Screening & Confirmatory Cut-Points
- Brief Overview & Answers to some FAQs

V. Devanarayan, Ph.D.
Abbott Laboratories, USA

E-Mail: viswanath.devanarayan@abbott.com

EIP Symposium
Immunogenicity of Biopharmaceuticals
Leiden, NL, November 18-19, 2009

Abbott
A Promise for Life

Acknowledgements

Coauthors of the AAPS-sponsored white papers

- Mire-Sluis et al., 2004, JIM *(design of ADA immunoassays)*
- Koren et al., 2007, JIM *(immunogenicity testing strategy)*
- ***Shankar et al., 2008, JPBA (validation of ADA immunoassays)***
- Gupta et al., in preparation (NAb assay validation)

Other members of AAPS-LBABFG

Colleagues from the Drug Analysis group in Abbott Labs

About the statistical recommendations in our white papers

Goal is to reduce subjectivity & increase objectivity

Assumed that not everyone has access to a statistician on this topic.

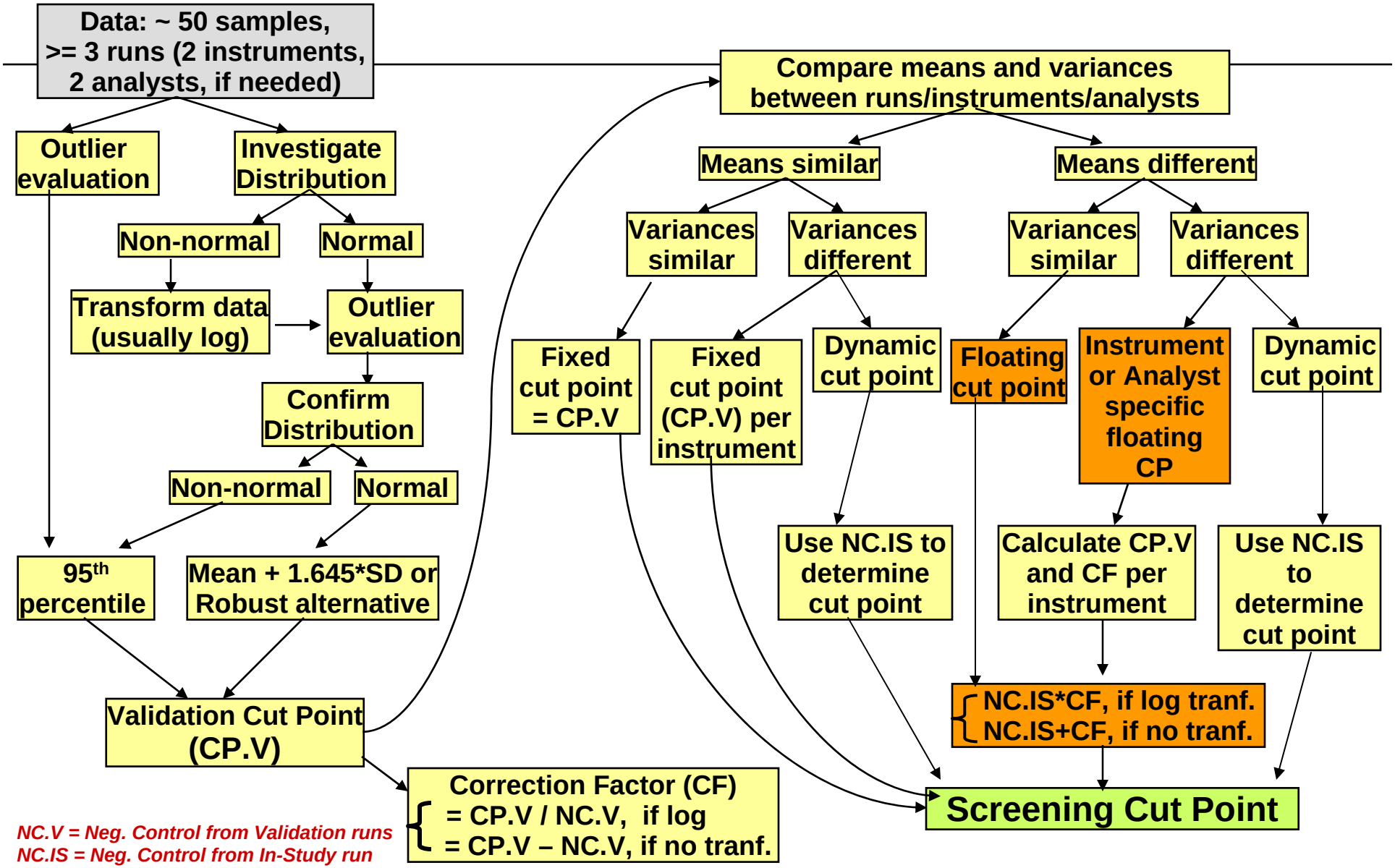
So the suggested stats are “relatively simple”, can be implemented with friendly software such as JMP.

- Compromise between optimal versus simple statistics.
- For many of the topics, more rigorous and elegant statistical methods are available, but not presented in this white paper.

When possible, consultation and engagement of statistician is highly recommended.

Screening Cut Point

Screening CP Evaluation Flow-Scheme



Unbalanced: Run-Confounded

Analyst	Assay Run	Assay Plate	Validation Serum Samples		
			$S_1 - S_{16}$	$S_{17} - S_{32}$	$S_{33} - S_{48}$
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

Not all samples get tested in every run.

→ Difference between runs are confounded with the difference between sample sets.

Unbalanced: Analyst-Confounded

Analyst	Assay Run	Assay Plate	Validation Serum Samples			
			$S_1 - S_{12}$	$S_{13} - S_{24}$	$S_{25} - S_{36}$	$S_{37} - S_{48}$
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

Not all samples get tested by the two analysts.

