Influence of biological variability, assay signal, and outlier criteria on Immunogenicity cut points and clinical relevance

V. Devanarayan, Ph.D., FAAPS

Charles River Laboratories

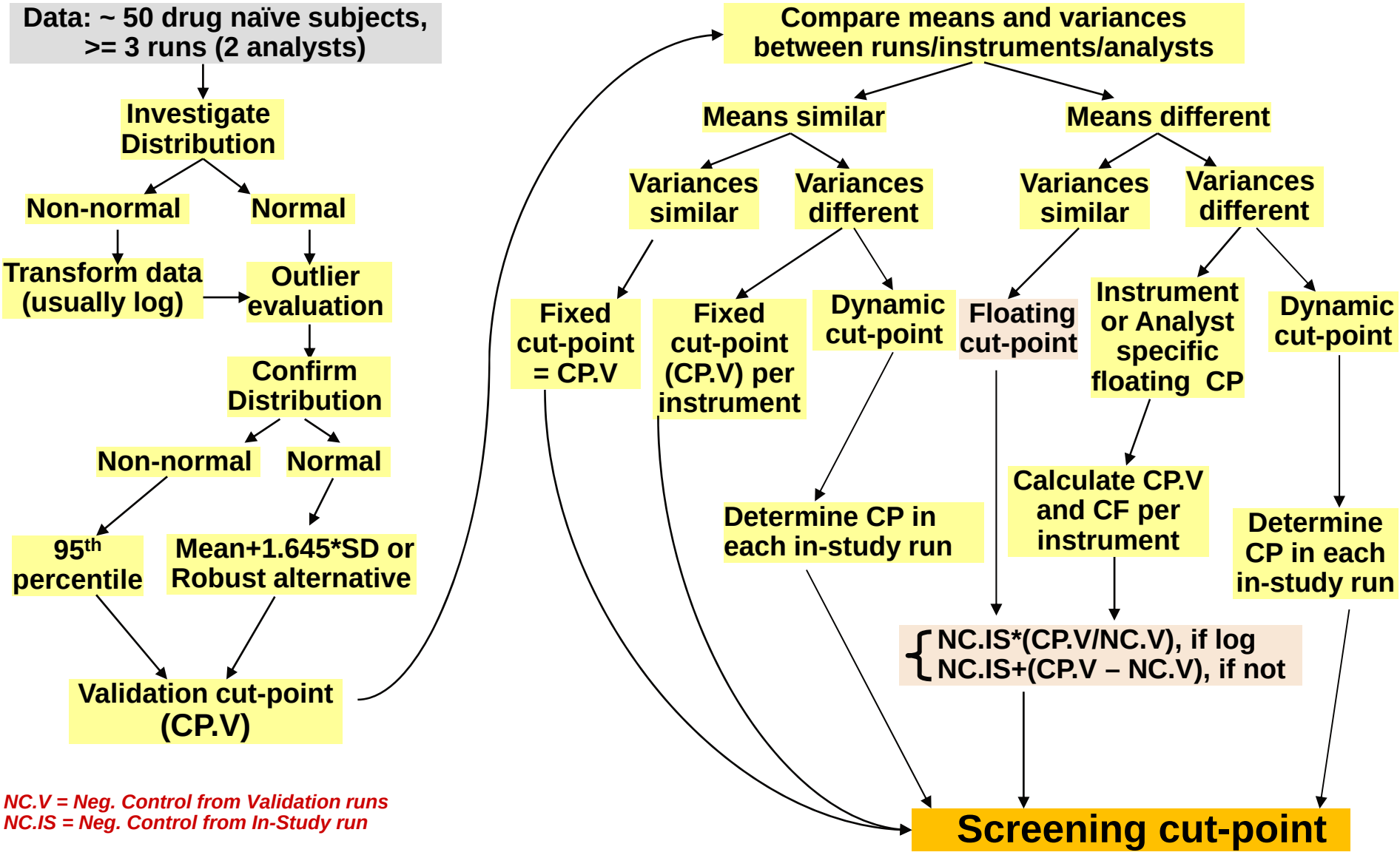
European Bioanalytical Forum - Focus Workshop on Immunogenicity

