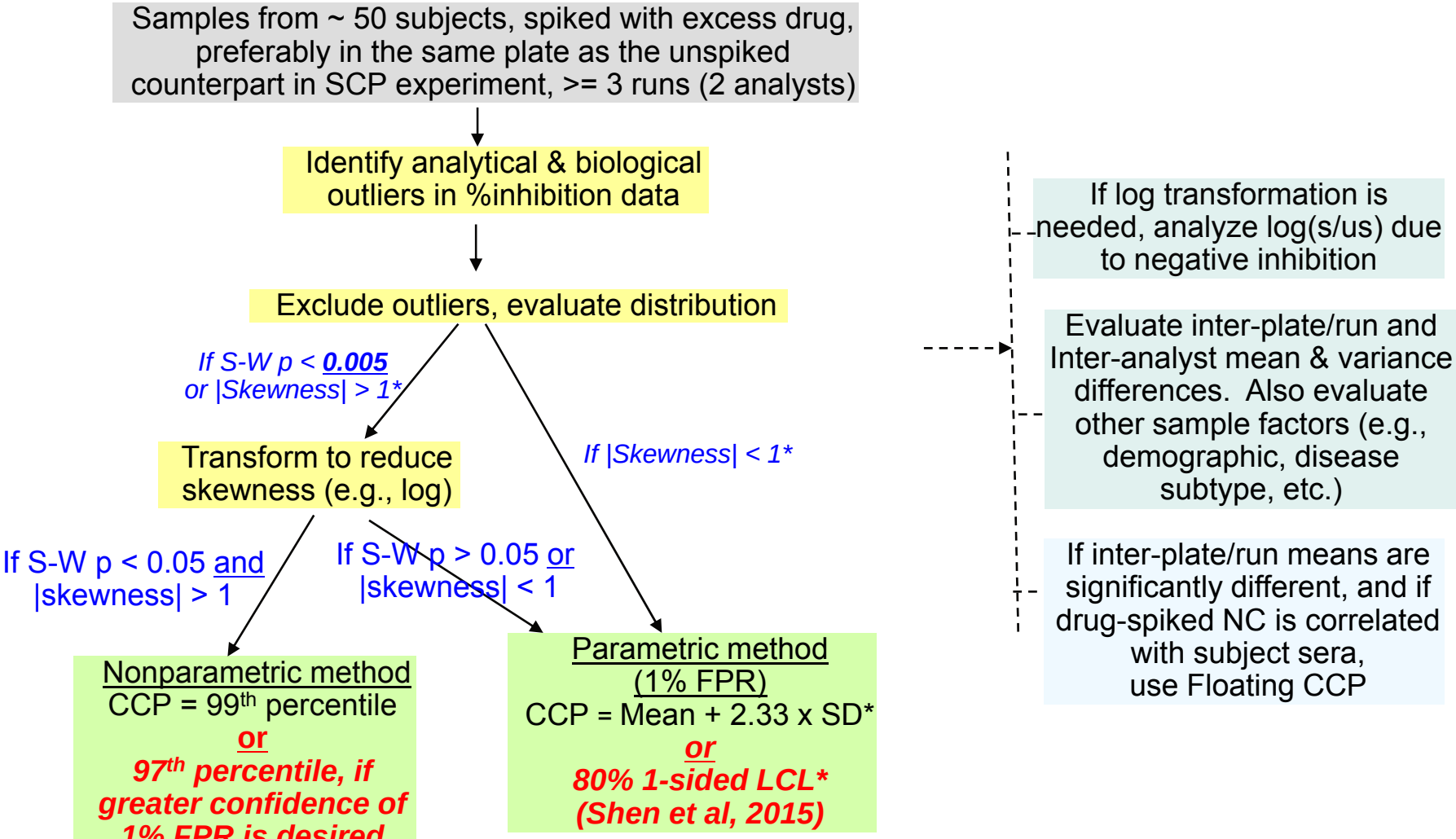


** Use Median instead of Mean, and 1.4826xMAD instead of SD to ensure robustness when there are borderline outliers.*

Devanarayan et al., 2017

Confirmatory Cut Point (CCP) Evaluation

Simplified flow-scheme that should work in most cases



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

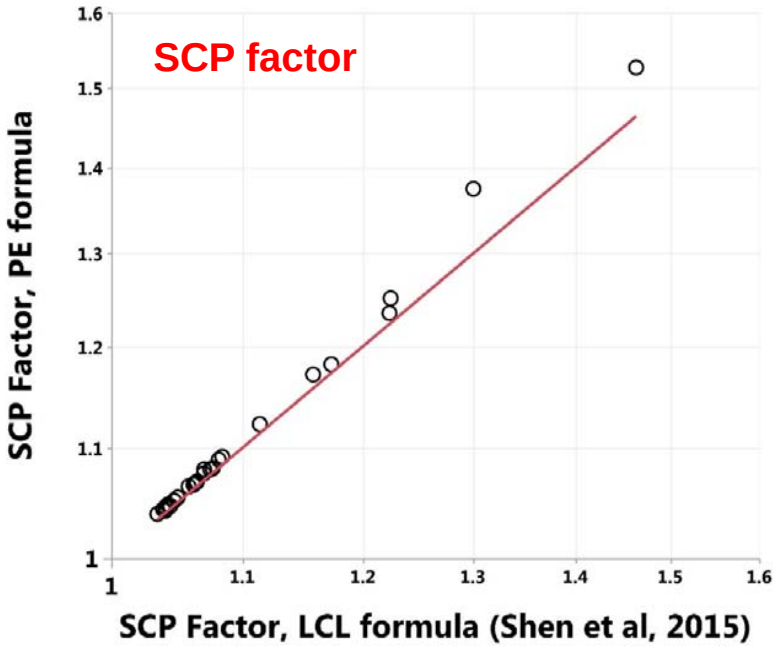
Devanarayan et al., 2017

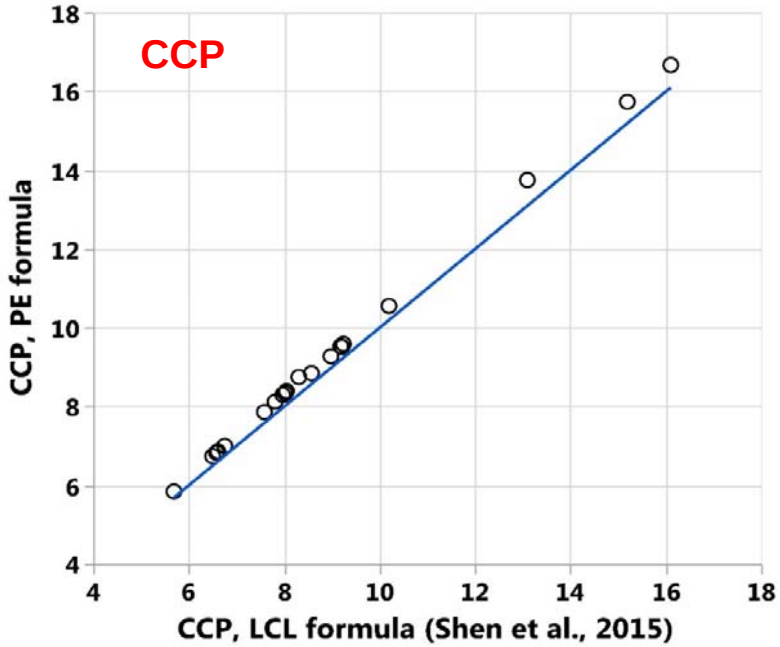
Point estimate (PE) vs. Lower confidence limit (LCL)

- Screening CP was defined to yield *approximately* 5% FPR.
- "Point Estimate" (**PE**) of 95th 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 95th percentile (SCP) and 99th 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 difference with several datasets….

SCP & CCP results from PE vs. LCL formulae

Data from ~ **25 assays** (mostly ECLs).





- CP values from the PE & 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 small sample size (e.g., early-phase in-study samples)*

Outlier criteria & misconceptions

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.)

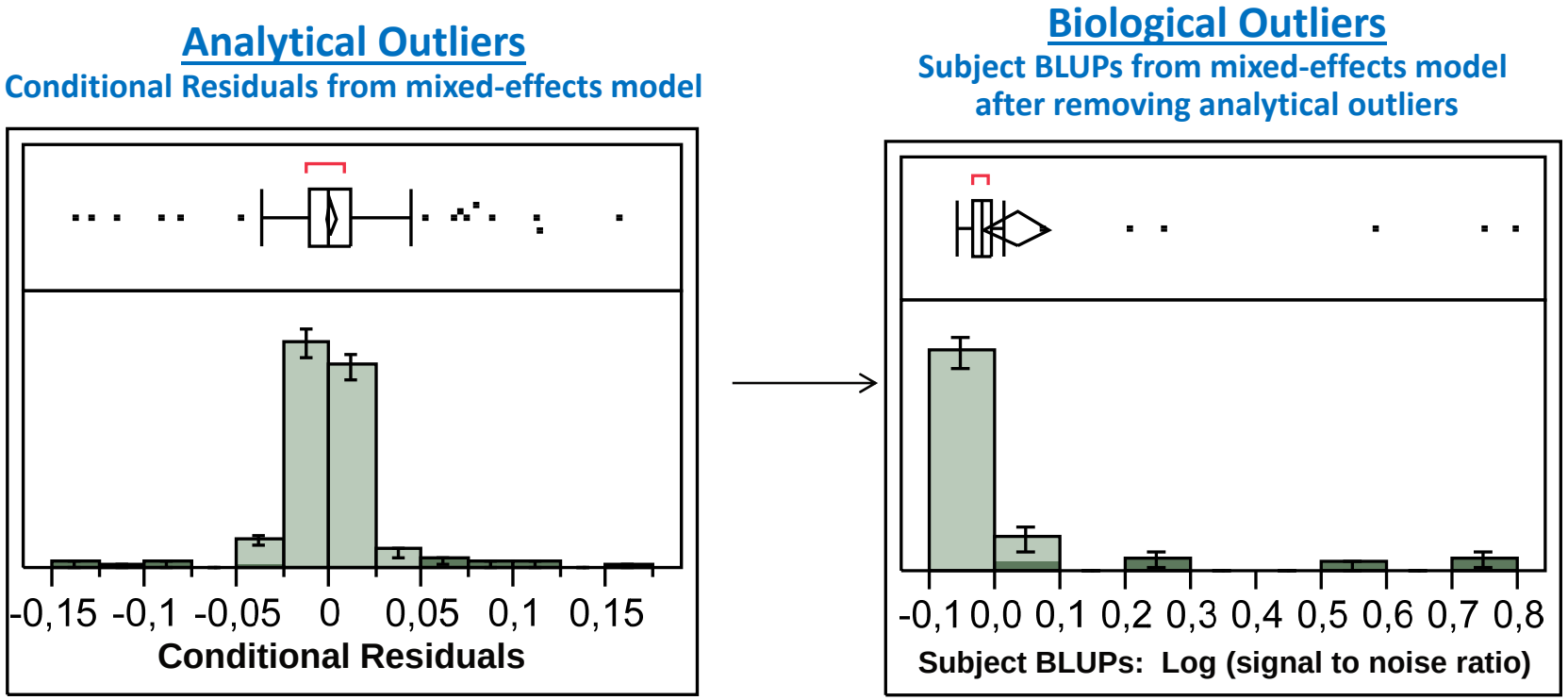
5. Refit the model without these analytical outliers, and then obtain Best Linear Unbiased Predictor (BLUP) for each subject.
6. Apply the "outlier box-plot" criteria on these subject BLUPs to identify the outlier subjects → ***Biological*** *outliers*
7. Refit the model without all the analytical & 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



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:

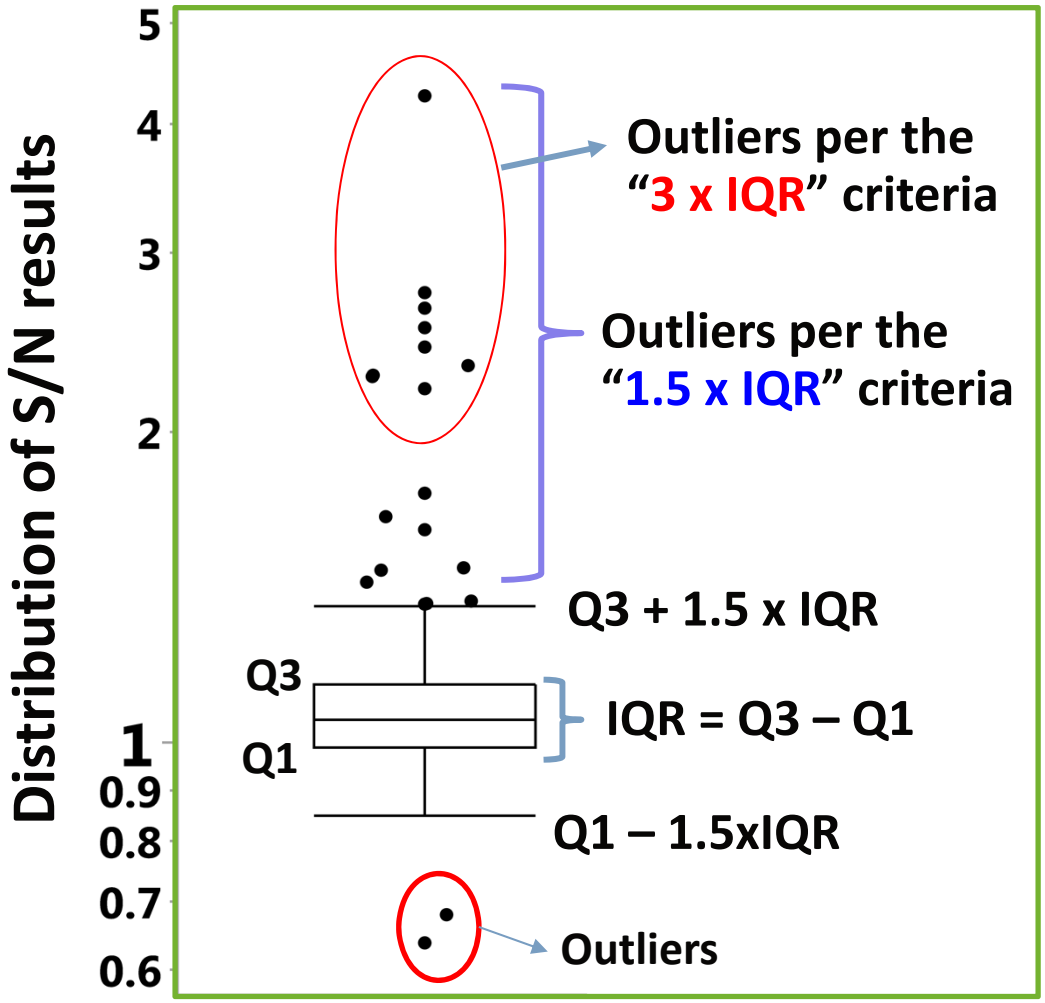
Due to concerns about "low cut points", "too many outliers", etc., this criteria gets subjectively changed to **3xIQR**.