→ Difference between analysts are confounded with the difference between sample sets.

Balanced design (plate-balanced)

Analyst	Assay Run	Assay Plate	Validation Serum Samples		
			$S_1 – S_{16}$	$S_{17} – S_{32}$	$S_{33} – S_{48}$
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	

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

Some FAQs about Screening CP

Why normal distribution? Are low outliers important?

Fixed vs. Floating CP?

Additive vs. Multiplicative Correction Factor?

How to calculate SD in CP formula?

Is our NC appropriate for the Floating-CP evaluations?

Can we use the same CP for a different study/population?

Distribution & Outliers (contd.)

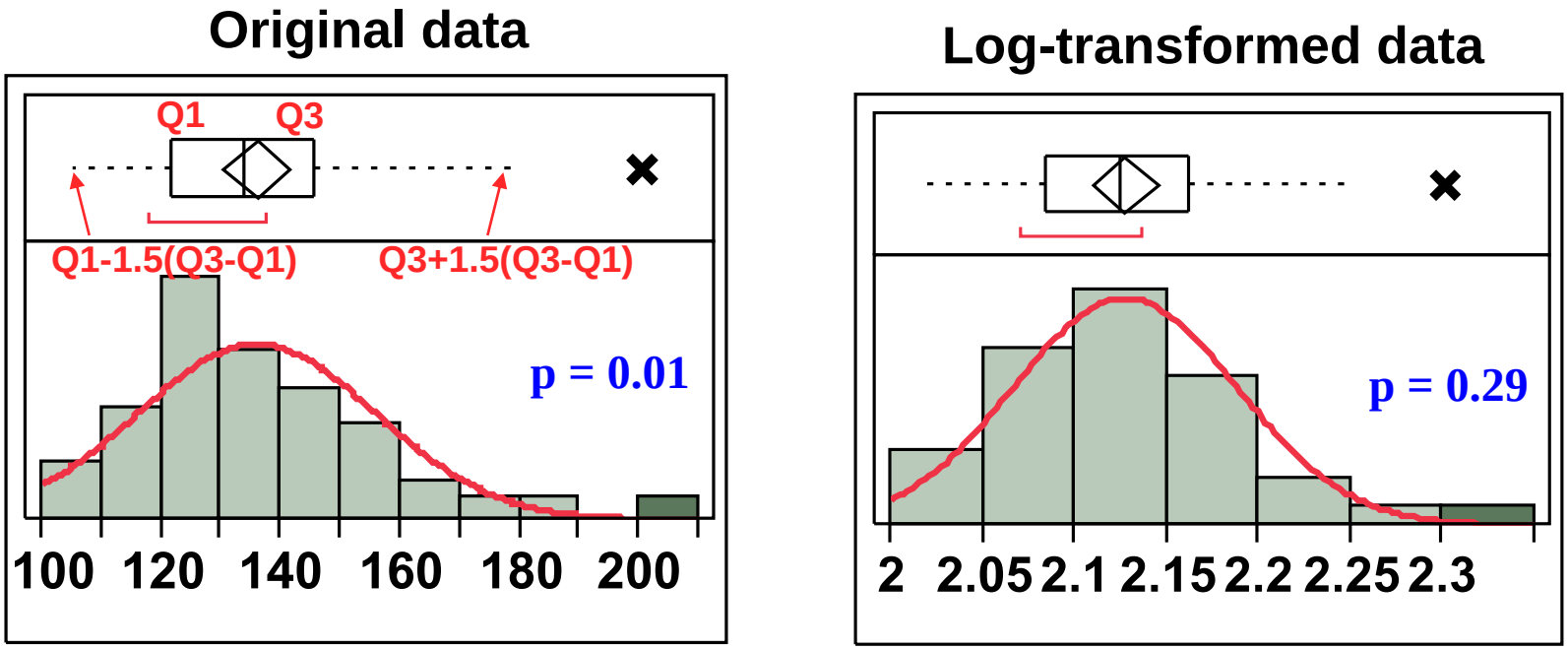
"Outlier Box-Plot" rule for excluding outliers from each run

- Samples **> Q3 + 1.5*(Q3-Q1)** or **< Q1 - 1.5*(Q3-Q1)**
  - Q3 = 75th percentile
  - Q1 = 25th percentile (Q2 = median)

Evaluate the distribution of data from each run, and for data averaged across runs, after excluding all the outliers.

- If the Shapiro-Wilk test for normality fails ($p < 0.05$), use non-parametric method (95th percentile) to evaluate screening cut-point.
- Otherwise, use the parametric method (Mean + 1.645xSD).
- But statistically preferred approach would be to use the residuals from the ANOVA.

Distribution & Outliers



- **Distribution of data from 48 subjects, averaged from 6 runs.**
- **Log-transformed data are relatively less non-normal.**
- **Further analysis should be performed in log-scale.**

Cut Point: Distribution & Outliers

Why ~ Normal distribution?

- Comparison of means and variances between assay run means
- Parametric approach to calculate cut point.

Why not just use Nonparametric (95th percentile) method?

- Requires larger sample size than parametric method.

Why exclude/down-weight outliers?