Lisbon, Portugal, September 19-20, 2018

charles river

Shankar et al., 2008

Screening CP Evaluation *(2008 white paper)*

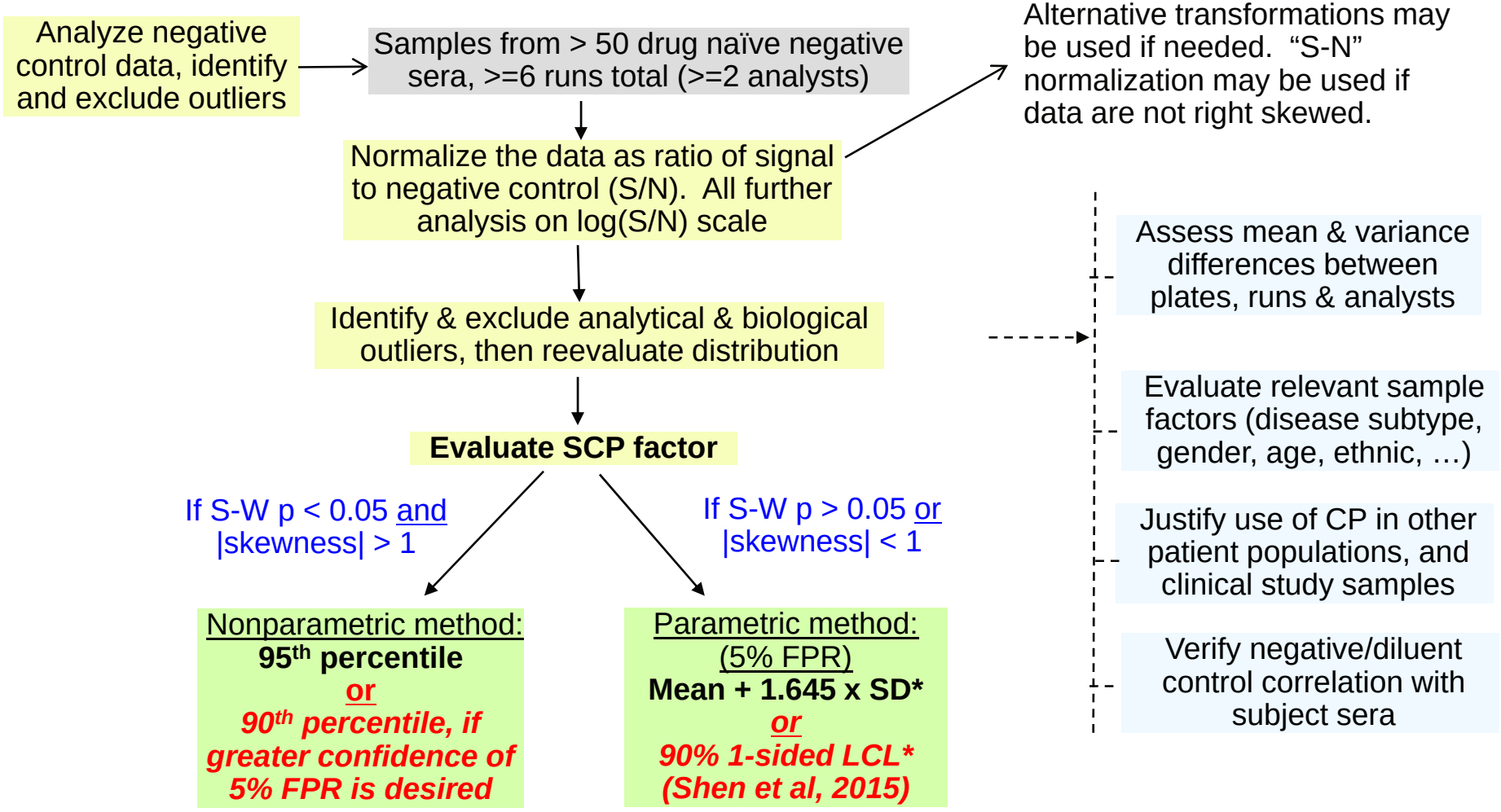


NC.V = Neg. Control from Validation runs
NC.IS = Neg. Control from In-Study run

Devanarayan et al., 2017

Screening CP Evaluation

Simplified flow-scheme that should work in most cases

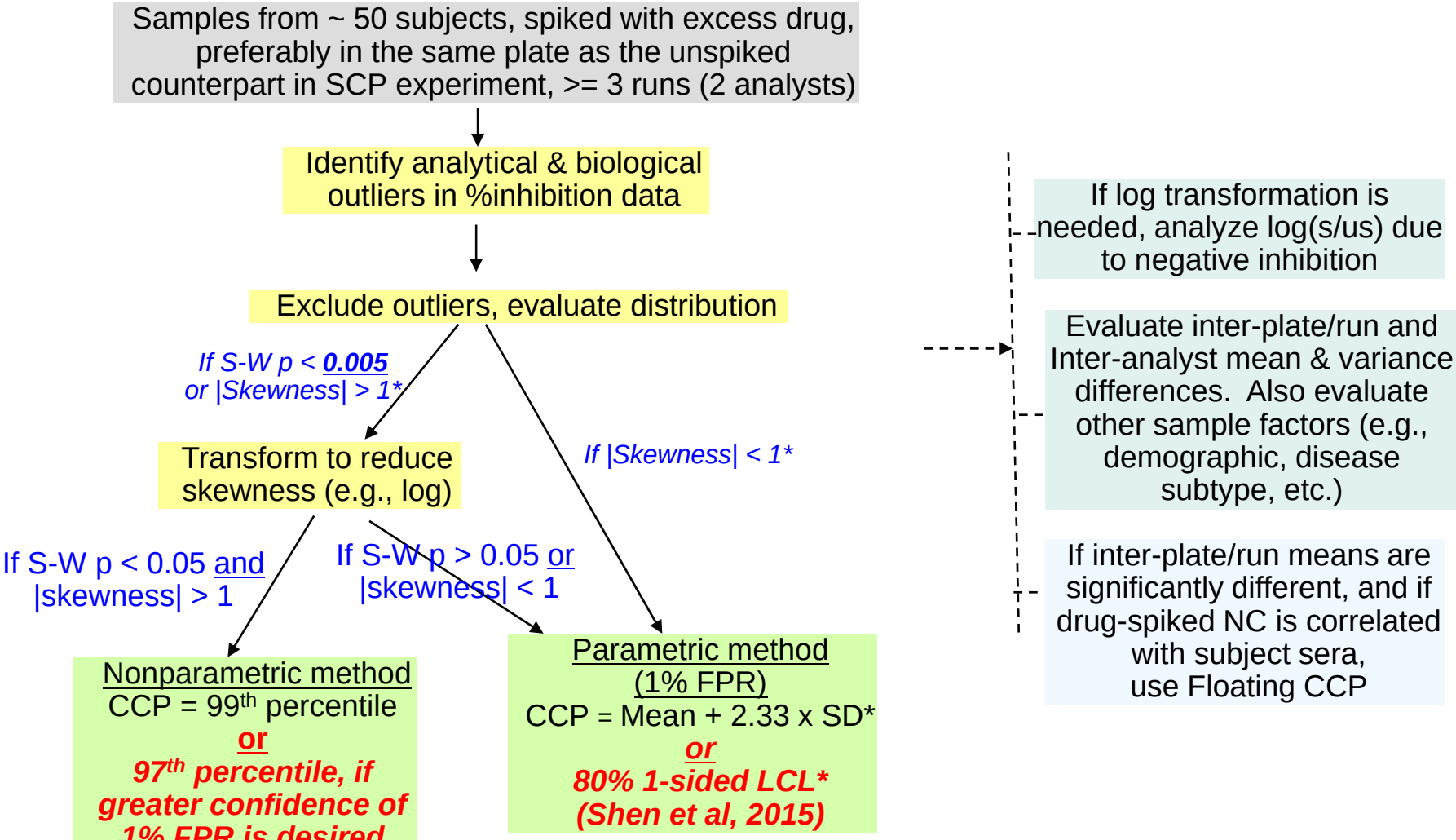


** Use Median instead of Mean, and 1.4826xMAD instead of SD to ensure robustness when there are borderline outliers.*

Devanarayan et al., 2017

Confirmatory Cut Point 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.*

Summary of some changes in the SCP recommendations

Characteristic	Shankar et al, 2008	Devanarayan et al., 2017
Type of Cut Points (CPs)	Fixed, floating and dynamic CPs were suggested for specific situations.	Floating CP is recommended for all scenarios.
Data Transformation	None specifically recommended	Log S/N ratio is offered as default transformation for routine application
Sources of variation	Only the intra-assay and inter-subject variability are used in the CP evaluation.	All relevant sources of variation are considered for CP calculation (inter-assay, intra-assay, inter-analyst, inter-subject, etc.).
Outlier identification	Outliers are not differentiated as being analytical versus biological.	A systematic analysis process is suggested to yield differentiation for analytical and biological outliers.
Outlier removal	Outliers are evaluated from a single iteration of ANOVA, without consideration of analytical versus biological outliers.	Analytical outliers are first identified and removed, and then the model is refit to the remaining data to remove biological outliers.
CP determination	Analysis is done directly on the assay signal (S), instead of S/N. Normalization with NC for floating CP is done only in the end using pooled data from each run.	Analysis is done directly on Log(S/N) from the corresponding plate. Fixed CP on S/N ratio equates to a floating CP on the signal.
CP normalization	Floating CP normalization using NC is computed at the run-level.	Floating CP normalization is computed at the plate-level to adjust for plate-to-plate variability.
Parametric vs. Nonparametric CPs	Nonparametric percentile method was recommended when the S-W p-value is < 0.05.	Nonparametric method is recommended only when both the S-W test p-value is < 0.05 and the absolute value of skewness coefficient is > 1.
Application of in-study CPs	No recommendations on when and how to evaluate in-study CPs.	Recommendations are provided