- Several talks at conferences a recent publication (e.g., Kubiak et al, 2018) 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?

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 **variability** of continuous measurements.

Equals the median of the absolute deviations from median:

- *MAD = median(|Xi – median(x)|)*
- *SD is replaced in the formula by **1.4826 x MAD***

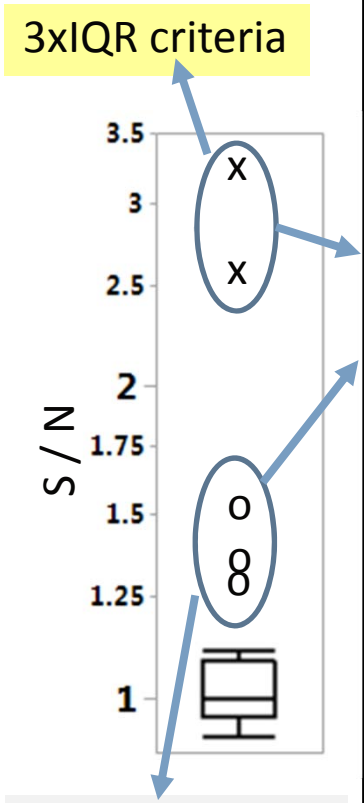
More resistant to the outliers than SD

- 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. *Subjective manipulations of outlier criteria will not matter much.*

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)) |

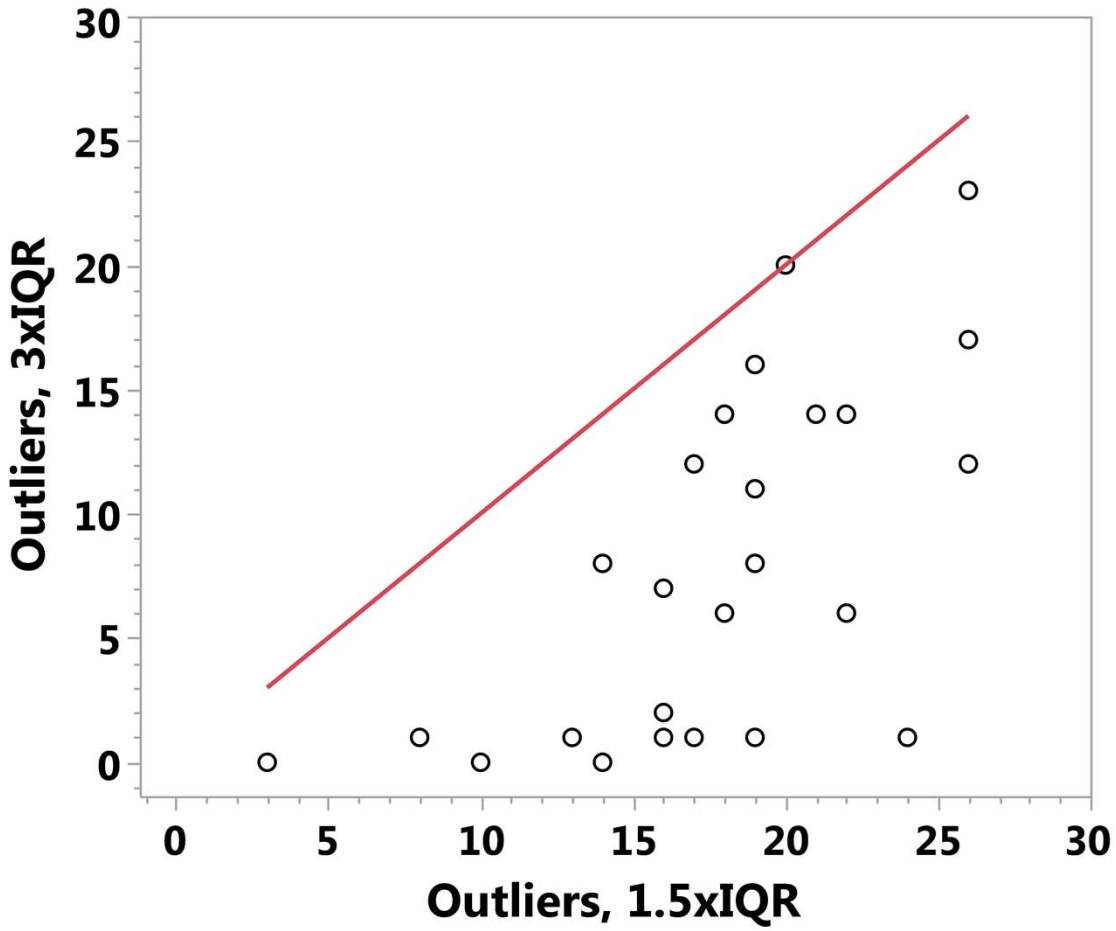
MAD = Median of all these absolute deviations

	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*

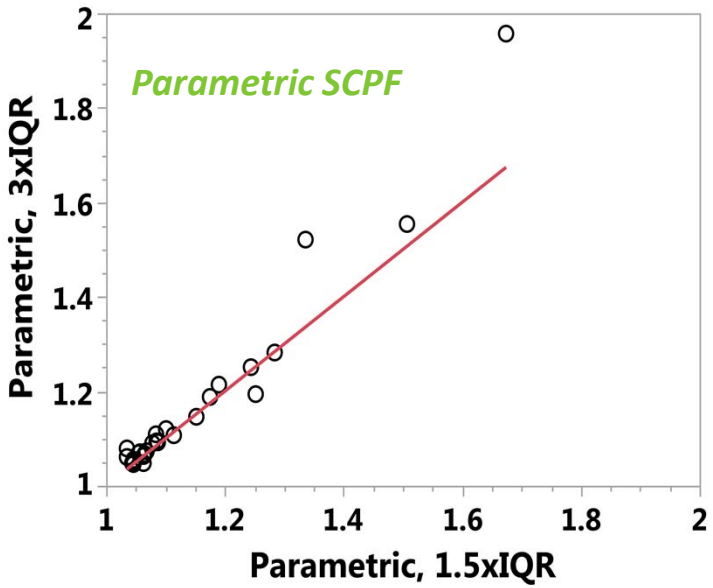
"Robust parametric" is more resistant to borderline outliers.

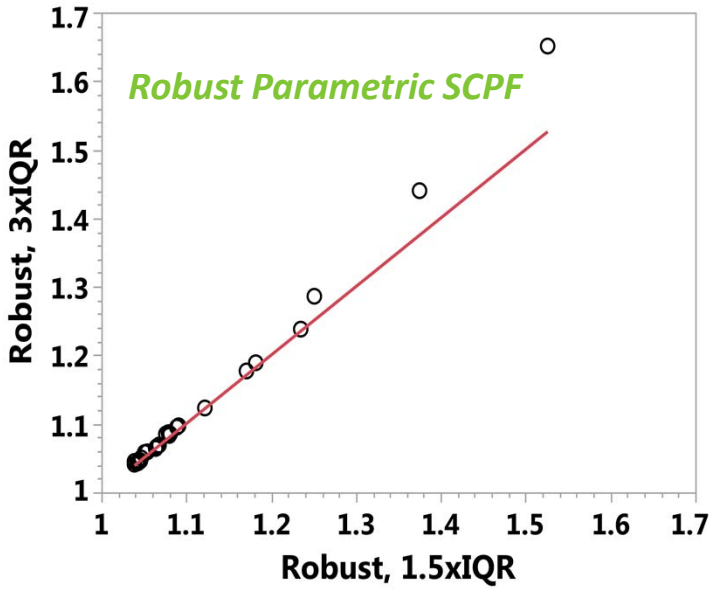
1.5xIQR vs. 3xIQR criteria for 25 assays



- 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.**

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 leaves out 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 and non-normal distribution.
- *Be Robust or Go-bust* ☺

Low cut points, Low signal, FPR & clinical relevance

"Low" Cut Points

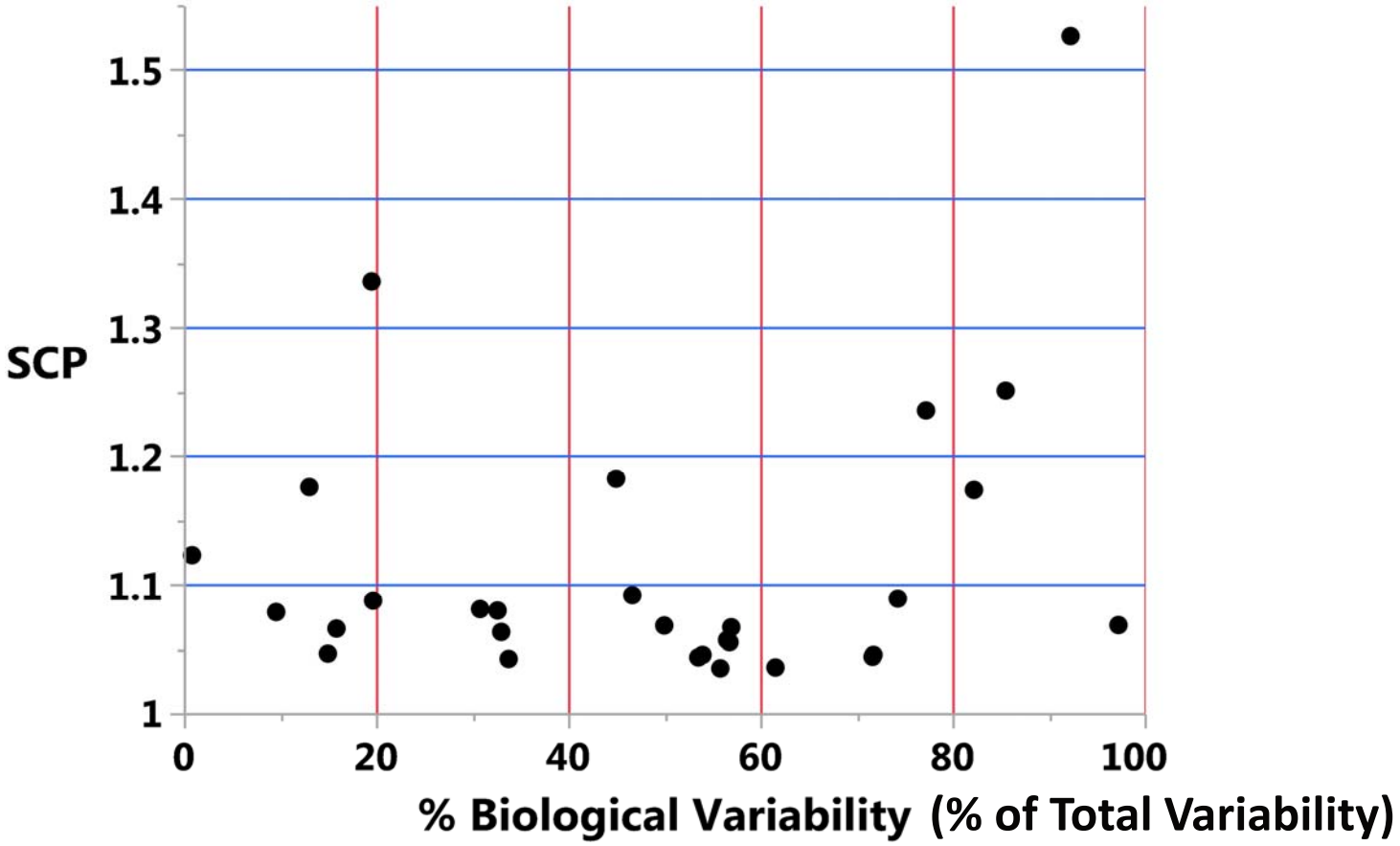
SCPF can be quite low, sometimes < 1.1 and often < 1.2

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 previous section!*
- Will this *dilute the overall clinical relevance* of the ADA results?