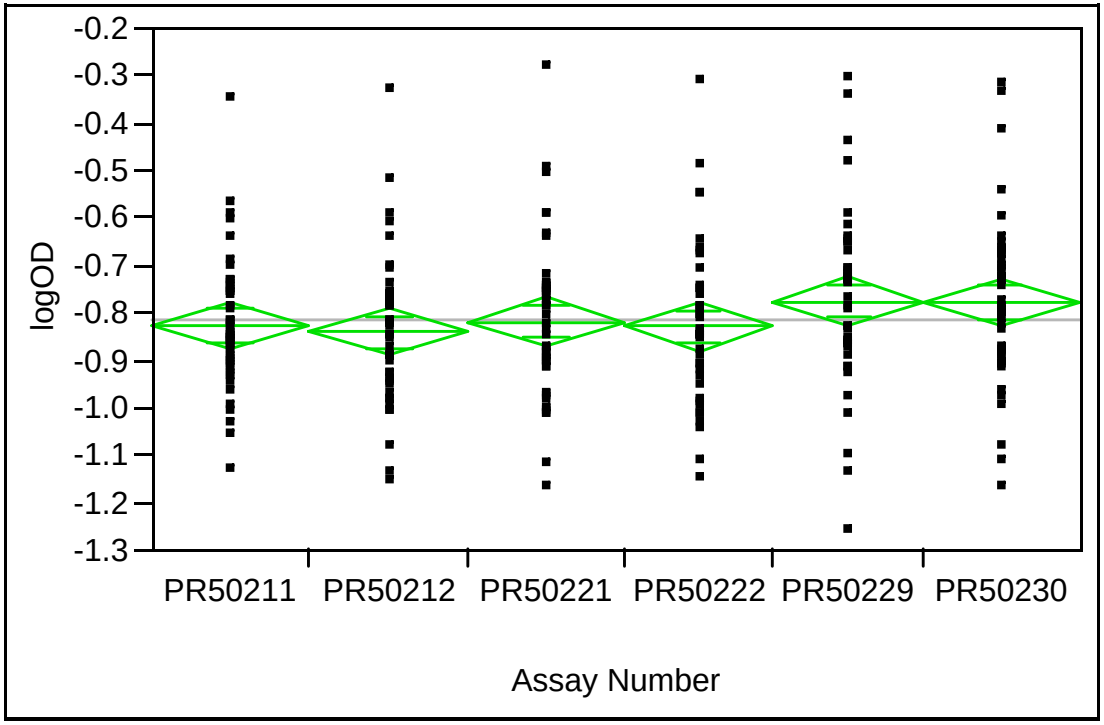
- To ensure that the false positive rate applies only to the **expected range of ADA negative response** (also, due to risk-based strategy).

Why worry about low outliers? (for risk-based strategy)

- While Low outliers deflate the mean, they may significantly inflate the variability, and hence the cut point.

Fixed or Floating?

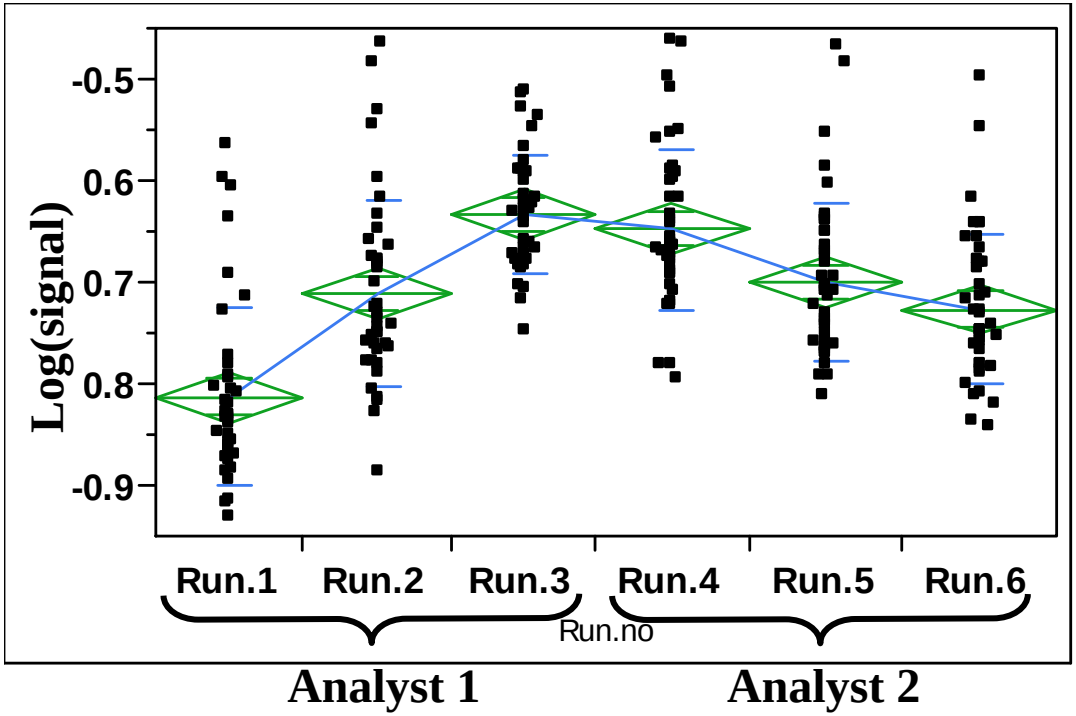
Example 1



Means not significantly different (p=0.31)
Variability not significantly different (Levene's test, p=0.66)
→ Fixed cut point method can be used.

Fixed or Floating?