"Low" Cut Points

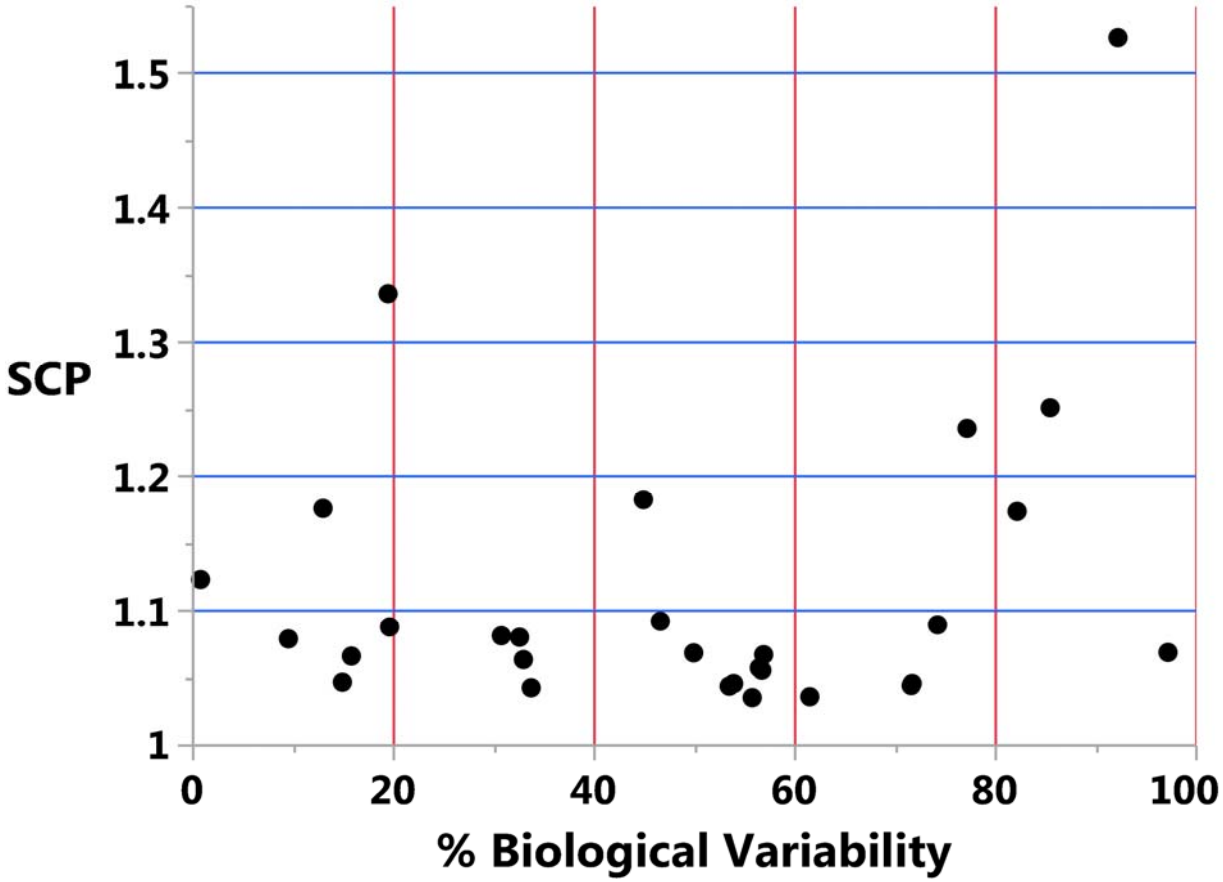
SCPF can be quite low, sometimes < 1.1 and often < 1.2

Frequently asked questions/concerns:

- Is this due to *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?
- Will this *dilute the overall clinical relevance* of the ADA results?

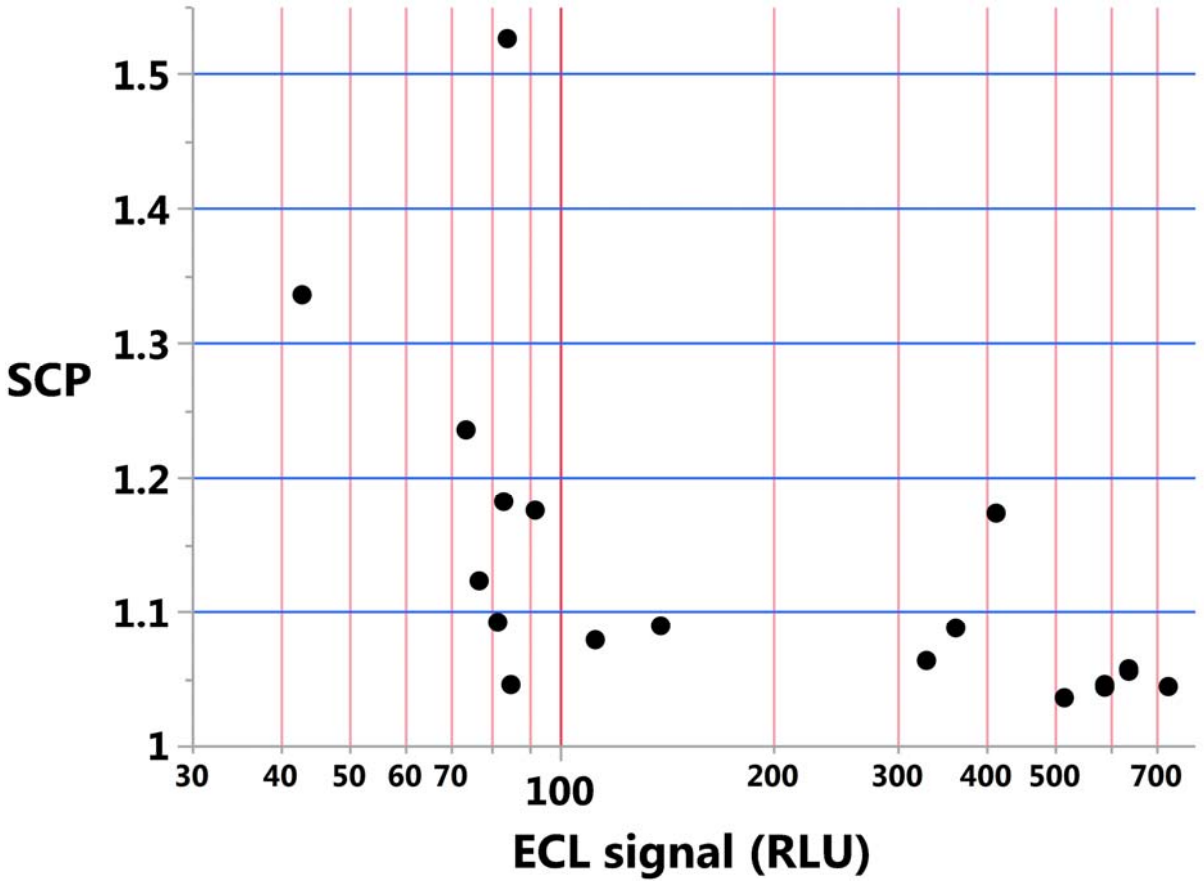
These questions will be addressed via retrospective evaluation of 25-30 assays; most of these assays have SCPF < 1.2