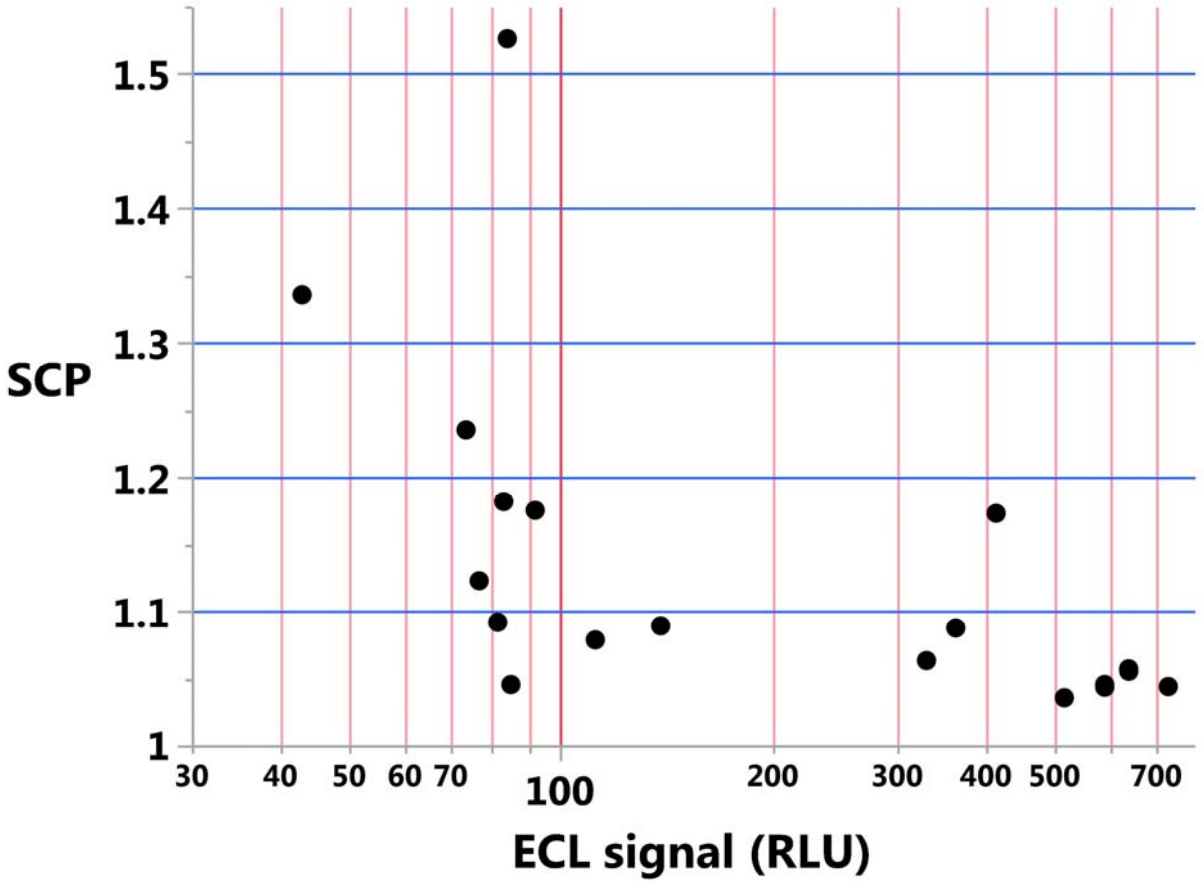
These questions will be addressed via retrospective evaluation of 25-30 assays; most of these assays have SCPF < 1.2

SCP factor vs. Biological variability



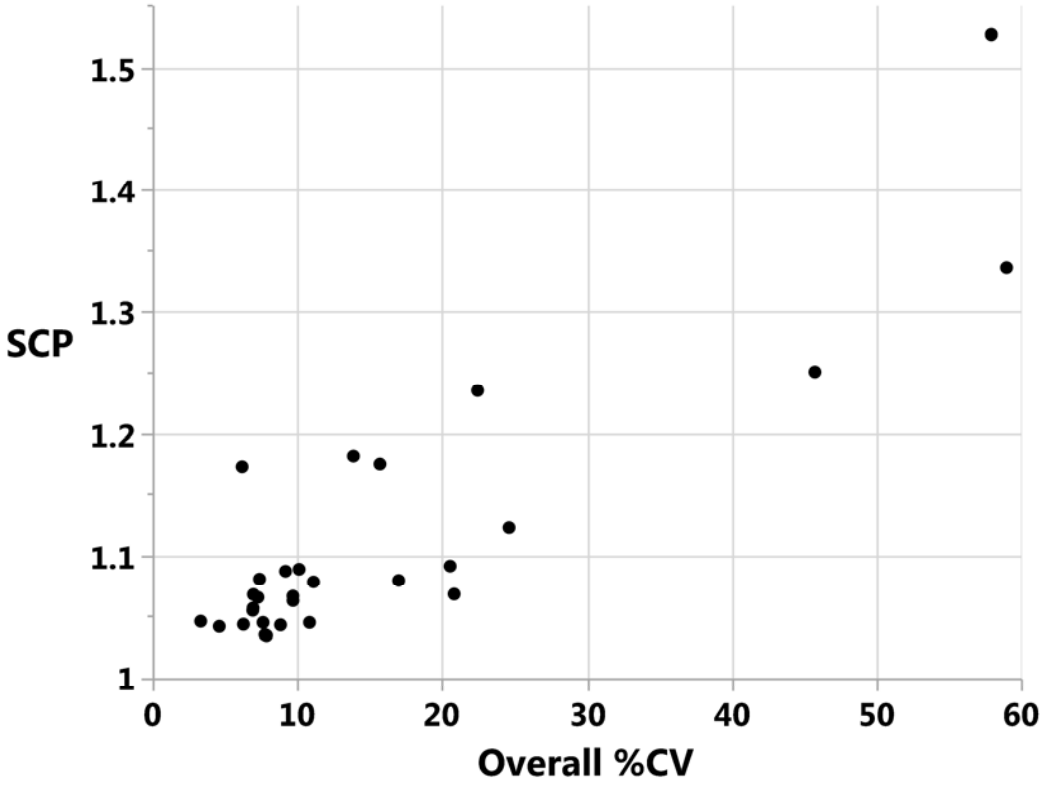
Assays with high biological variability can also have low SCP factors.

SCP factor vs. Assay signal



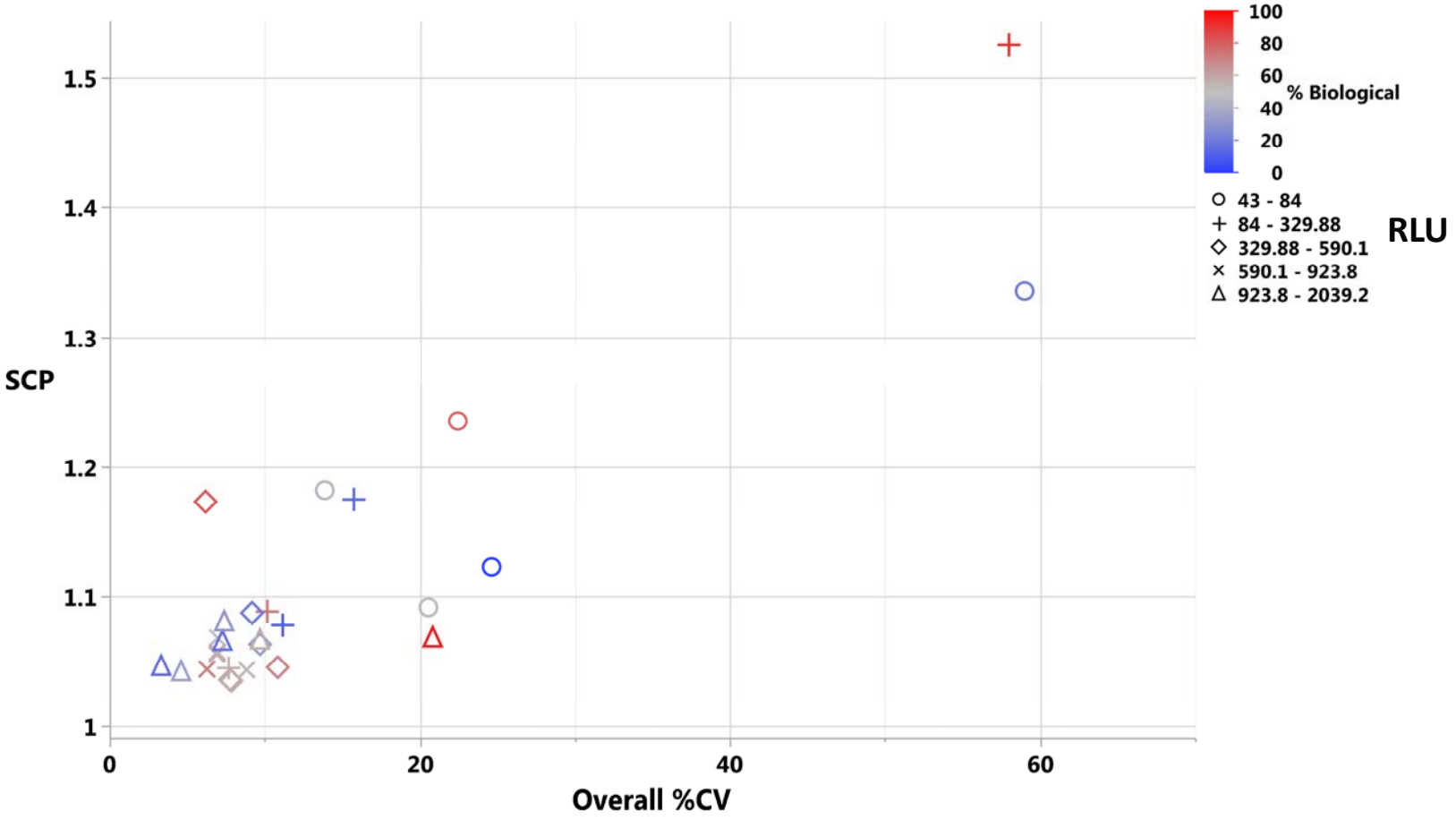
- Assays with high RLUs can also have low SCP factors.
- Low RLU (<100) does not always imply low SCP.

SCP factor vs. Total Variability (biological + analytical)



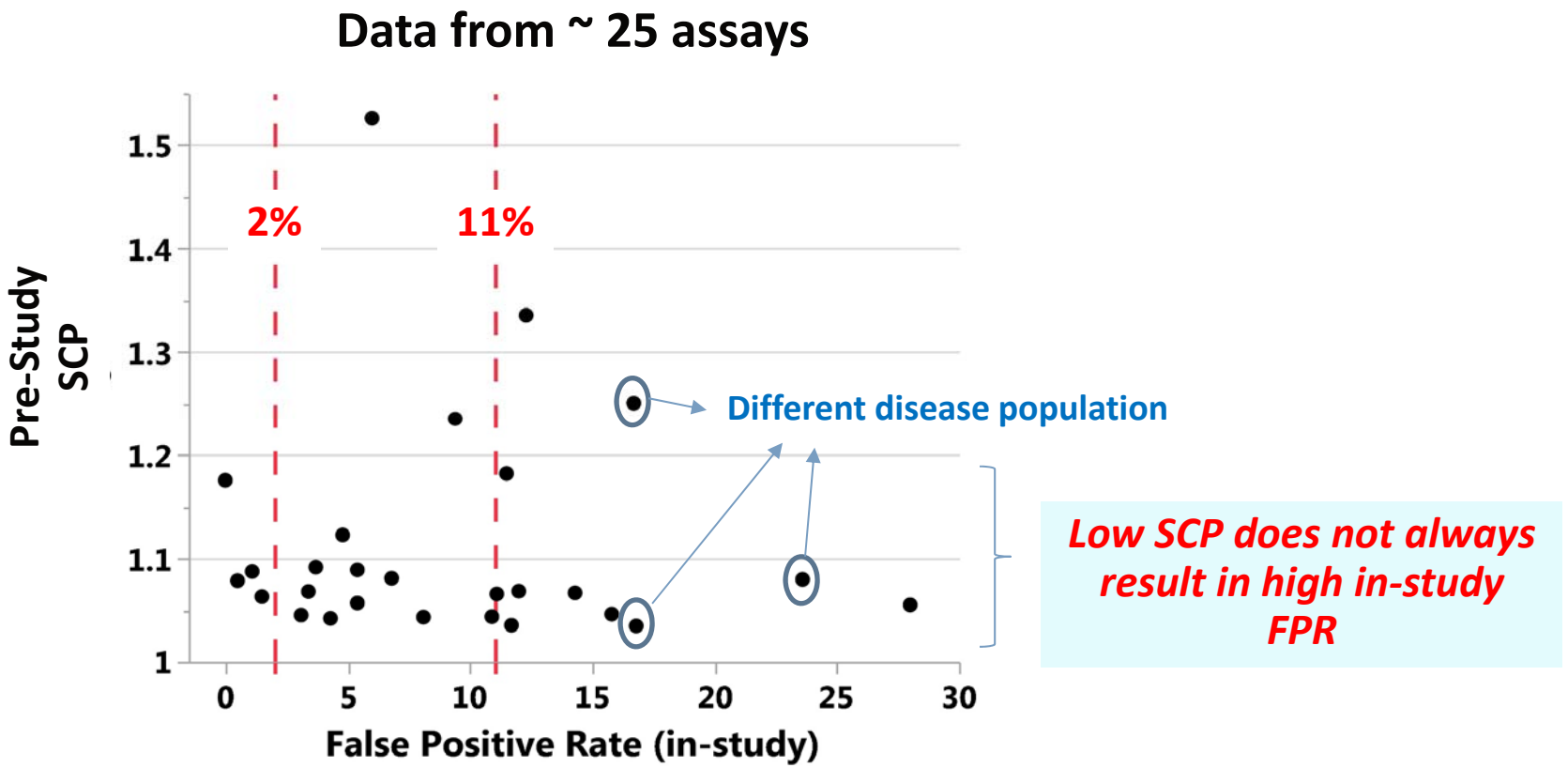
- **As expected, assays with higher total variability have higher SCP factor.**
- **Some departure from correlation is due to the differences between S and N**

SCP factor vs. Total Variability, by % biological & RLU



This shows that the SCP factor depends mostly on the total variability, regardless of the % biological variability and RLU level.