Example 2

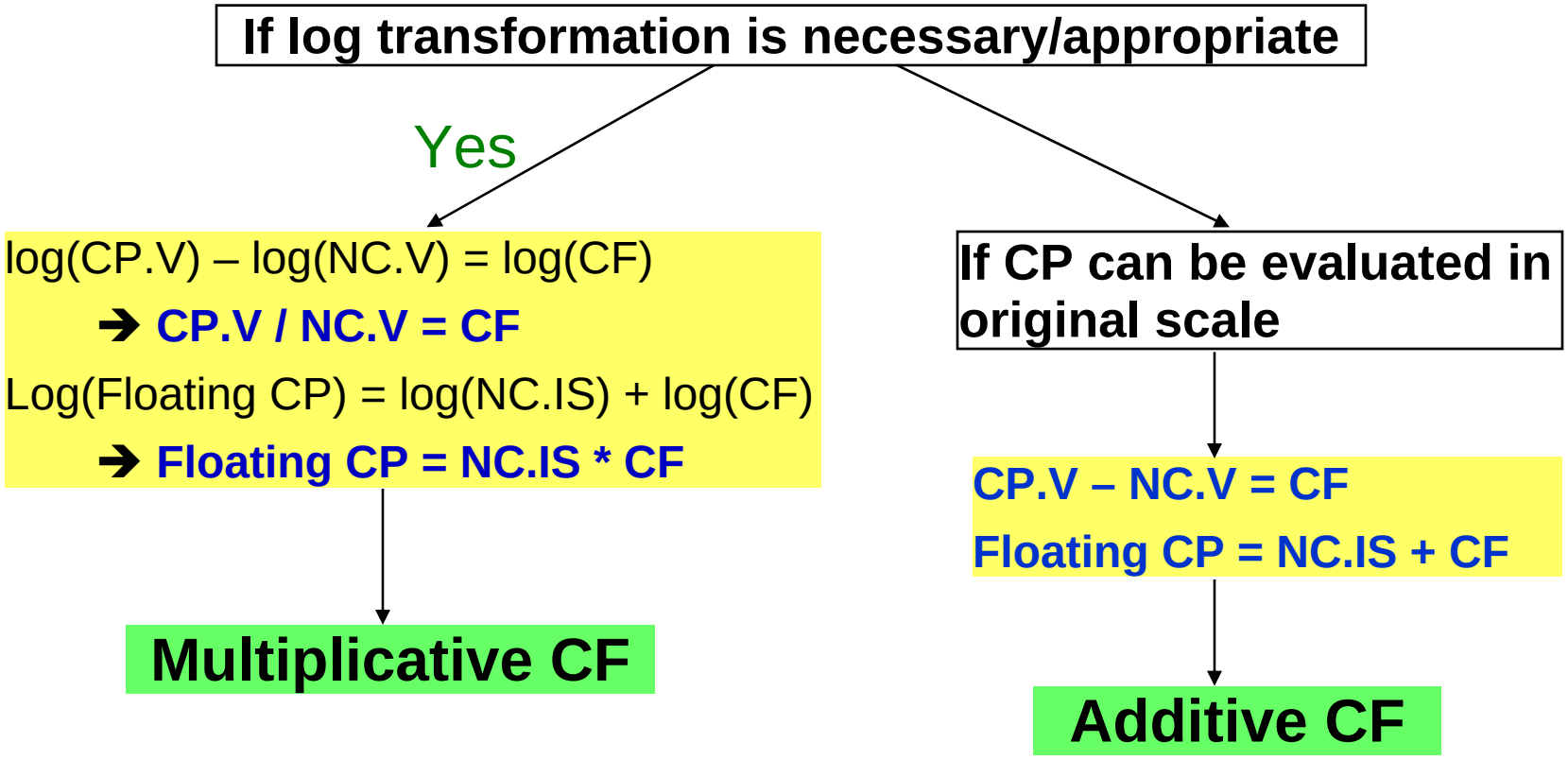


Means significantly different (p<0.0001)
Variability not significantly different (Levene's test, p=0.37)
→ Floating cut point method should be used.

Additive or Multiplicative Correction Factor ?

(for Floating Cut Point)

CF = Correction Factor
CP.V = Cut Point from validation data
NC.V = Neg. Control from Validation runs
NC.IS = Neg. Control from In-Study run



Floating Cut Point

Is our Negative Control (NC) appropriate?

Use of plate-specific NC with the correction factor (CF) **assumes** that NC trends with subject samples.

In order to ensure the validity of this approach:

- Plot the NC mean versus NHS mean from six runs. Assess significance of the linear correlation ($p<0.05$, and/or say, $R^2>80\%$.
- Alternatively, include NC values as covariates in the model. Assess significance of the NC term ($p<0.05$).

When this assumption fails, the use of negative control for defining a floating-CP might be futile.

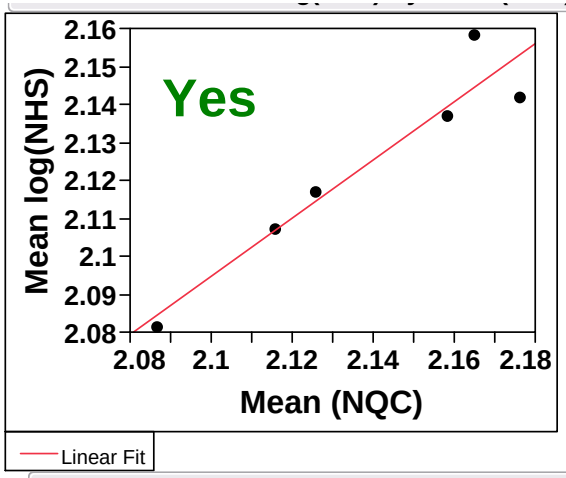
- More likely when analytical variability exceeds biological variability.

Alternatives: New pool, Subject specific cut-point, other controls?

Floating Cut Point (contd.)

Is our Negative Control (NC) appropriate?