SCP factor vs. Biological variability



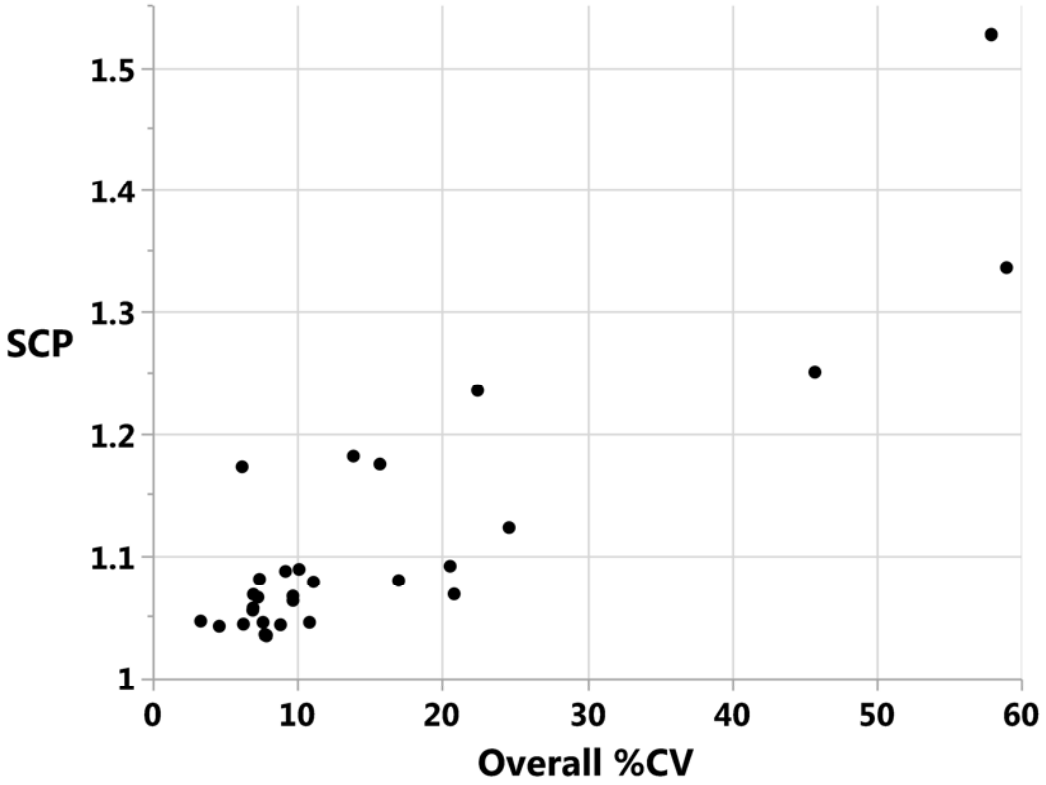
Assays with high biological variability can also have low SCP factors.

SCP factor vs. Assay signal



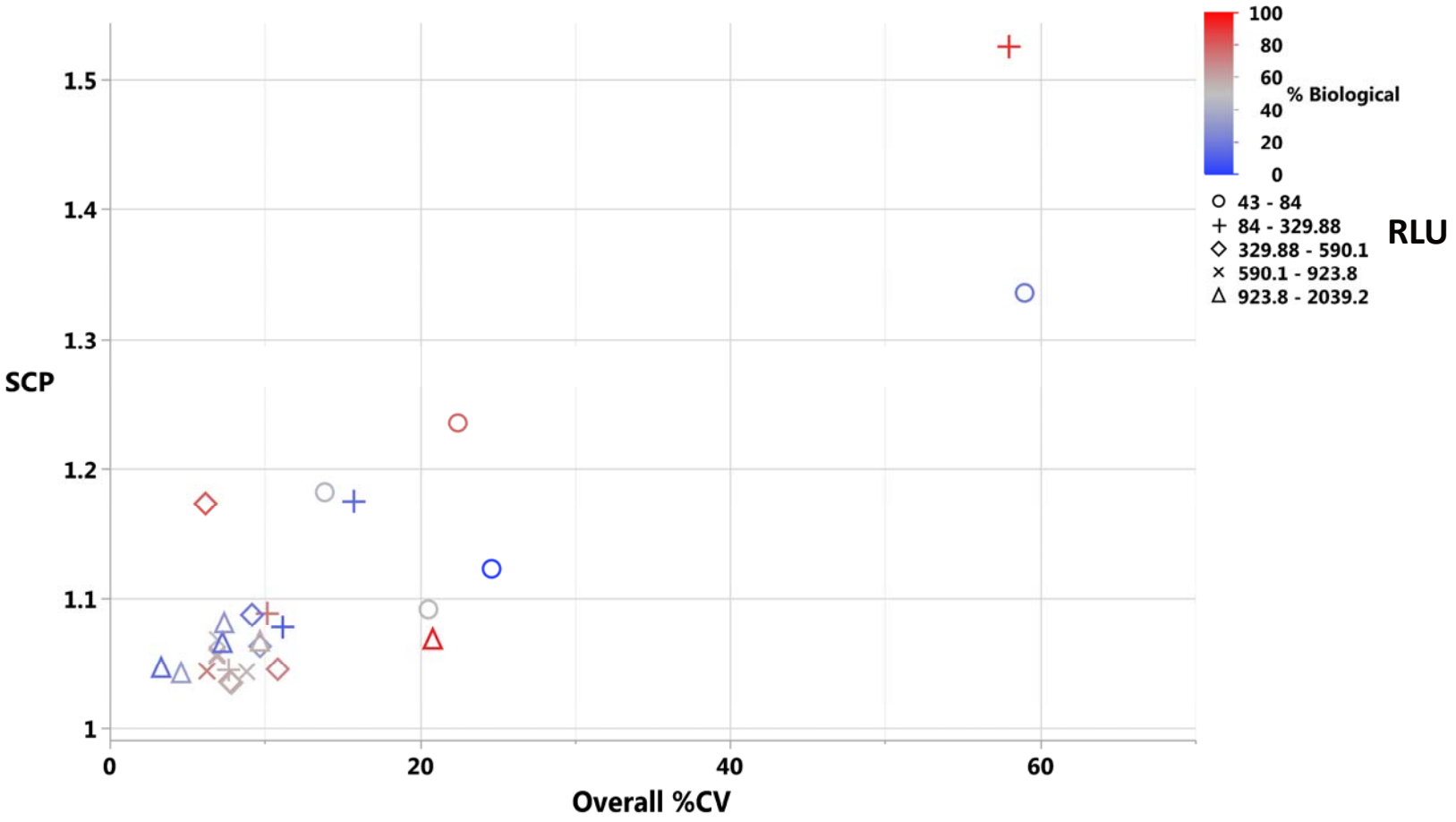
- Assays with high RLUs can also have low SCP factors.
- Low RLU (<100) does not always imply low SCP.