Low cut points & in-study FPR



Impact of low cut points on clinical relevance ??

Lower cut points may imply higher incidence of low-titer ADAs that are not clinical relevant, but….

*This will **not** dilute the clinical relevance of other ADA positive samples (i.e., those with higher titers, etc.).*

On the contrary, this may strengthen the evidence around clinical relevance of ADA results. Examples in next few slides.

Example 1

Impact of increased incidence of low-titer ADAs…

	Titer < 50	Titer > 50
CI no	8	10
CI yes	2	20
Total	10	30

p-value ~ 0.01

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

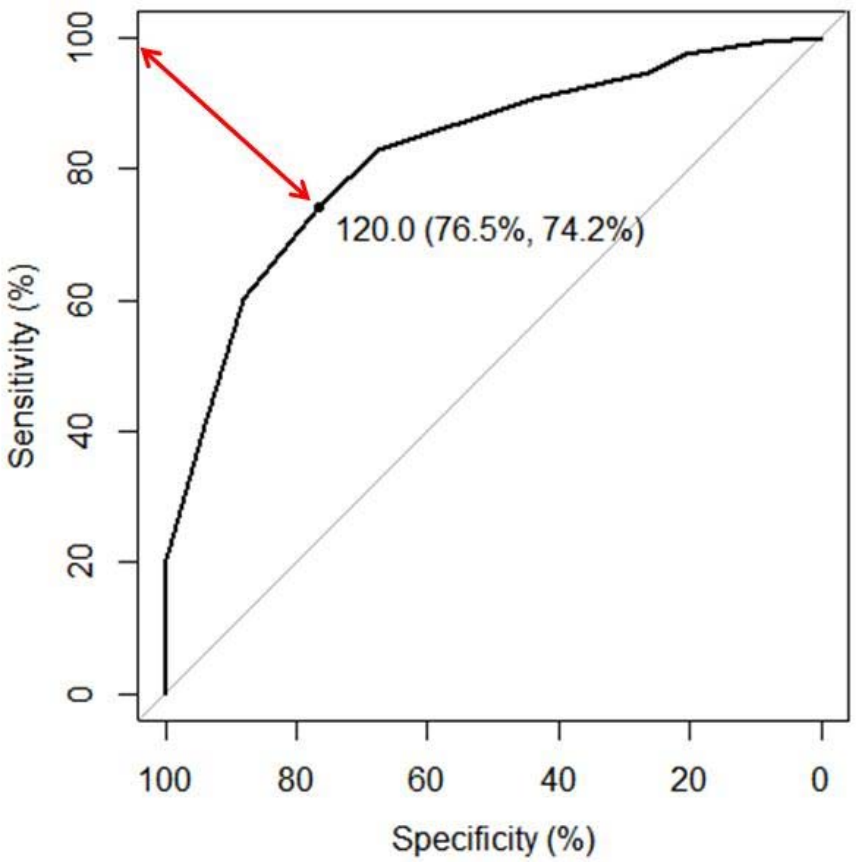
	Titer < 50	Titer > 50
CI no	80	10
CI yes	20	20
Total	100	30

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



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.

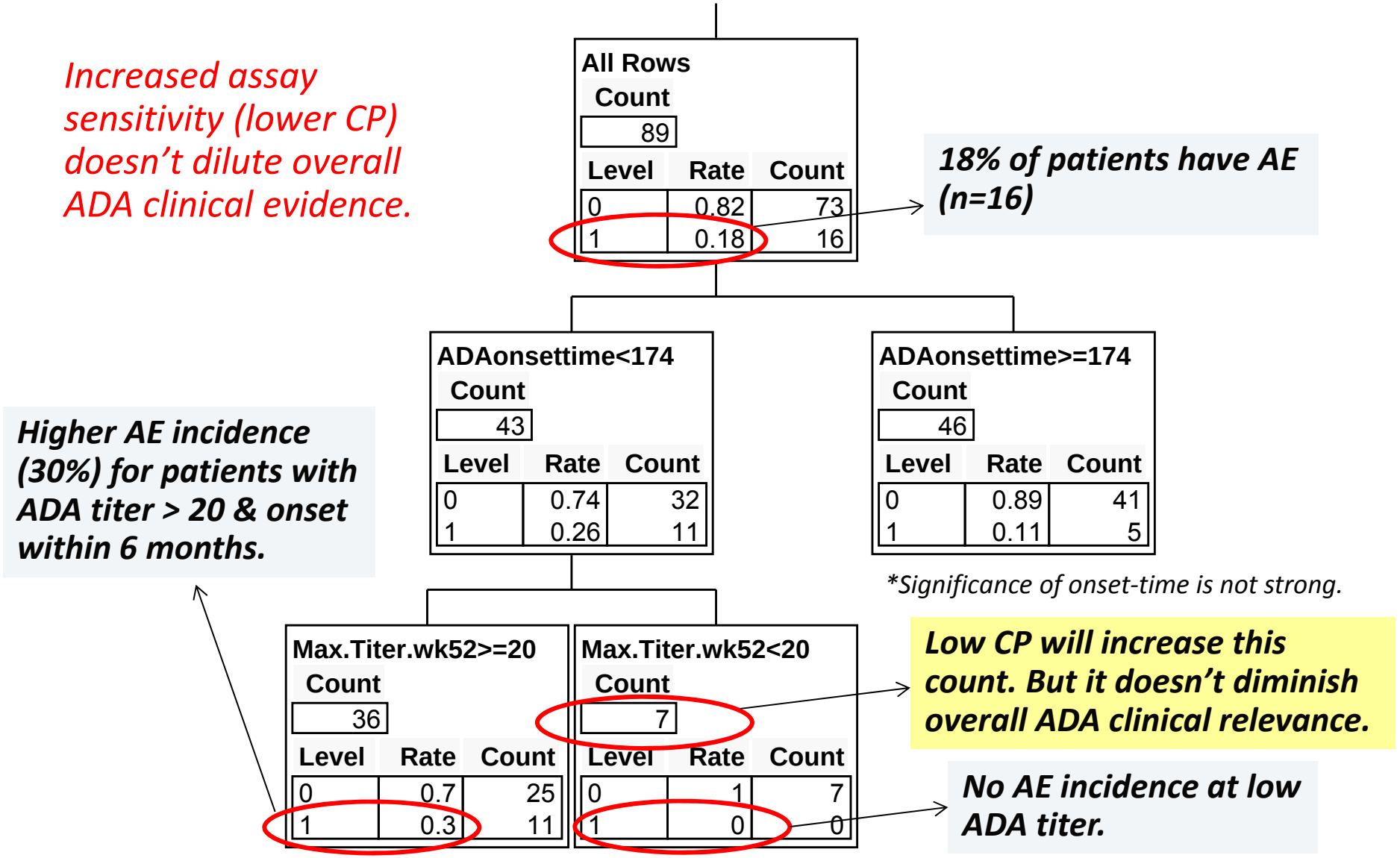
Increased assay sensitivity, with higher incidence of low-titer ADAs, will not dilute this clinical impact.

Low CP actually helps strengthen/refine the cumulative ADA clinical relevance.

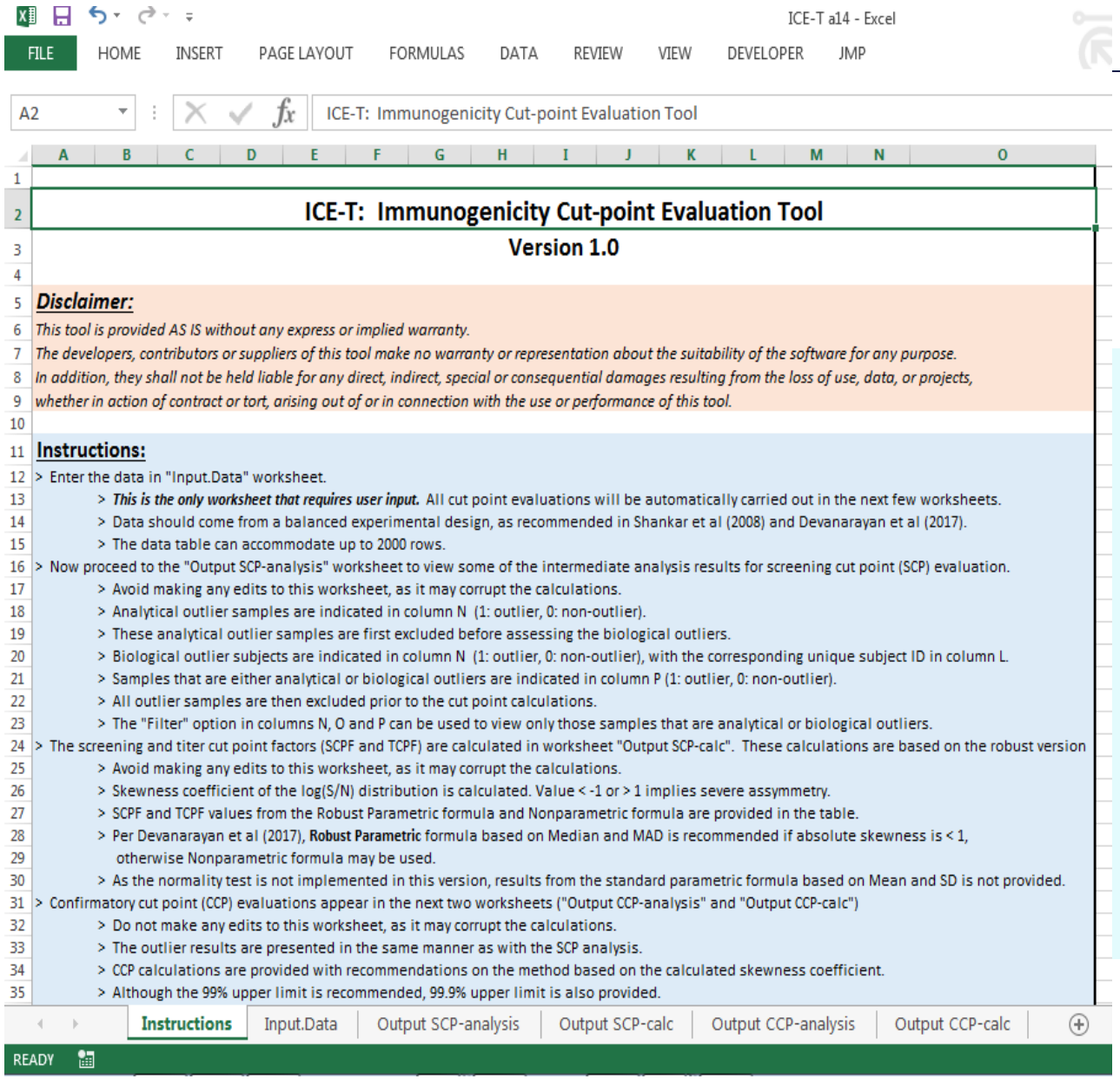
Combining the Titer results with ADA kinetics such as "Onset Time" and "Duration" (transient vs. degree of "persistence") via multivariate analysis may provide additional insights on clinical relevance (Shankar et al., 2014)

Example 3

Increased assay sensitivity (lower CP) doesn't dilute overall ADA clinical evidence.