Example-1



Linear Fit

Mean log(NHS) = 0.4838325 + 0.7670203*Mean (NQC)

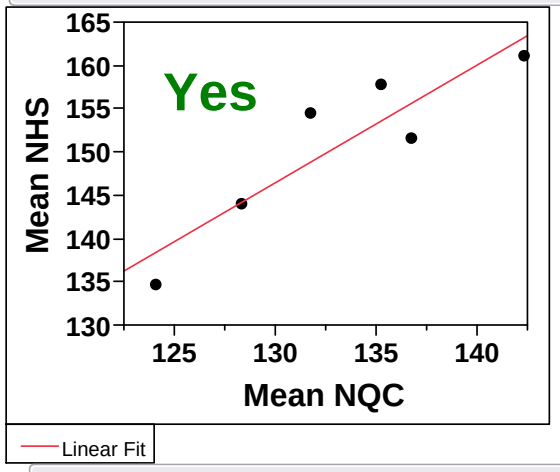
Summary of Fit

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

Mean NHS = -31.16837 + 1.3659528*Mean NQC

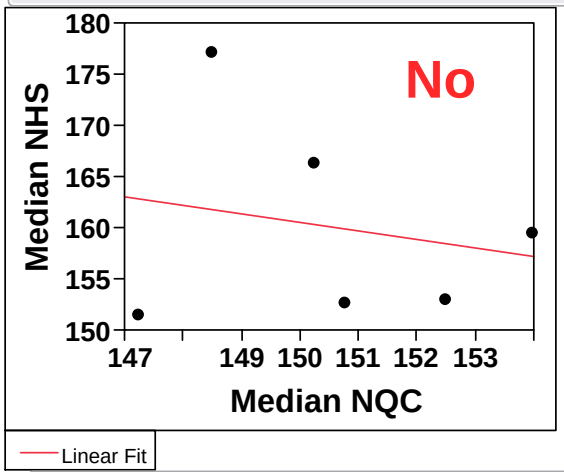
Summary of Fit

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

Median NHS = 287.03853 - 0.8435759*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

Evaluation of SD for Cut Point

Need to consider both Biological & Analytical variability.

- That is, **Inter-Subject** variation & **Intra-Run** variation respectively.

SD that incorporates both biological & analytical variation can be obtained from random effects ANOVA, or from a simple spreadsheet calculation.

*If calculated this way, it shouldn't really matter whether it is a low or high background assay (*ECL, ELISA, etc.)

Evaluation of SD for Cut Point

Simple approach

Subject #	Run 1	Run 2	Run 3	……..	Run 6
1	185	191	182	……..	189
2	186	181	187	……..	192
3	182	185	190	……..	183
.					
.					
.					
50	178	180	182	……..	187

Common mistake:

- Mean for each subject is first determined.
- SD of these means is used in CP formula.
- ***Ignores analytical variability***!

Recommendation:

- **Calculate Variance of the donor sample results from each run.**
- **Pool these variances across runs.**
  - **Weight each run's variance using its degrees of freedom (DF).**
    - **This will be different if outliers were eliminated.**
    - **Just a simple average if no outliers were removed.**
- **SD for cut point = sqrt(pooled variance)**

Evaluation of **SD** for Cut Point

More rigorous/correct approach

JMP output from "random effects ANOVA" (variance-components)

REML Variance Component Estimates

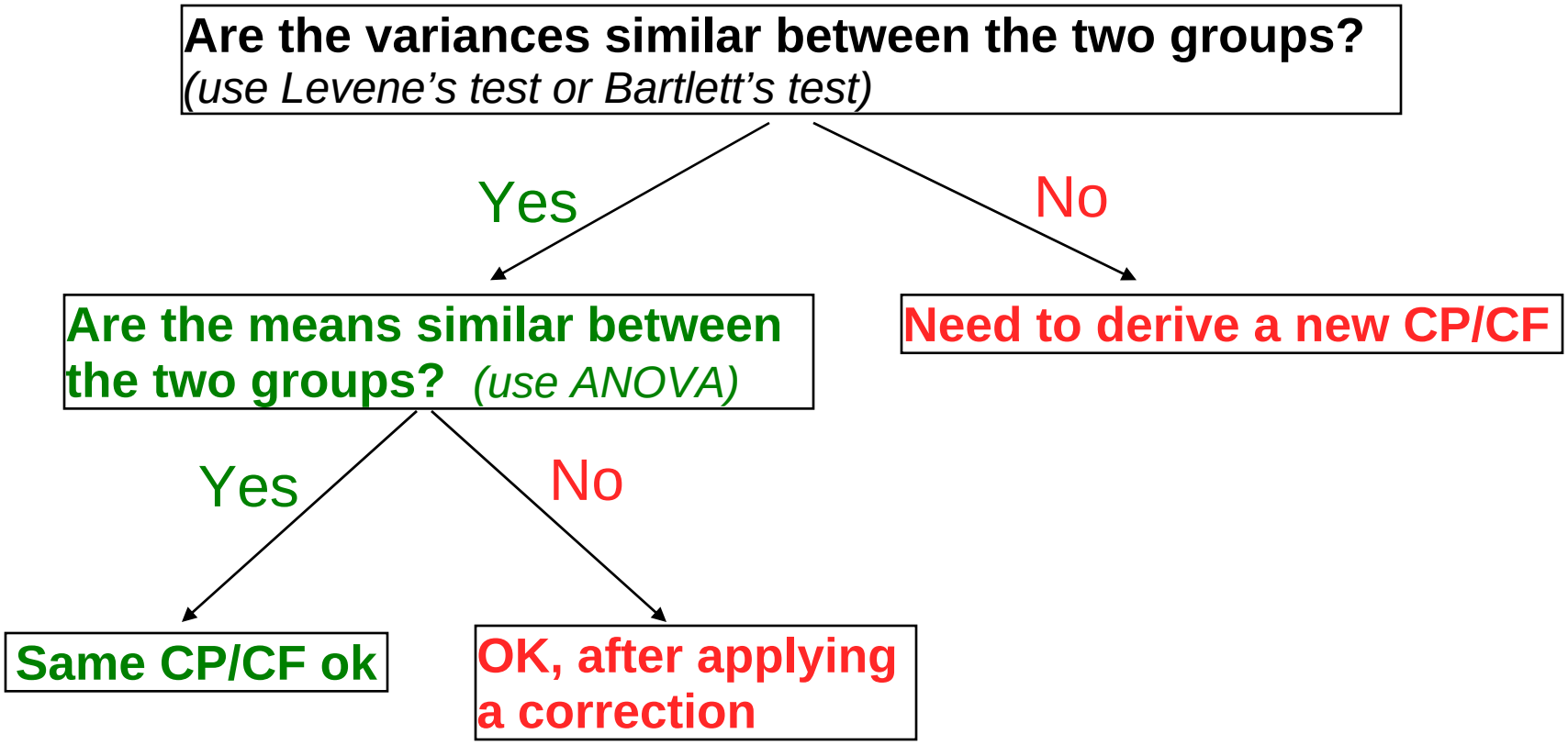
Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Pct of Total
Run.a[Analyst]	0.2564771	55.331561	42.313672	-27.60324	138.26636	7.638
Sample.ID	2.1016299	453.39896	101.11769	255.20829	651.58962	62.584
Residual		215.73682	19.988906	181.35419	260.97284	29.779
Total		724.46734				100.000