SCP factor vs. Total Variability (biological + analytical)



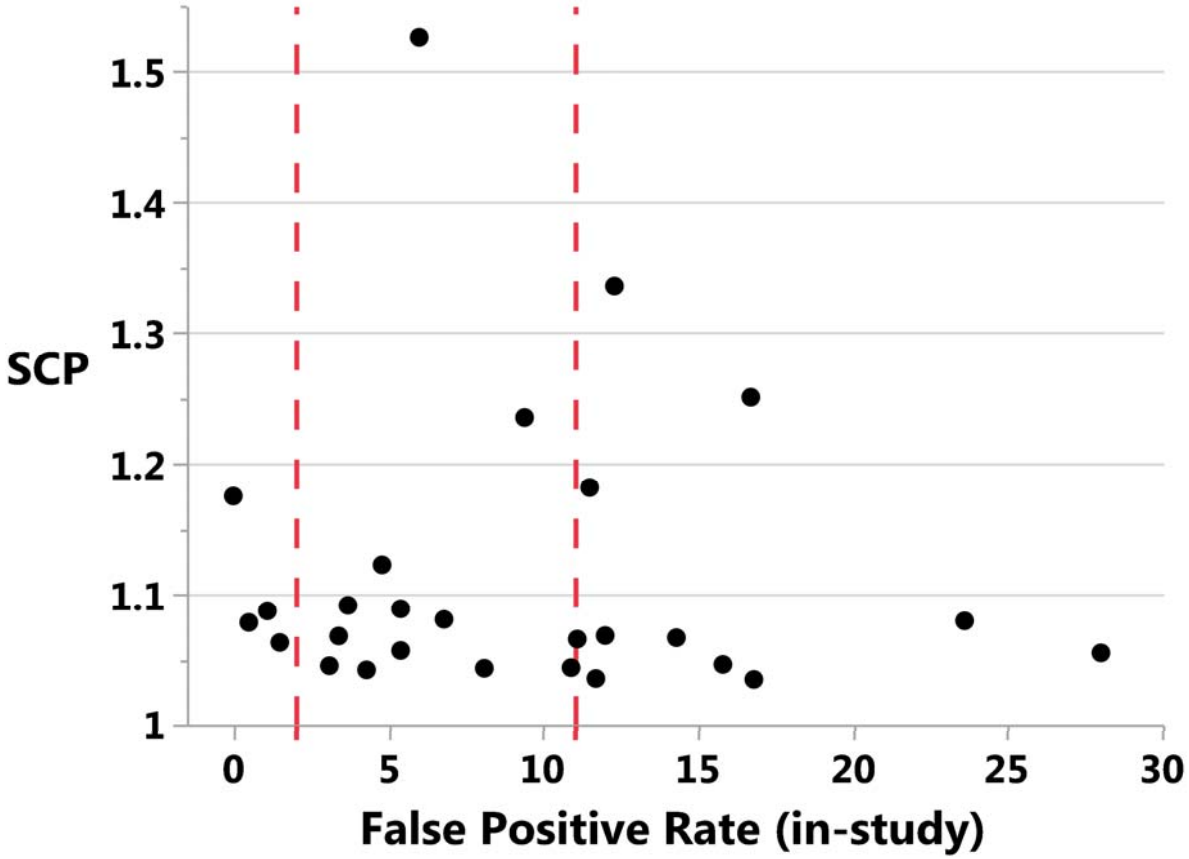
- **As expected, assays with higher total variability have higher SCP factor.**
- **Some departure from correlation is due to the differences between S and N**

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.

SCP vs. False Positive Rate (in-study)



- Low SCP does not always result in high in-study FPR, and therefore does not always require in-study cut point evaluation.
- Some of the high FPR observations are due to different disease populations.

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 2xIQR or 3xIQR.

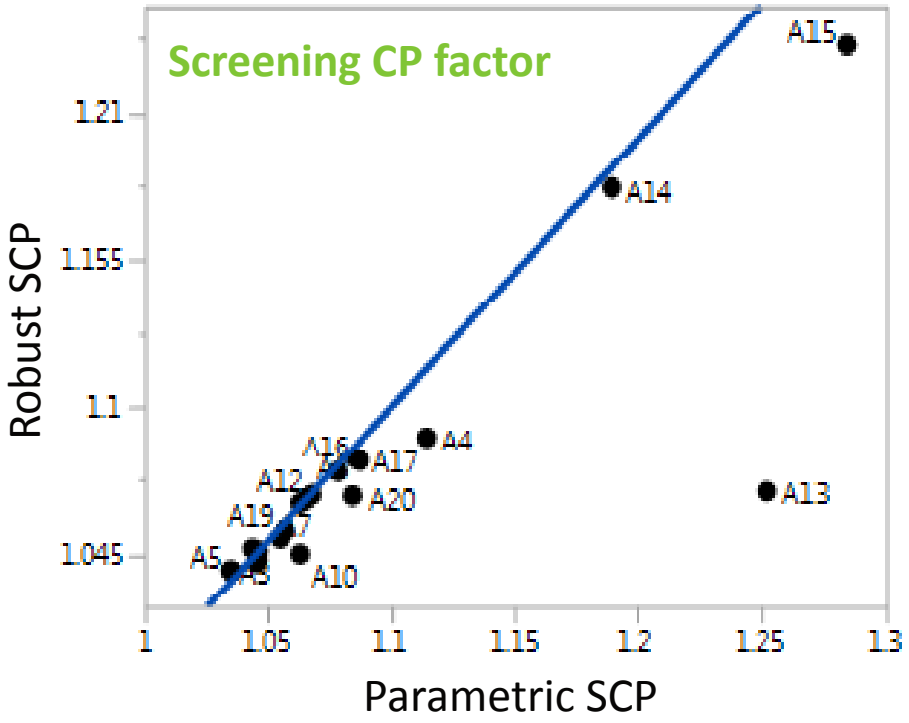
- Several presentations at WRIB-2018 and other meetings argue in favor of 3xIQR or other criteria that lead to higher CP values.
- This is due to the use of Mean/SD instead of Median/MAD.

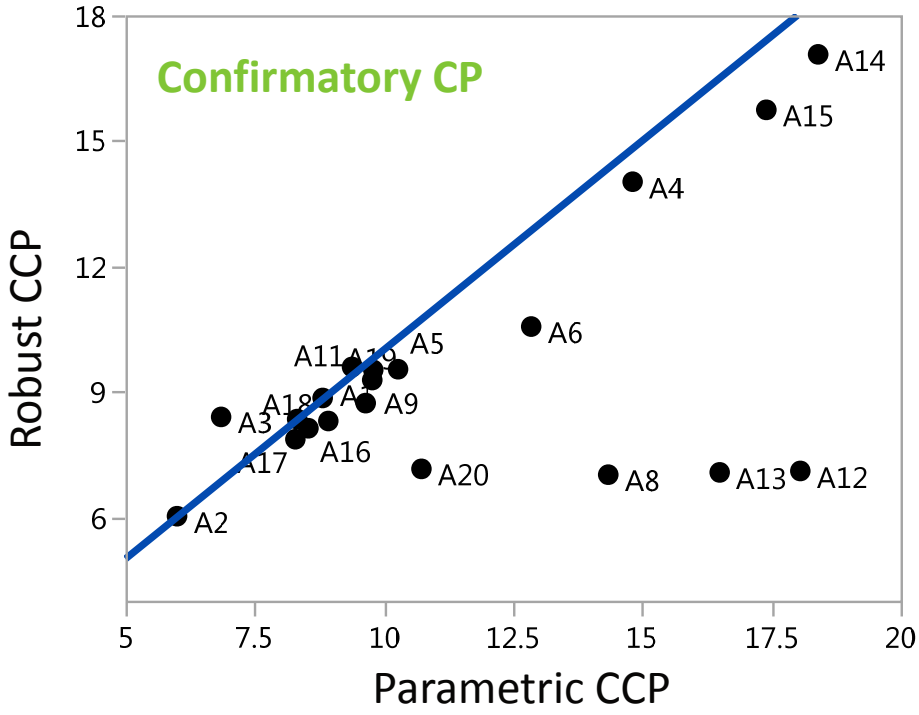
Slides that follow provide background on these methods, and demonstrate the use of robust methods to avoid this subjectivity.