Upcoming Excel-based tool for "preliminary/informal" CP calculations

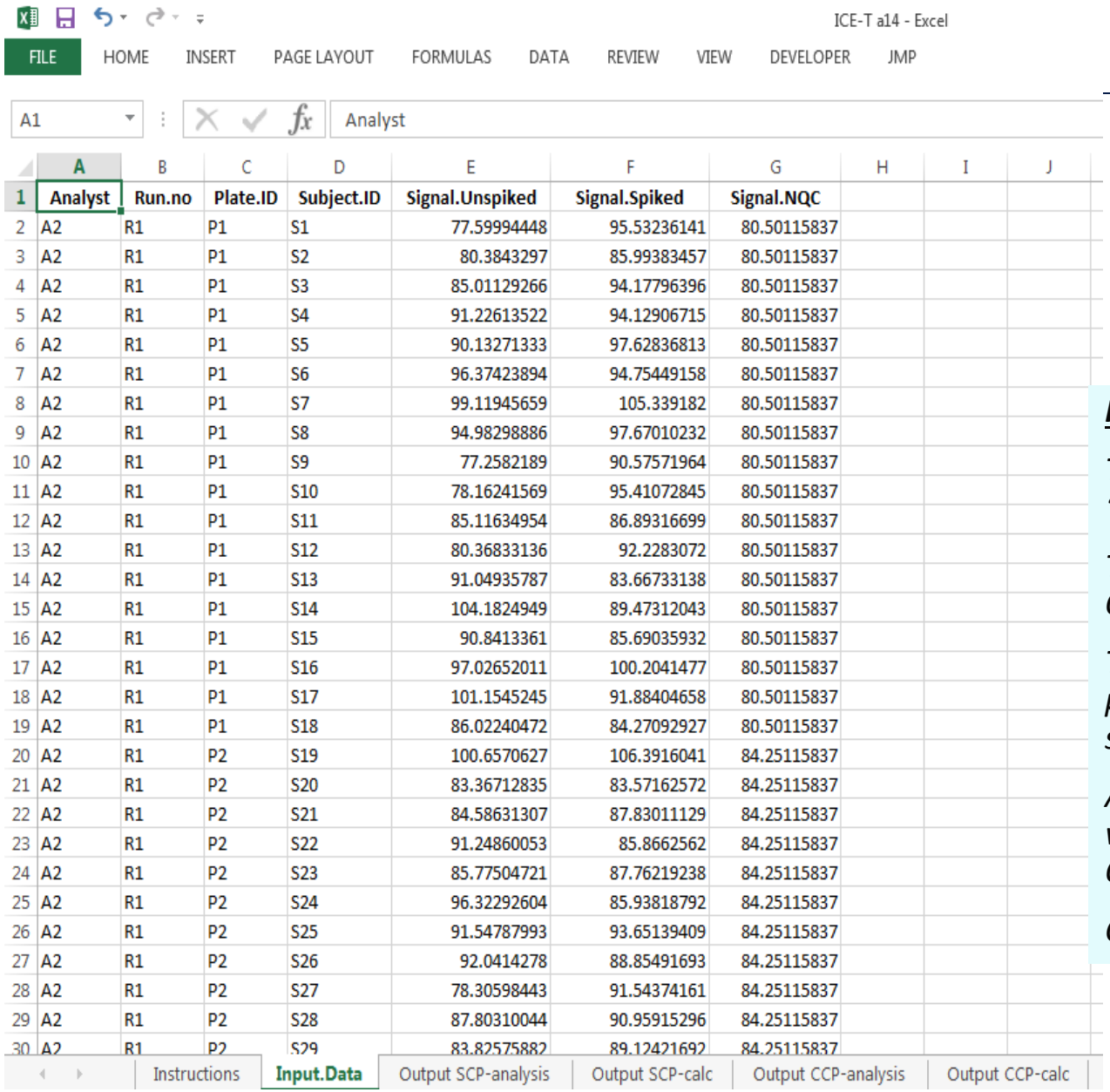


Instructions worksheet:

This includes standard disclaimers and cautionary statements that this tool is strictly for "exploratory" and "informal" use only.

It is not GxP validated, not CFR compliant, etc.

It implements a simpler option described in Devanarayan et al, 2017. The more rigorous method described in the paper for more formal GxP use is not possible in Excel, and requires SAS or other specialized programs.



	A	B	C	D	E	F	G	H	I	J
1	Analyst	Run.no	Plate.ID	Subject.ID	Signal.Unspiked	Signal.Spiked	Signal.NQC			
2	A2	R1	P1	S1	77.59994448	95.53236141	80.50115837			
3	A2	R1	P1	S2	80.3843297	85.99383457	80.50115837			
4	A2	R1	P1	S3	85.01129266	94.17796396	80.50115837			
5	A2	R1	P1	S4	91.22613522	94.12906715	80.50115837			
6	A2	R1	P1	S5	90.13271333	97.62836813	80.50115837			
7	A2	R1	P1	S6	96.37423894	94.75449158	80.50115837			
8	A2	R1	P1	S7	99.11945659	105.339182	80.50115837			
9	A2	R1	P1	S8	94.98298886	97.67010232	80.50115837			
10	A2	R1	P1	S9	77.2582189	90.57571964	80.50115837			
11	A2	R1	P1	S10	78.16241569	95.41072845	80.50115837			
12	A2	R1	P1	S11	85.11634954	86.89316699	80.50115837			
13	A2	R1	P1	S12	80.36833136	92.2283072	80.50115837			
14	A2	R1	P1	S13	91.04935787	83.66733138	80.50115837			
15	A2	R1	P1	S14	104.1824949	89.47312043	80.50115837			
16	A2	R1	P1	S15	90.8413361	85.69035932	80.50115837			
17	A2	R1	P1	S16	97.02652011	100.2041477	80.50115837			
18	A2	R1	P1	S17	101.1545245	91.88404658	80.50115837			
19	A2	R1	P1	S18	86.02240472	84.27092927	80.50115837			
20	A2	R1	P2	S19	100.6570627	106.3916041	84.25115837			
21	A2	R1	P2	S20	83.36712835	83.57162572	84.25115837			
22	A2	R1	P2	S21	84.58631307	87.83011129	84.25115837			
23	A2	R1	P2	S22	91.24860053	85.8662562	84.25115837			
24	A2	R1	P2	S23	85.77504721	87.76219238	84.25115837			
25	A2	R1	P2	S24	96.32292604	85.93818792	84.25115837			
26	A2	R1	P2	S25	91.54787993	93.65139409	84.25115837			
27	A2	R1	P2	S26	92.0414278	88.85491693	84.25115837			
28	A2	R1	P2	S27	78.30598443	91.54374161	84.25115837			
29	A2	R1	P2	S28	87.80310044	90.95915296	84.25115837			
30	A2	R1	P2	S29	83.82575882	89.12421692	84.25115837			

Instructions | **Input.Data** | Output SCP-analysis | Output SCP-calc | Output CCP-analysis | Output CCP-calc

Input.Data worksheet:

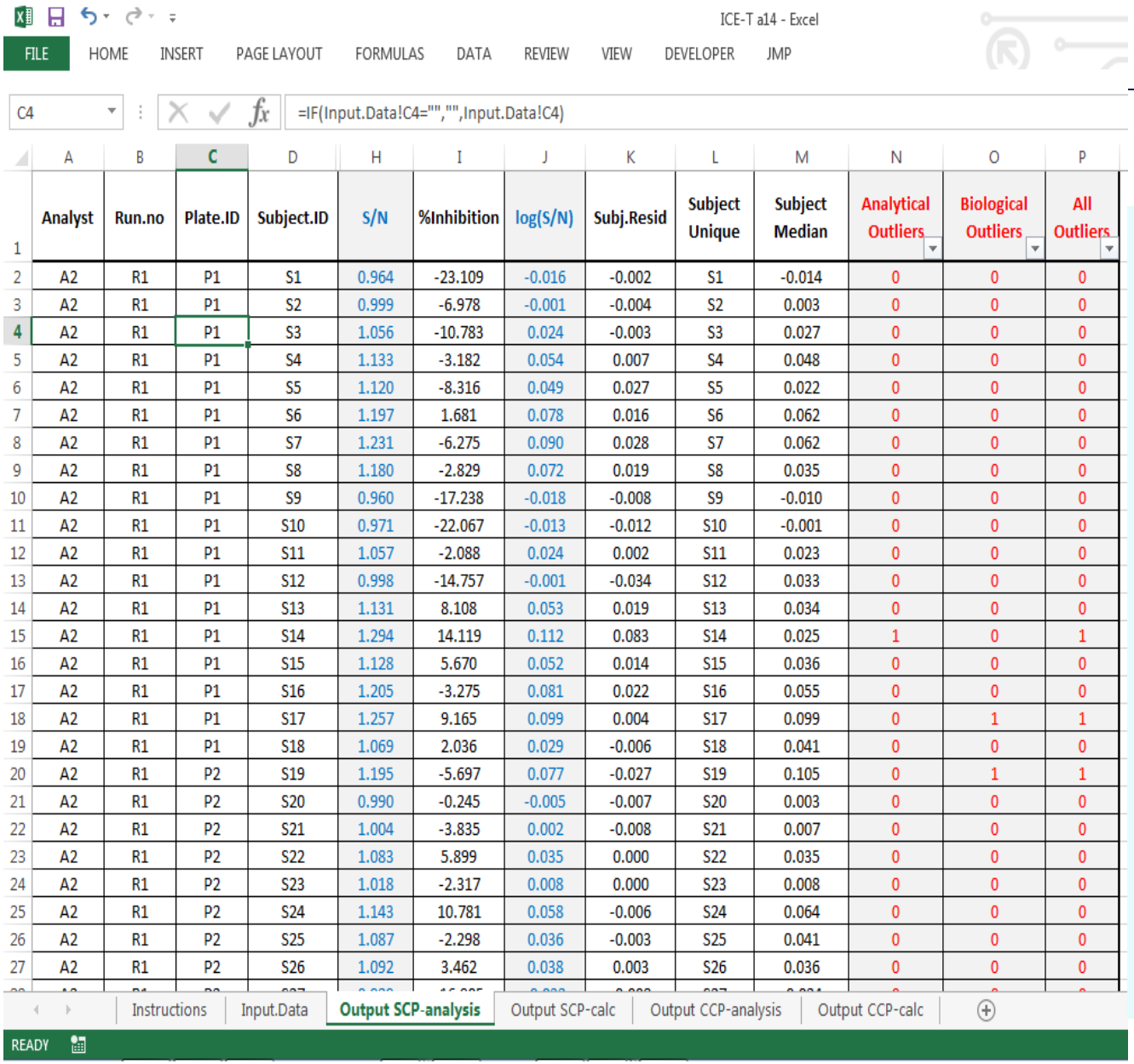
This is where you input the data in "long" format.

The other sheets are related to Output of SCP & CCP analyses.

Those sheets will be blank before pasting data in this Input.Data sheet.

After pasting the data, the results will automatically appear in the Output sheets.

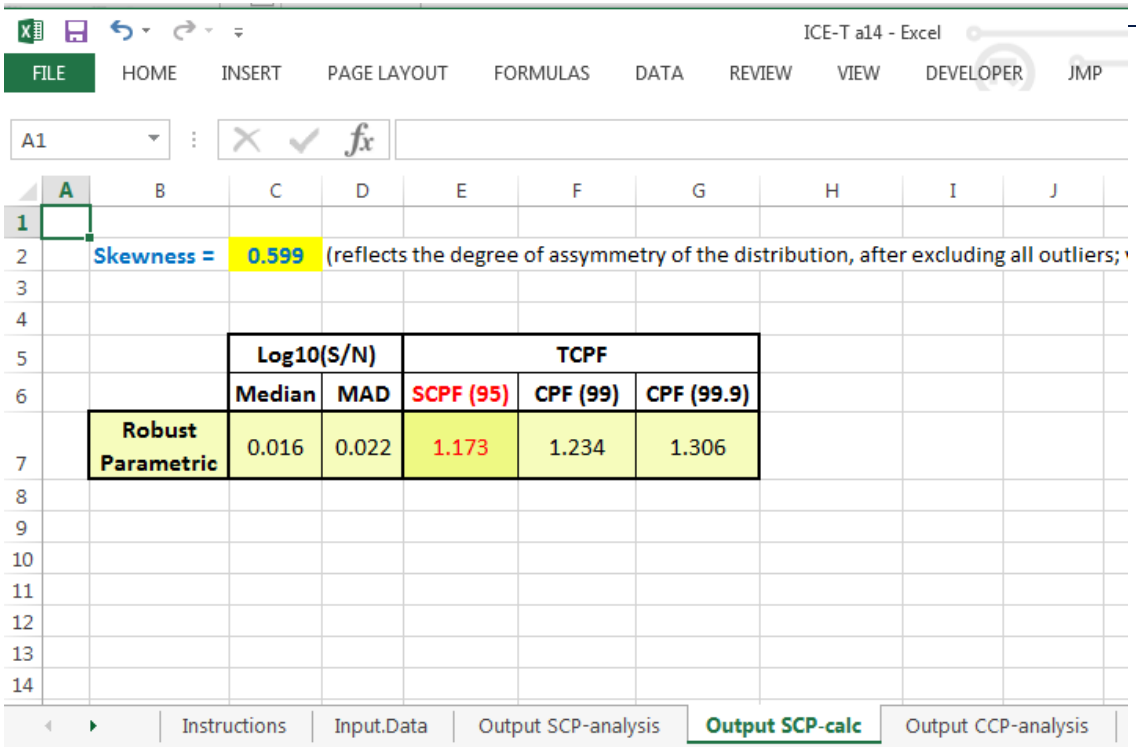
CP analysis in < 2 minutes! ☺



	A	B	C	D	H	I	J	K	L	M	N	O	P
1	Analyst	Run.no	Plate.ID	Subject.ID	S/N	%Inhibition	log(S/N)	Subj.Resid	Subject Unique	Subject Median	Analytical Outliers	Biological Outliers	All Outliers
2	A2	R1	P1	S1	0.964	-23.109	-0.016	-0.002	S1	-0.014	0	0	0
3	A2	R1	P1	S2	0.999	-6.978	-0.001	-0.004	S2	0.003	0	0	0
4	A2	R1	P1	S3	1.056	-10.783	0.024	-0.003	S3	0.027	0	0	0
5	A2	R1	P1	S4	1.133	-3.182	0.054	0.007	S4	0.048	0	0	0
6	A2	R1	P1	S5	1.120	-8.316	0.049	0.027	S5	0.022	0	0	0
7	A2	R1	P1	S6	1.197	1.681	0.078	0.016	S6	0.062	0	0	0
8	A2	R1	P1	S7	1.231	-6.275	0.090	0.028	S7	0.062	0	0	0
9	A2	R1	P1	S8	1.180	-2.829	0.072	0.019	S8	0.035	0	0	0
10	A2	R1	P1	S9	0.960	-17.238	-0.018	-0.008	S9	-0.010	0	0	0
11	A2	R1	P1	S10	0.971	-22.067	-0.013	-0.012	S10	-0.001	0	0	0
12	A2	R1	P1	S11	1.057	-2.088	0.024	0.002	S11	0.023	0	0	0
13	A2	R1	P1	S12	0.998	-14.757	-0.001	-0.034	S12	0.033	0	0	0
14	A2	R1	P1	S13	1.131	8.108	0.053	0.019	S13	0.034	0	0	0
15	A2	R1	P1	S14	1.294	14.119	0.112	0.083	S14	0.025	1	0	1
16	A2	R1	P1	S15	1.128	5.670	0.052	0.014	S15	0.036	0	0	0
17	A2	R1	P1	S16	1.205	-3.275	0.081	0.022	S16	0.055	0	0	0
18	A2	R1	P1	S17	1.257	9.165	0.099	0.004	S17	0.099	0	1	1
19	A2	R1	P1	S18	1.069	2.036	0.029	-0.006	S18	0.041	0	0	0
20	A2	R1	P2	S19	1.195	-5.697	0.077	-0.027	S19	0.105	0	1	1
21	A2	R1	P2	S20	0.990	-0.245	-0.005	-0.007	S20	0.003	0	0	0
22	A2	R1	P2	S21	1.004	-3.835	0.002	-0.008	S21	0.007	0	0	0
23	A2	R1	P2	S22	1.083	5.899	0.035	0.000	S22	0.035	0	0	0
24	A2	R1	P2	S23	1.018	-2.317	0.008	0.000	S23	0.008	0	0	0
25	A2	R1	P2	S24	1.143	10.781	0.058	-0.006	S24	0.064	0	0	0
26	A2	R1	P2	S25	1.087	-2.298	0.036	-0.003	S25	0.041	0	0	0
27	A2	R1	P2	S26	1.092	3.462	0.038	0.003	S26	0.036	0	0	0