Var(Inter-subject) ~ 453.399, Var(Intra-run) ~ 215.737

SD for cut point = sqrt(453.399 + 215.737)

If log transf. is necessary, SD and CP should be calculated in log scale first, and then converted back.

Can we use the same CP or CF for a different disease/target population?



Confirmatory Cut Point

FAQs about Confirmatory Cut Point

How to handle negative inhibition when log transformation is necessary?

Fixed vs. Floating?

- Due to within-plate normalization, what should we expect?
- Is it necessary to require the same # of runs (6) for this CP?

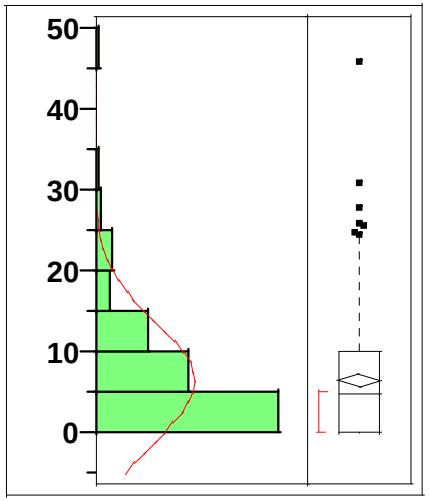
We need **non-specific** ADA positives, but we use drug-naïve ADA negatives for this CP evaluation.

- Is this reasonable? What are the assumptions? How to verify?
- Alternative sampling strategies?
- Alternative approaches to confirmatory CP? Which method is less subjective?

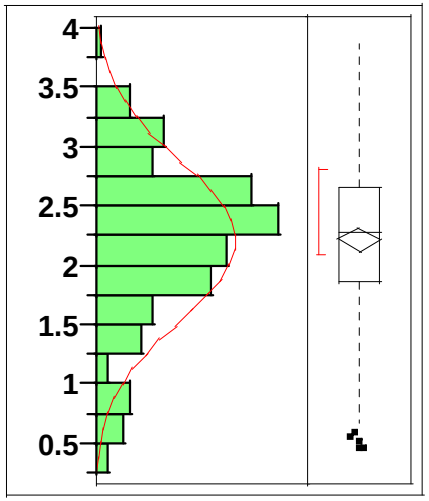
Confirmatory Cut Point

$$\%\text{Inhibition} = 100 * \left(1 - \frac{\text{Signal of sample with Inhibitor(s)}}{\text{Signal of unspiked sample(us)}}\right)$$

%Inhibition



after log tranf.



Subjective thresholds (e.g., 50%) were widely used in the past (< 2006).

Devanarayan, 5-2006, IIR conference: proposed a cut-point strategy.

Shankar et al (2008) published the objective cut-point approach based on biological & analytical variability.

Confirmatory Cut Point

Experiment, Data & Calculations

Data:

~ 50 drug naïve samples spiked with drug (preferably with the same samples & in the same Screening CP experiment)

~ 4-6 runs total, >= 2 instruments/analysts when appropriate

Compute % inhibition for each sample. ***Take log if needed***.

Confirmation CP = Mean (% Inhibition) + 2.33*SD (% Inhib.)

Distribution, outliers, variability, etc., should be investigated.

2.33 ➔ 1% false positive rate (use 3.09, if 0.1% is desired)

How to handle samples with negative inhibition? (next slide)

Confirmatory Cut Point

Samples with negative inhibition?

When log transformation is necessary and there are negative inhibition samples:

- Determine CP_1 of log(s/us), then apply the conversion formula below to determine CP_{99} of %inhibition.
- Use parametric method if distribution is adequately normal.

$$\%Inhibition = 100 * \left(1 - \frac{\text{Signal of sample with Inhibitor (s)}}{\text{Signal of unspiked sample (us)}}\right) = 100 * \left(1 - \frac{s}{us}\right)$$

$$= 100 * \left(1 - 10^{\log\left((\frac{s}{us})\right)}\right)$$

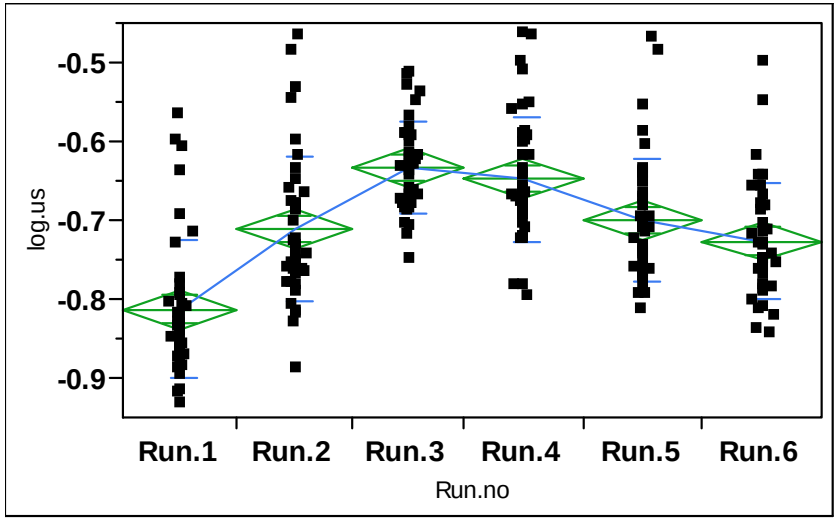
$$CP_{99}(\%\text{inhibition}) = 100 * \left(1 - 10^{CP_1\left(\log\left(\frac{s}{us}\right)\right)}\right)$$

where, $CP_{99}(.) = 99th$ percentile cutpoint, $CP_1(.) = 1st$ percentile cutpoint