Parametric vs. Robust-Parametric

Parametric: Mean + 1.645 x SD

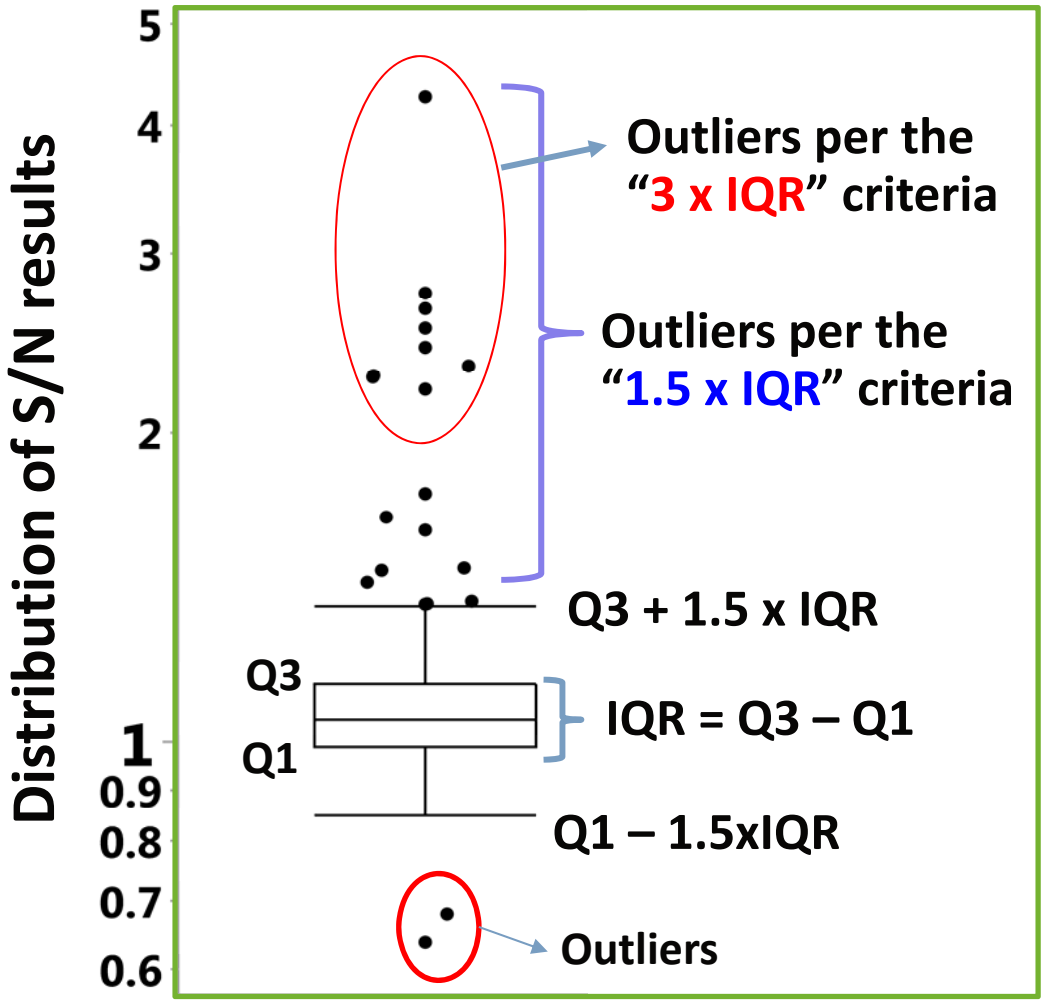
Robust Parametric: *Median + 1.645 x 1.4826 x MAD*





- Borderline outliers in some assays skew the parametric CP values (Mean/SD).
- Robust parametric (Median/MAD) is not impacted by these outliers, and is thus safer.
  - It is also more immune to arbitrary outlier decisions (e.g., 1.5xIQR vs. 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**

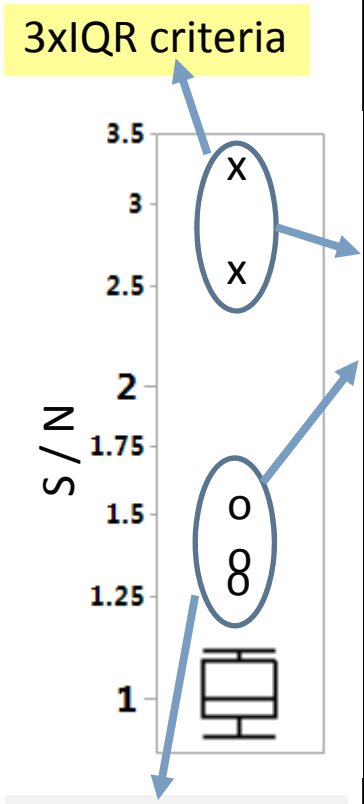
- ~ 99.9997% of the samples

Heavily debated at recent conferences.

- ***Such debate and arbitrary manipulation is unnecessary if robust methods are used.***