Output SCP-analysis worksheet:

- *This output sheet includes the evaluation of outliers.*
- *Analytical and Biological are listed (0/1 for no/yes).*
- *Excel Filter option can be used to evaluate the outliers closely.*
- *The analytical outliers are identified and excluded first, then the biological outliers are identified and excluded.*



"Output SCP-calc" worksheet:

- *SCP and TCP factor results appear here.*
- *TCPF can be set at 95th, 99th or 99.9th, depending on the measurability and precision of Titers.*
- *Similar output results appear for CCP analysis in the next two worksheets.*
- *A separate worksheet will include various graphs as well.*

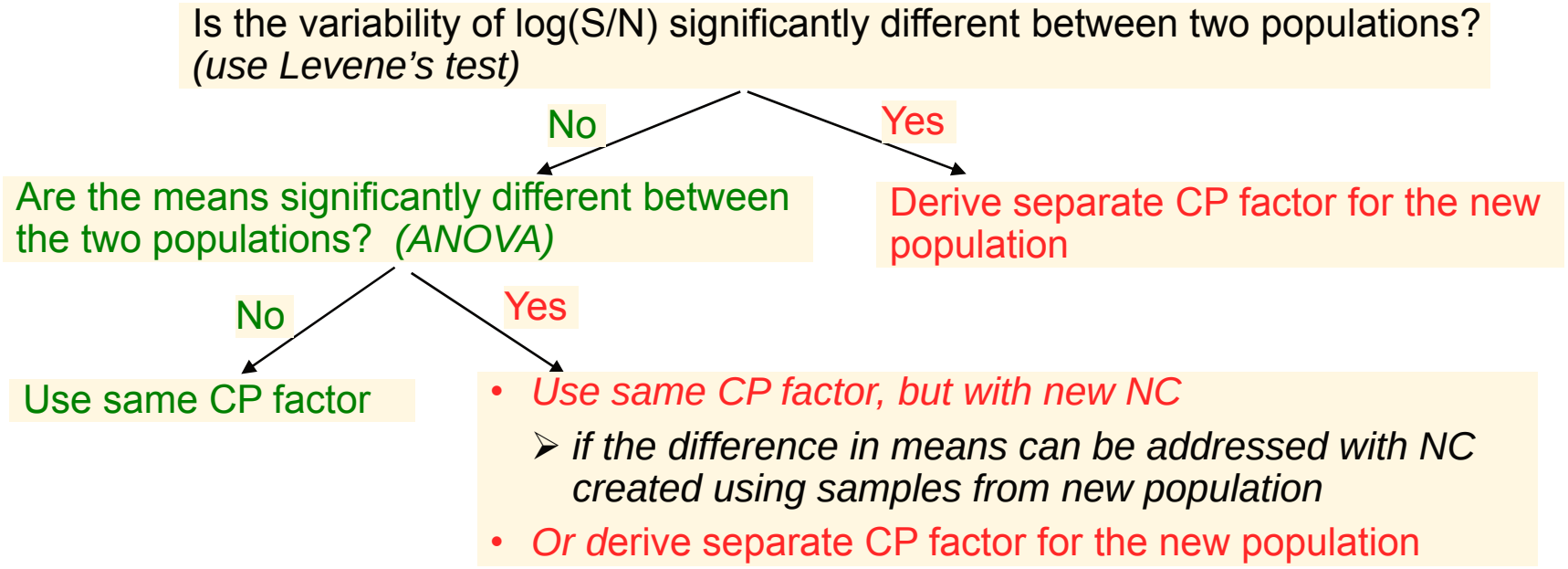
Release Plans:

- *By April 2019, hopefully.*
- *I will be placing it in the **NIH Assay Guidance Manual** (NIH-AGM).*
  - *Free open access publication.*
  - *https://www.ncbi.nlm.nih.gov/books/NBK53196/*

Extensions to other populations

Extensions of CP factor to other populations

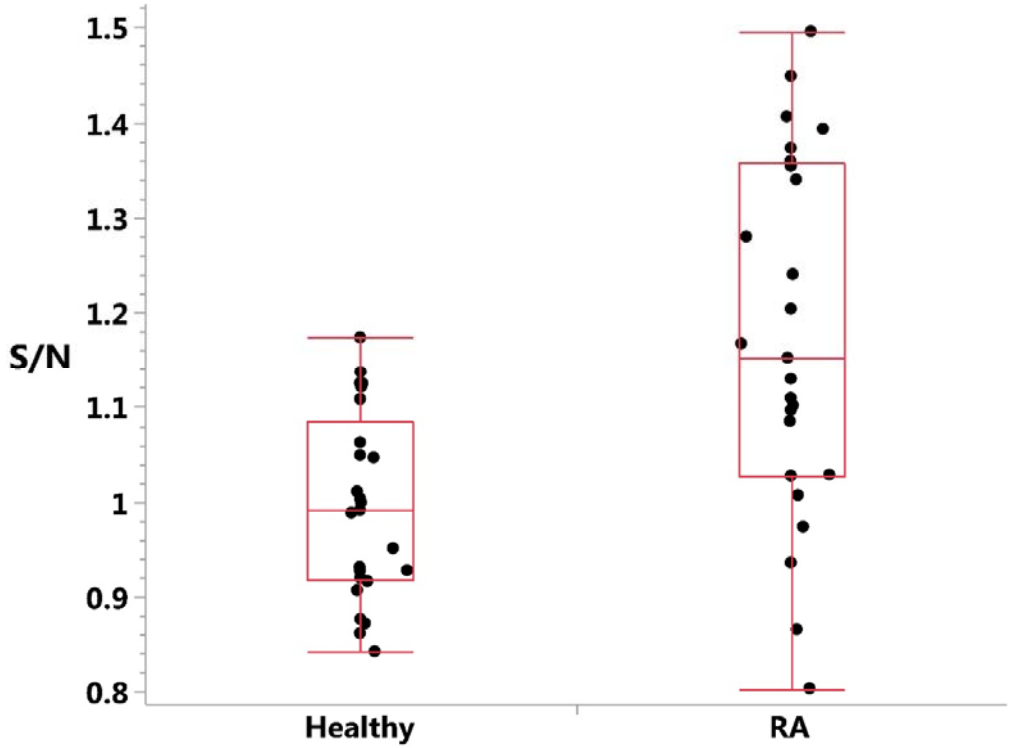
- *For extending the CP to other populations (different disease, demographic, etc.), this decision tree may be useful.*
- *Need > 20 drug naïve individual sera from each population; assess outliers, distributions, etc.*



Devanarayan et al., 2017

Extensions of CP factor to other populations

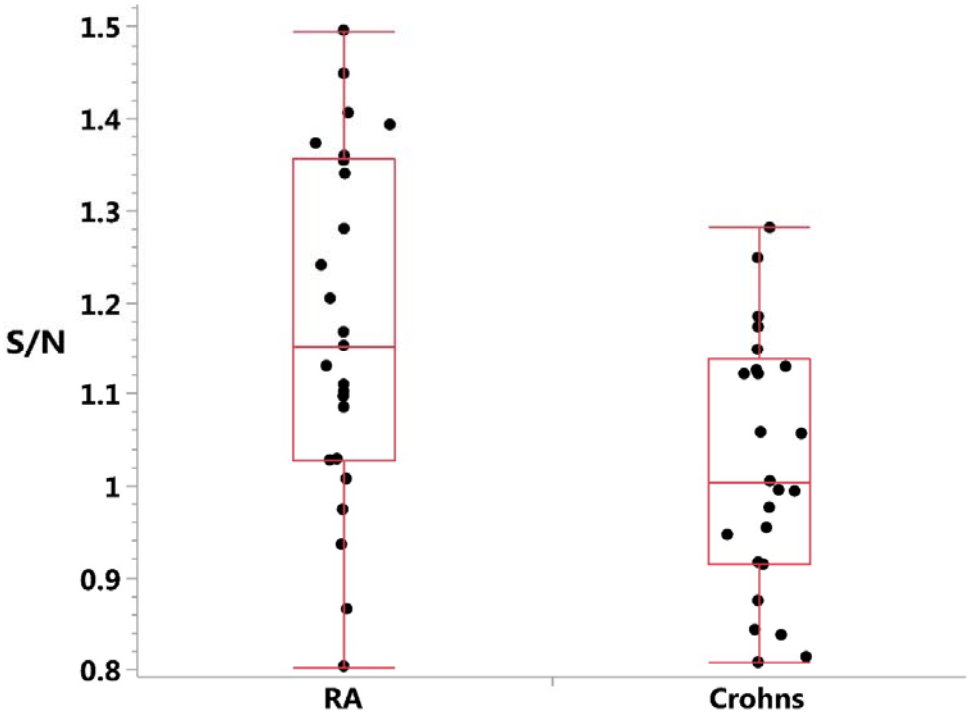
Case study 1



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

Extensions of CP factor to other populations

Case study 2



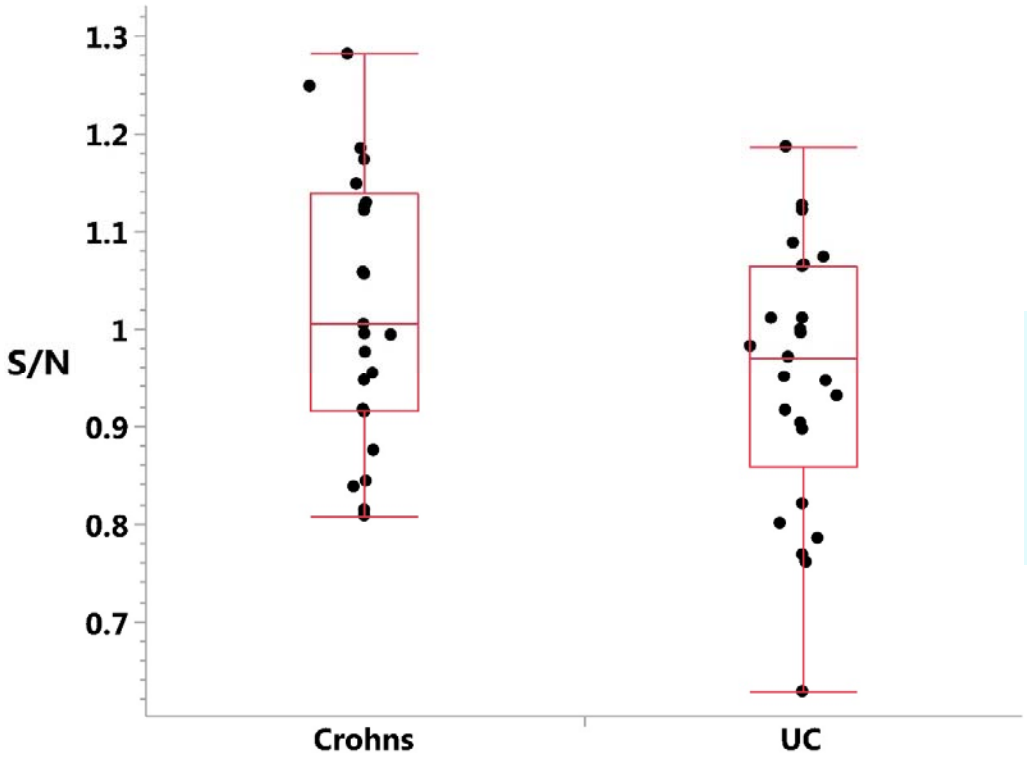
Variances are not significantly different

But means are significantly different (p=0.016)

- *If this difference can be accounted for 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*
- *Same CP factor can be used for UC population.*

In-study cut points

In-study cut points

Based on simulations, with the sampling design of Shankar et al (2008):

- FPR of SCP factor can vary between 2 to 11% (Devanarayan et al., 2017)
  - After excluding samples with pre-existing Ab.
- FPR < 2% or > 11% can trigger the need for in-study cut point.