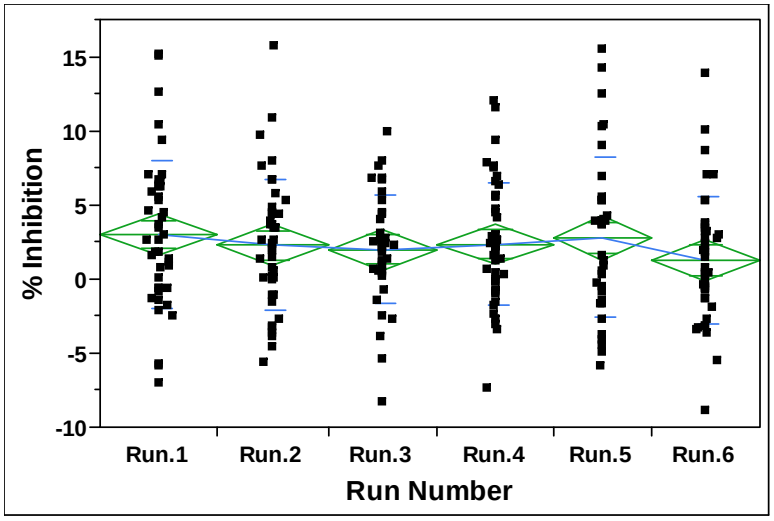
Confirmatory Cut Point

Fixed or Floating?

Assay signal of unspiked samples



% inhibition of spiked samples



Unspiked signal is sig. different across runs, but % inhibition is not.

This is due to the internal intra-plate normalization, i.e., unspiked and spiked samples were run in the same plate.

- **Fixed cut point will usually suffice. But need to verify!**

6 runs of data shouldn't be required. 2-4 runs may suffice.

Confirmatory Cut Point

Assumption

Target population:

- Non-Specific ADA Reactive samples

  (i.e., drug naïve samples with signal > screening cut point)

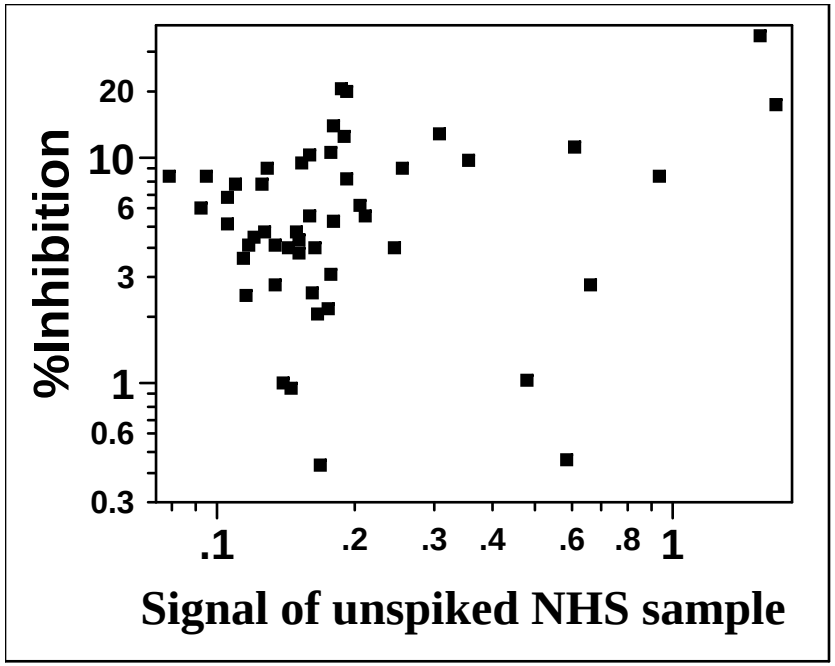
We usually don't have enough of these samples.

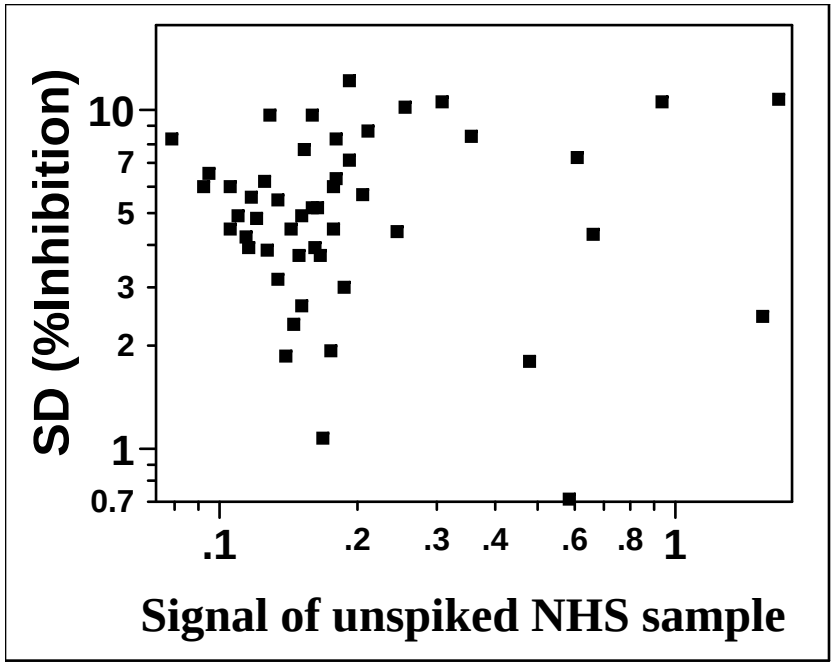
So we assume that

- % inhibition of drug naïve samples spiked with the inhibitor has similar distribution (mean & variance) to the inhibition of non-specific ADA positive samples.
- Is this a reasonable assumption???
  - Probably Yes for most assays, but may be useful to check this on a case by case basis.