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

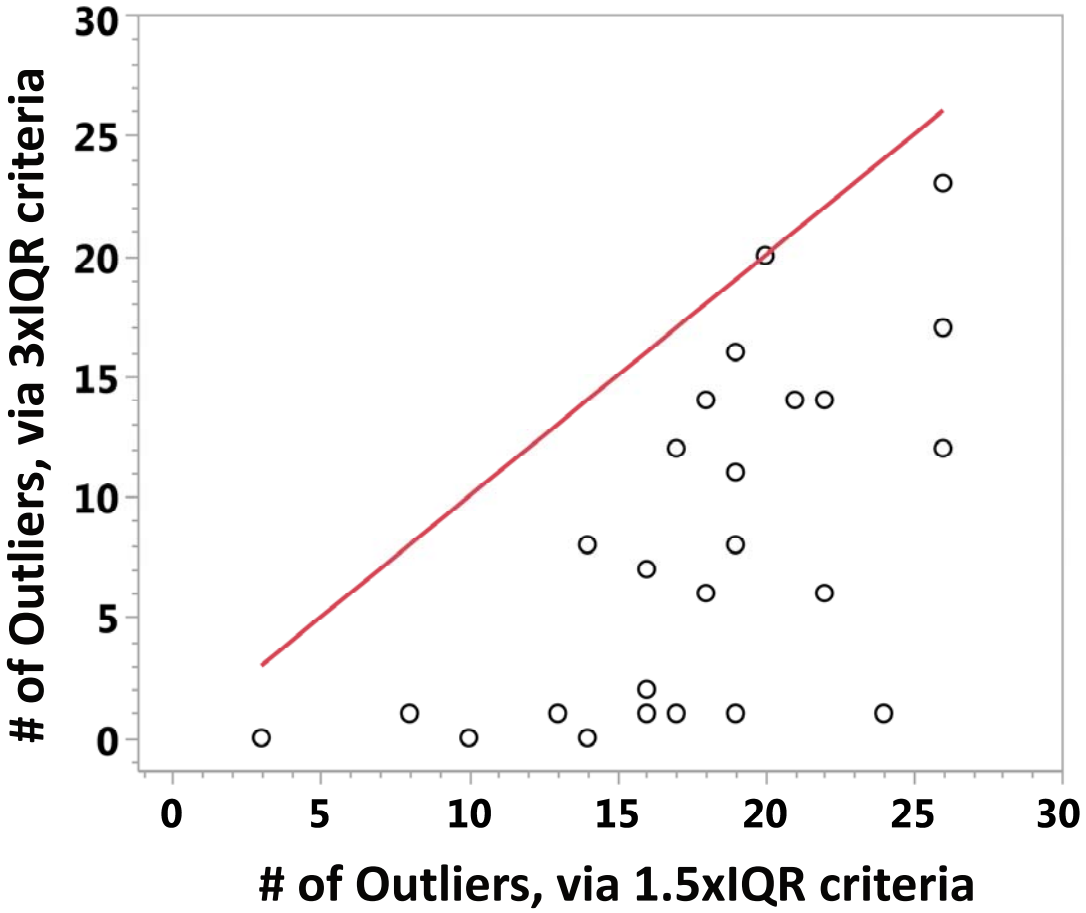
0.497 =| 0.514 – median(log(S/N)) |

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

MAD = 1.4826 x MAD*

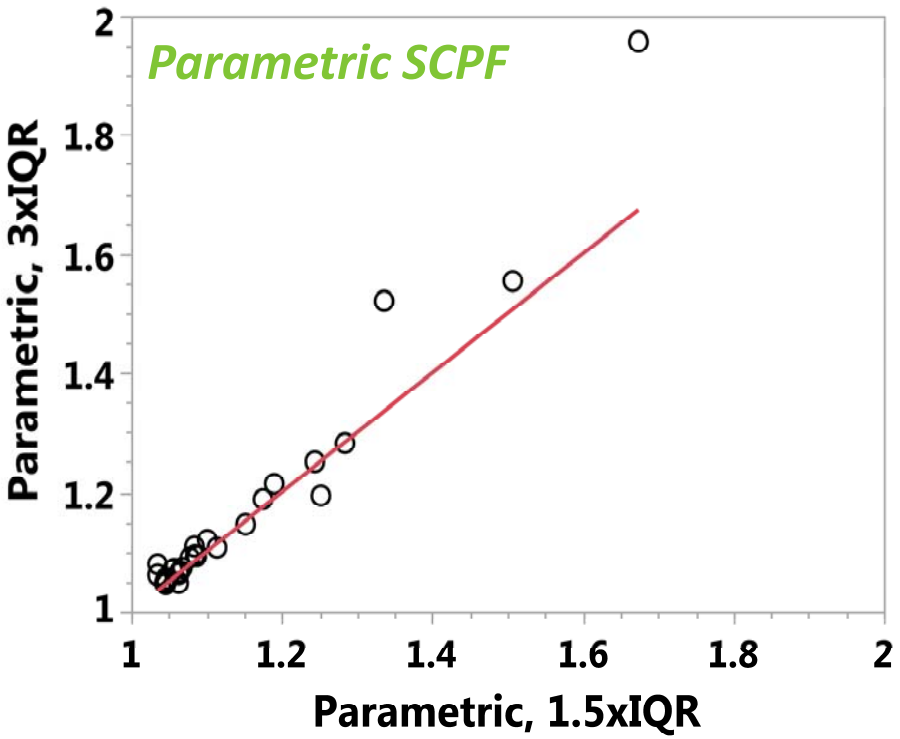
"Robust Parametric" SCP is highly resistant to different outlier identification criteria.

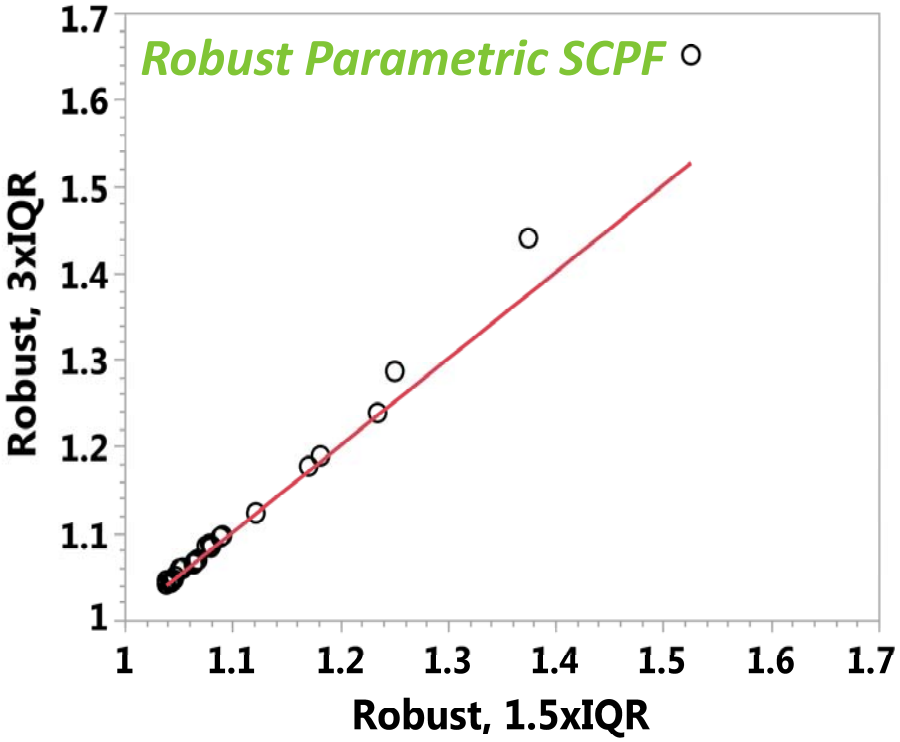
1.5xIQR vs. 3xIQR criteria; data from 25 assays



- 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 (contd.)