This may occur if:

- 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

Visual and statistical assessment may provide some insights.

- Box plots, comparison of means and variances (ANOVA, w/o outliers), etc.

In-study cut points

Design / data requirements & analysis

Pre-dose samples from > *50 subjects* (phase-II), > 100 subjects (phase-III)

- One reportable result per subject will suffice.
- Samples should be tested across > 3 plates/runs and >=2 analysts.
- Stratification with respect to key baseline characteristics (esp. phase-III).
- *Overall variability from these data will reflect all the relevant sources of variability (analyst, inter-run, intra-run, etc.). Therefore, it is not necessary to test the samples multiple times via a balanced design format.*

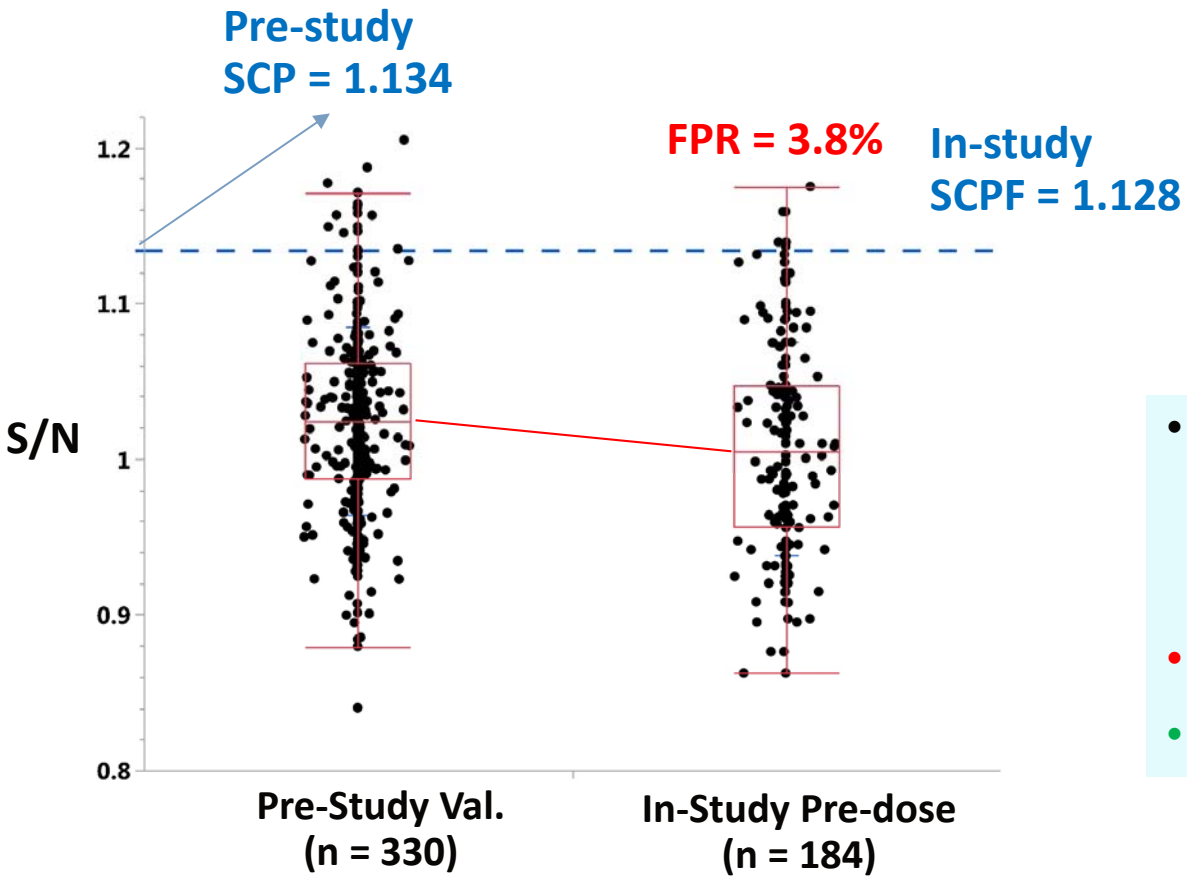
For rare disease or pediatric studies where samples are limited:

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

Analysis: *Similar as pre-study CP analysis (assess distribution, exclude outliers, etc.)*