Verifying Assumption

Case Study 1

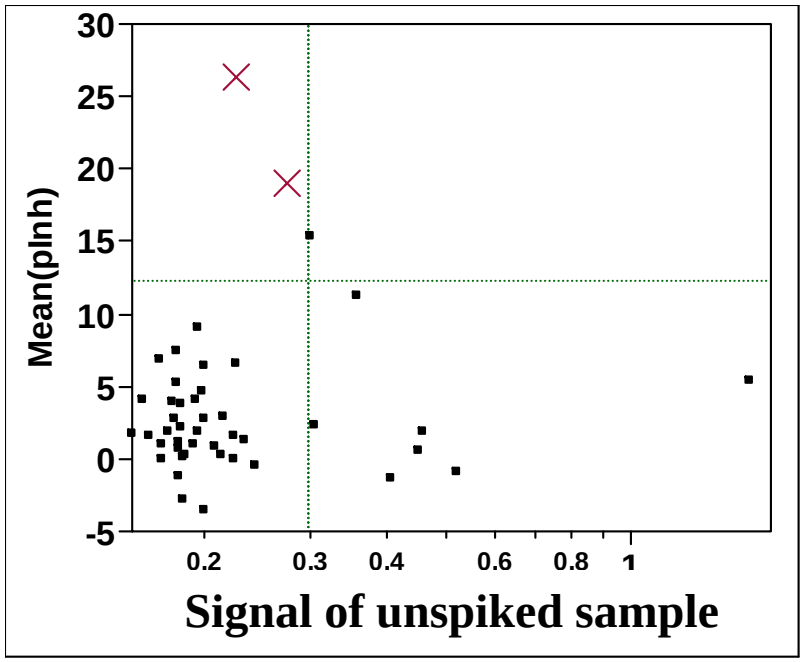


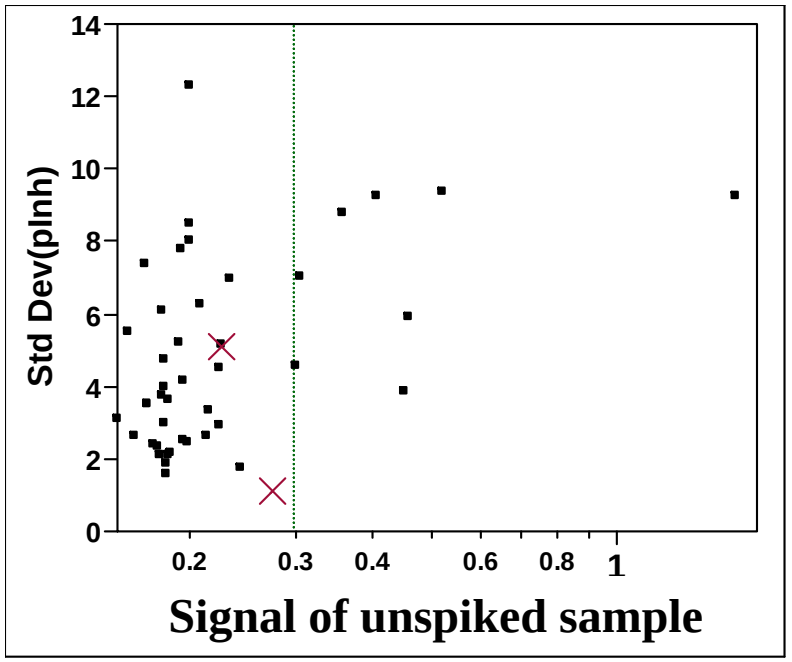


The mean and variability of % inhibition appear similar between the low and high samples.

Verifying Assumption

Case Study 2





Mean % inhibition is fairly similar across the range, but variability appears to be higher among the high samples, but there aren't that many.

Alternative sampling strategy when this assumption does not hold…

Sort all samples (~50) from the screening cut point experiment with respect to assay signal, from high to low.

Select the top ~25 samples (with highest signal).

Analyze the %inhibition data from only these ~25 samples.

Determine the CP of %inhibition as described earlier.

Alternative approach to Confirmation CP

Spike all individual negative samples in the experiment described in previous slides with low conc. of Control Ab.

- ~ assay sensitivity or low positive control ("**mock positive**" samples)

Calculate % inhibition of these Ab spiked samples

- Perform similar evaluations as described earlier for cut point analysis
- Determine Mean and SD of the % Inhibition values

Then determine the **lower 99th percentile** (~1st percentile) using either the parametric or nonparametric approach.

- **Mean – 2.33xSD**
- Controls for false negative rate to a desired level (e.g., ~1%).

Samples that fall below this cut point are considered as non-specific.

Comparison of the two approaches

Using individual negative samples (not spiked with Ab):

- Representative of the non-specific samples needed for setting the threshold (cut point) above which we can say with confidence that the screen-positive samples are specific to the drug.
- ***Controls the false positive rate at a desired level***.

Using mock positives (Control Ab-spiked individual neg samples):

- *These are not non-specific samples.*
- ***Cut point may vary depending on the affinity & other issues of the chosen Control Antibody.***
- ***Also, depends on the concentration chosen for the control Ab.***
- ***Controls the false negative rate, but not the false positive rate.***

"Ask Not about validating a developed Cut Point; Ask about developing a valid Cut Point."

Thank You for your attention!