- **Robust method is not impacted by subjective changes and use of different outlier criteria such as 1.5xIQR or 3xIQR.**
- **As 1.5xIQR corresponds ~ 2.7xSD rule, this is a reasonable default.**

Clinical relevance

Lower cut point & therefore better sensitivity

- May imply more low-titer ADAs that are not clinical relevant, but….
- *Does not dilute overall clinical relevance of all ADA results!*

On the contrary, it 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…

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

p-value ~ 0.01

Suppose improved assay sensitivity led to 90 additional low-titer ADAs:

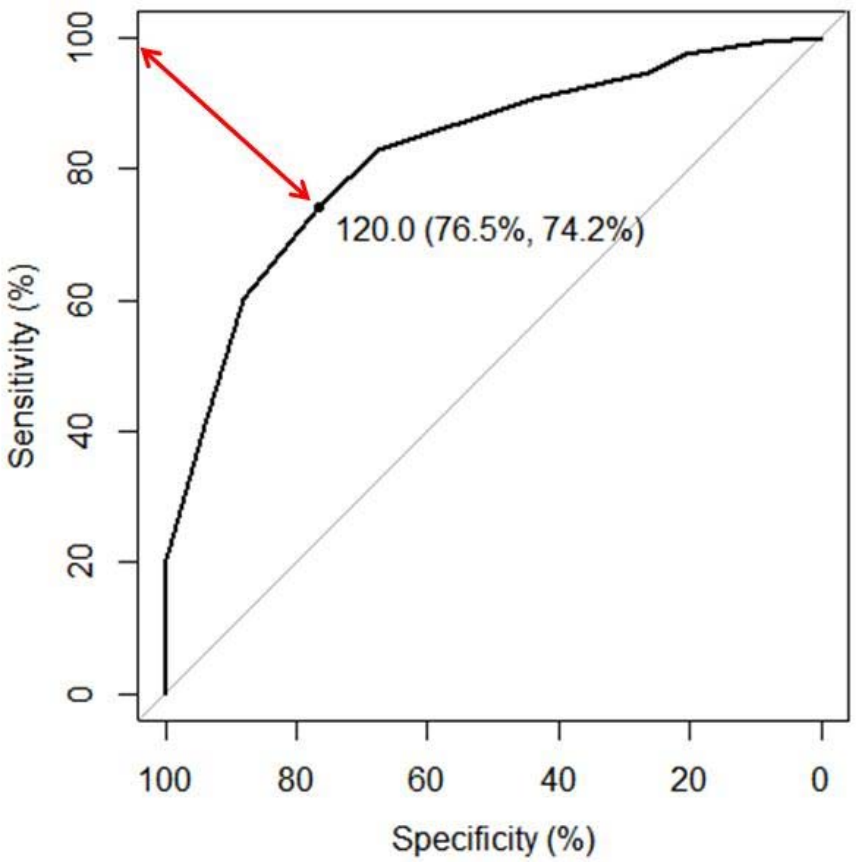
Clinical Impact	Titer < 50	Titer > 50
no	80	10
yes	20	20
Total	100	30

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.

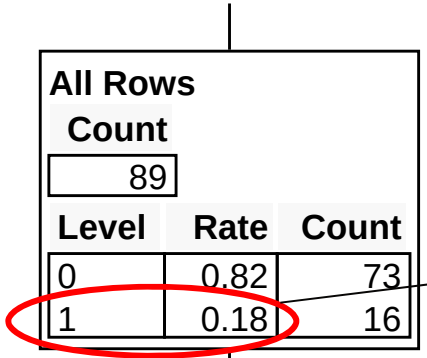
Increased assay sensitivity, with higher incidence of low-titer ADAs, will not dilute this clinical impact.

In fact, it helps strengthen/refine the overall evidence.

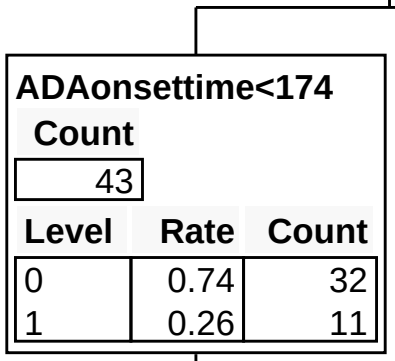
Combining the Titer results with ADA kinetics such as "Onset Time" and "Duration" (transient vs. degree of "persistence") via multivariate analysis may provide additional insights on clinical relevance (Shankar et al., 2014) ***– next slide.***

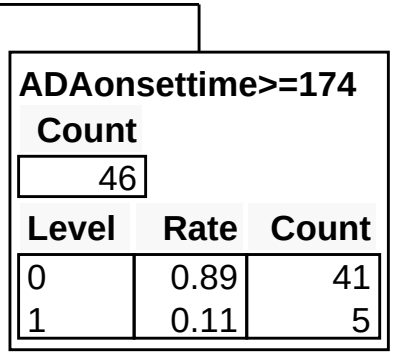
Example 3

Increased assay sensitivity (lower CP) doesn't dilute the clinical evidence. Multivariate evaluation of Titers with ADA Kinetics provides more insights on clinical relevance.



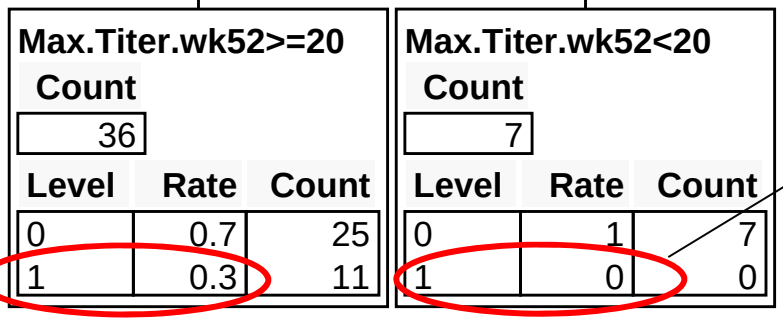
18% of patients have AE (n=16)





**Significance of onset-time is not strong.*

Higher AE incidence (30%) for patients with ADA titer > 20 & onset within 6 months.



No AE incidence at low ADA titer.

Summary

- Low SCPF is not always due to low biological variability!
  - Total variability is the key factor, regardless of the relative level of biological versus analytical variability.
- Assays with high signal (RLU) can also have low cut points.
- Low RLU (<100) does not always imply low SCPF.
  - Artificially inflating the signal by manipulating sample treatment may not help, and may not be necessary!
- Low SCPF does not always result in high in-study FPR, and therefore does not always require in-study cut points.
- Subjective manipulation of outliers is not necessary if robust (Median/MAD) method is used with 1.5xIQR criteria.
- *Overall* clinical relevance is not compromised by low SCPF.
- The updated flow-scheme in Devanarayan et al. (2017) tends to work for a wide variety of data scenarios.

Acknowledgments

Ron Bowsher & Wendell Smith (B2S), Rob Nelson (Novimmune)

Colleagues from Charles River Labs, and former colleagues from AbbVie.

Thanks also to the coauthors of these seminal papers from whom I have learned a lot over the past 15 years in this field.

1. Mire-Sluis et al., 2004, JIM *(ADA screening - design elements)*
2. Koren et al., 2007, JIM *(ADA testing strategy)*
3. Shankar et al., 2008, JPBA *(ADA screening – method validation)*
4. Gupta et al., *2011, JPBA (NAb – method validation)*
5. *USP chapter on Immunogenicity screening methods, 2013*
6. *USP chapter on Neutralizing Antibody methods, 2014*
7. Shankar et al., 2014*, AAPS Journal (Clinical ADA interpretation)*
8. Shankar et al., 2015*, Nature Biotech (Clinical ADA reporting)*
9. Amaravadi et al., 2015, Bioanalysis (*additional considerations*)
10. Devanarayan et al., 2017, *AAPS J.* (*updated CP recommendations*)