In-study cut points

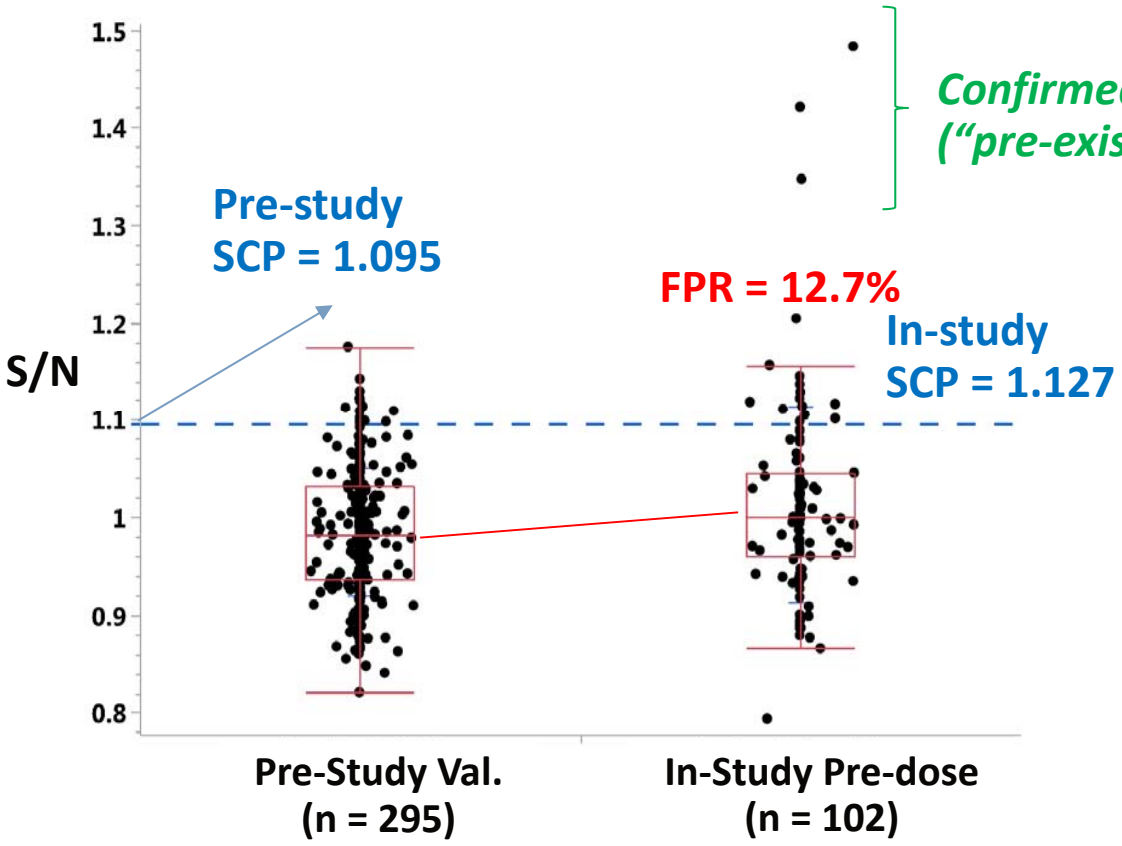
Case Study 1



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

In-study cut points

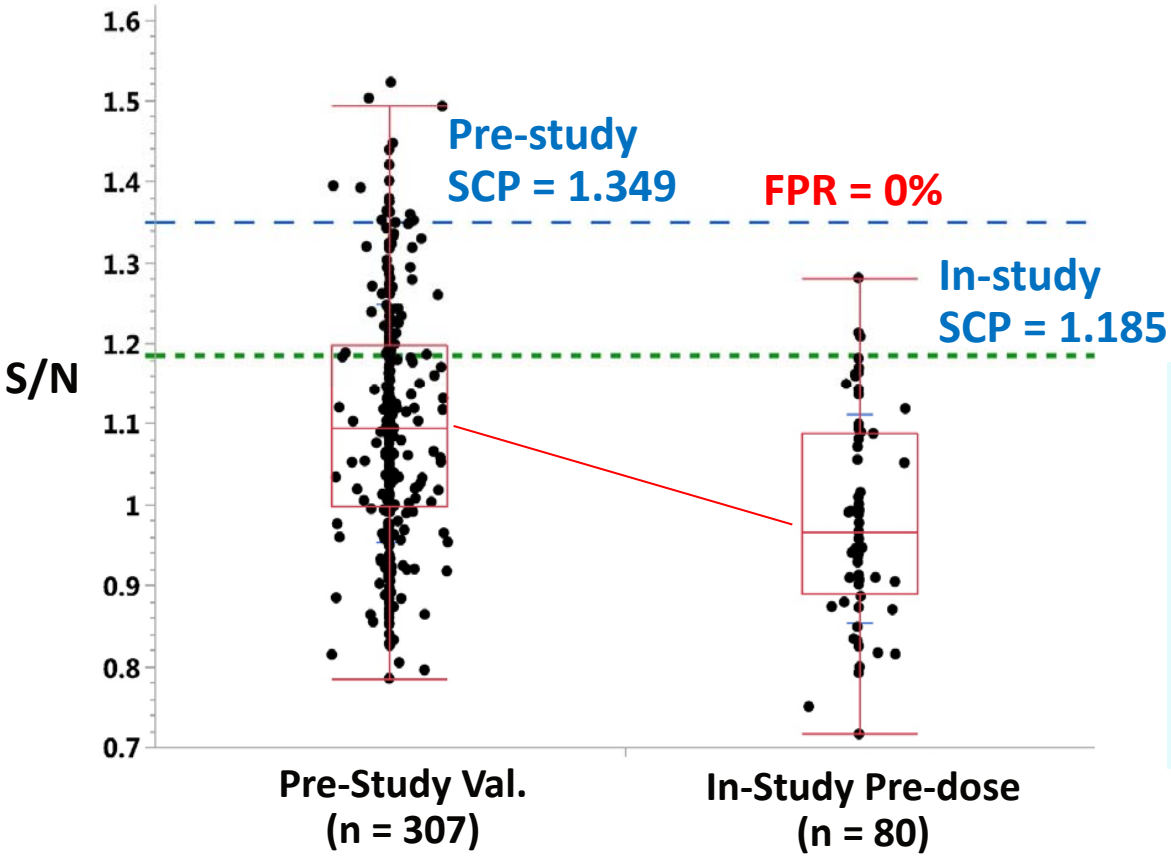
Case Study 2



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

In-study cut points

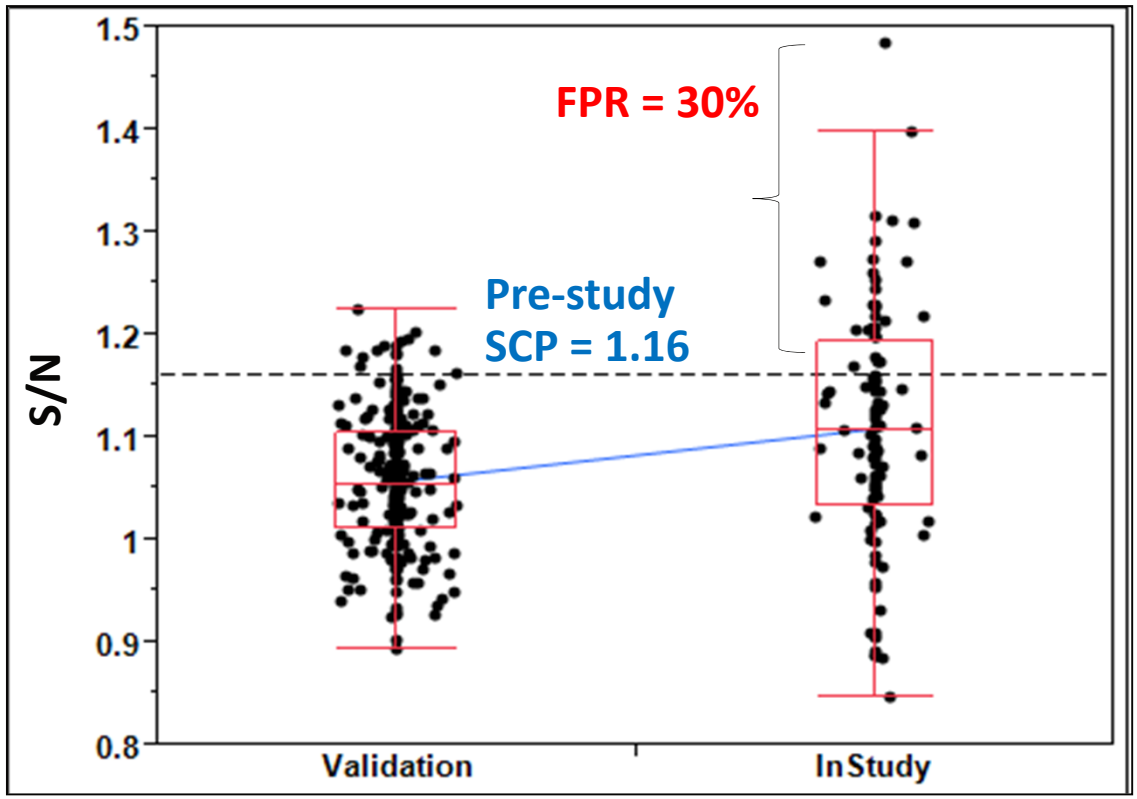
Case Study 3



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

In-study cut points

Case Study 4



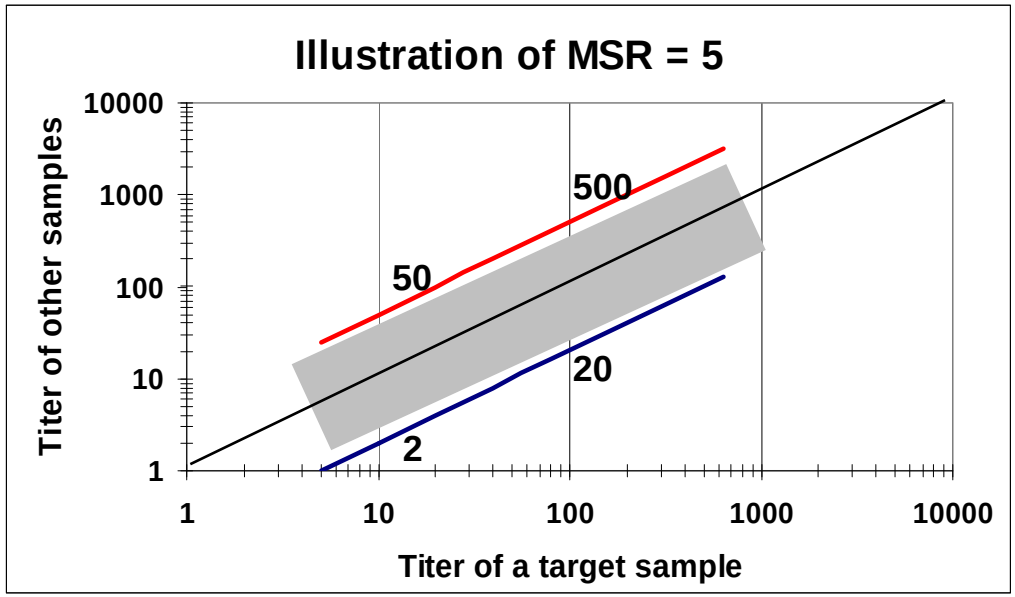
- *Variances are significantly different*
- **FPR = 30%**, after excluding pre-existing Ab
- ***Need in-study CP***

Titer Precision & Titer CP

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 for most assays from our experience, and is considered desirable.

Evaluation of Titer MSR

Use the data from 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 *SD of log(titer)* results, for use in the formula below.

$$\text{MSR} = 10^{t_{0.05,n\_1} * \sqrt{2} * SD}$$

- 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^) of the difference of log(titers) = Ratio of Titers.
- Hence this is the Minimum Significant Ratio 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 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.

Criteria for Treatment-boosted ADA

Criteria for identifying treatment-boosted ADA

For subjects with pre-existing ADA, i.e., with confirmed positive ADA at baseline, a criteria is needed to determine whether post-dose ADA positive results are treatment-boosted.

- *See Shankar et al (2014) for treatment-induced vs. treatment-boosted, and other important terminologies.*

As in early 2000s, when subjective criteria (e.g., 50% inhibition) was used to confirm ADA positive samples, subjective criteria are still common for declaring treatment-boosted ADAs.

- e.g., 2-fold difference in pre vs. post-dose signal from screening assay (the "2-fold" criteria is totally subjective).

Such subjective gut-driven criteria do not consider variability, and fail to control false positive/negative error rates.

Criteria for identifying treatment-boosted ADA

1. Dilution-dependent criteria (*adapted from clinical serology*)
   - If the titers are determined via 2-fold serial dilutions, 4-fold difference between pre-dose vs. post-dose titers is suggested as a criteria for treatment-boosted ADA. For 3-fold serial dilution, 9-fold difference is suggested, etc. This is apparently common in other applications.
   - ***Ignores assay & biological variability*** *& doesn't control error rates. May lead to under/over-reporting 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 criteria for defining treatment-boosted ADA *(Ref: USP chapter <1106.1>).*
   - Procedure for evaluation is explained earlier in this presentation.

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 data from in-study phase 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

Similar to the screening 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 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 Titers = 2.5
- Then the *Titer of HPC in every in-study run (plate) should be within 400 to 2500* (1000 / 2.5 = 400, 1000 X 2.5 = 2500)
- Alternatively, one dilution fold criteria can be applied, but note that it is very subjective and doesn't take into account of the Titer variability.

Sensitivity & Low QC

Sensitivity & Low QC assessment

Sensitivity is currently defined separately for Screening & Confirmatory assays.

- Sometimes simply interpolated from the SCP & CCP values. Implies only 50% confidence.
- 2016 FDA draft guidance calls for "consistently" observing positive result.
  - Implies > 50% confidence; perhaps 80%, 90%, 95%, 99%, ….

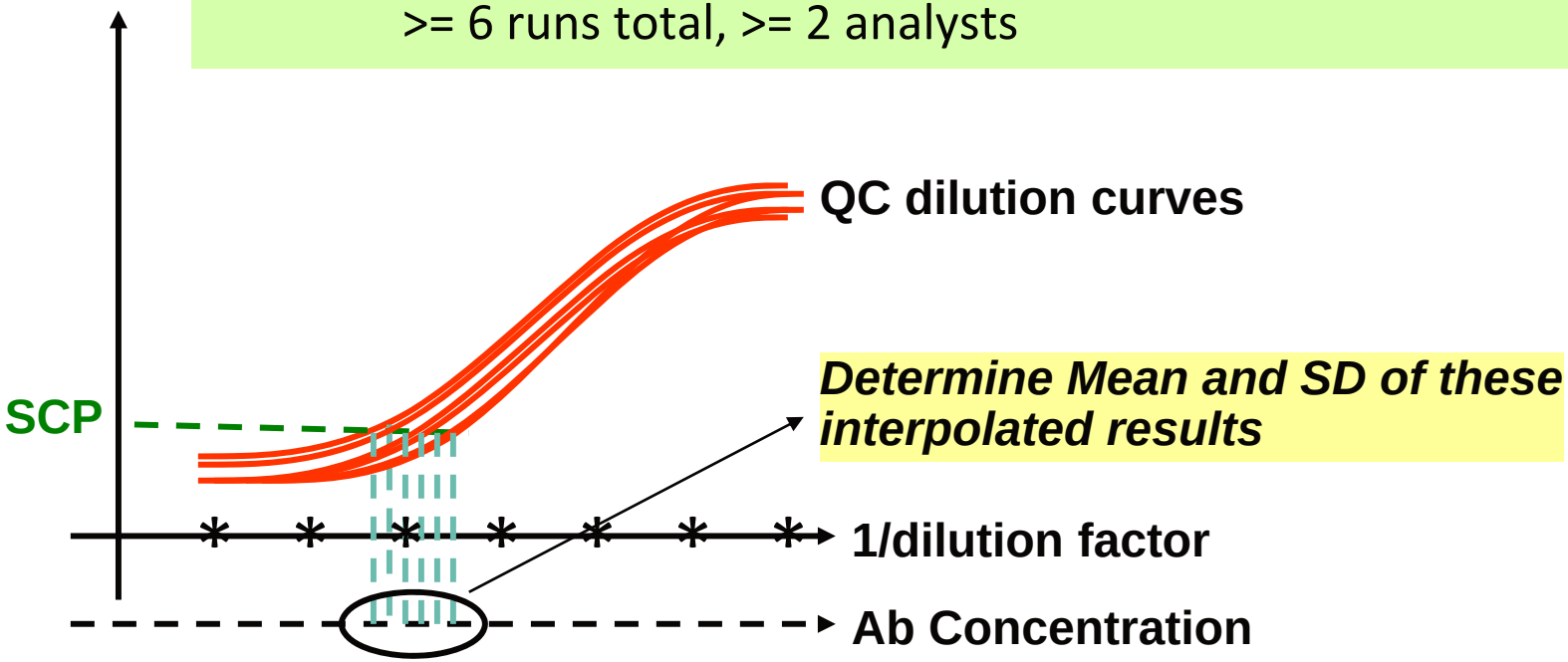
LQC is set at 1% failure rate in screening assays.

- This level may be inadequate for confirmatory assays.
- May pose challenges in Selectivity assessments.
- Drug-spiked LQC may not serve its purpose during in-study phase.

Proposal: Define Sensitivity as lowest ADA level that is consistently (>99%) detected (screening assay) **and** confirmed (specificity assay). This should also serve the purpose of LQC.

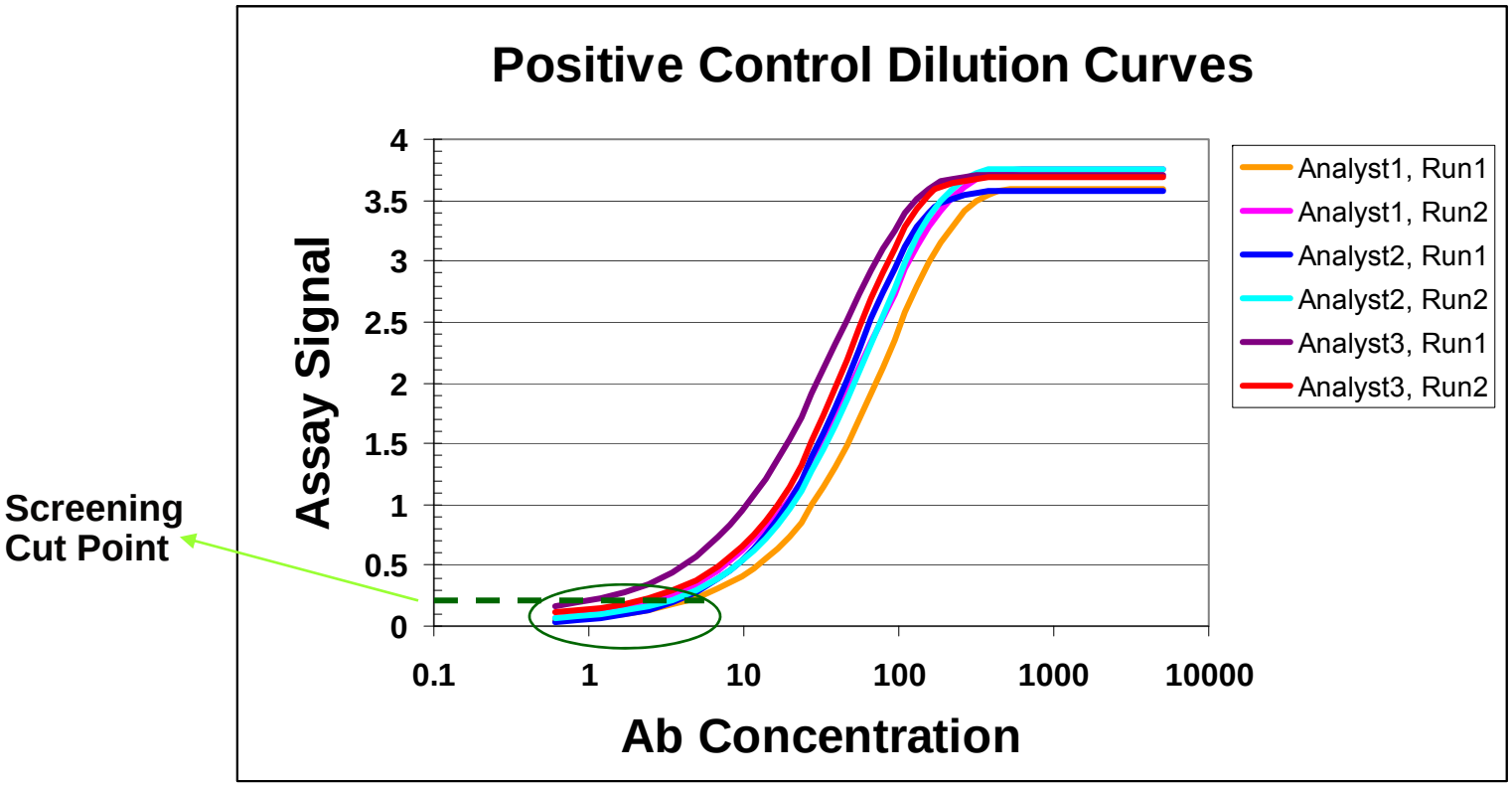
Sensitivity & Low QC assessment (contd.)

Data: >= 5 two-fold serial dilutions of **positive QC pool**
>= 6 runs total, >= 2 analysts



- Evaluate 99th percentile of these interpolated levels = **Mean + $t_{0.01,df}$ SD**
- If confirmed (drug spiked %inhibition > CCP), then this is Sensitivity & LQC.
- Otherwise, use the next dilution level that confirms as the Sensitivity & LQC
  - < 1% failure rate. Perhaps up to 0.1% is good enough?

Sensitivity & Low QC level assessment – illustration



Positive Control dilution curves from 3 analysts, 2 runs each.

Curves were fit and interpolation was done using a 5-parameter logistic model, with weighting. Simple linear interpolation between dilutions that flank the SCP might also be adequate.

Sensitivity & Low QC level assessment – contd.

	Analyst 1	Analyst 2
Run 1	2.232	2.635
Run 2	2.911	2.115
Run 3	1.757	3.855
Mean	2.58	
SD	0.74	
95% upper limit	**4.08**	
99% upper limit	**5.08**	

→ Sensitivity & Low QC ?

If the drug-spiked % inhibition at 5.08 ng/mL is > CCP, this can be reported as the Sensitivity & Low QC level.

If not, assess the drug specificity at the next dilution level, and use that level for Sensitivity & Low QC.

Thank you for your interest & attention!

Questions ?