Simplified strategy for Immunogenicity cut-point evaluations & some practical considerations

V. Devanarayan, Ph.D.

AbbVie Inc., USA

European Immunogenicity Platform (EIP) Workshop

February 24-26, 2014

abbvie

Recommendations presented here were influenced by the following collaborations:

Wendell Smith & Ron Bowsher, B^2S Consulting

DMPK-BioAnalysis colleagues at AbbVie.

Coauthors of the AAPS white-papers & USP chapters:

- Mire-Sluis et al., 2004, JIM *(ADA screening - design elements)*
- Koren et al., 2007, JIM *(ADA testing strategy)*
- *Shankar et al., 2008, JPBA (ADA screening – method validation)*
- *Gupta et al., 2011, JPBA (NAb)*
- *USP chapter on Immunogenicity screening methods, 2013*
- *USP chapter on Neutralizing Antibody methods, 2014*

Outline

Background *(will skip this as Dr. Shankar has already covered it)*

Screening Cut-Point

Confirmatory Cut-Point

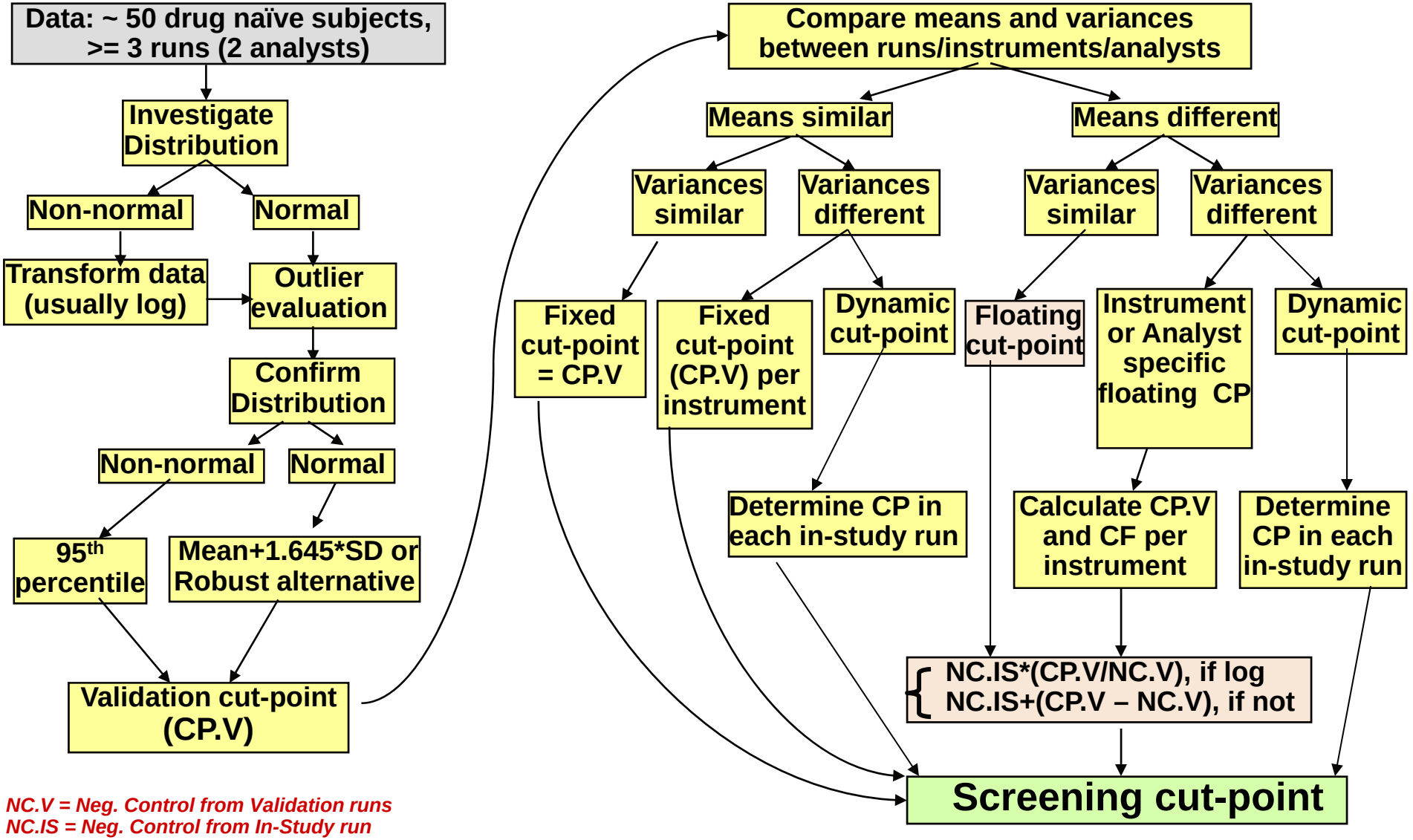
Titer Cut-Point

Some Practical Considerations

ADA Four-tiered Testing Strategy

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

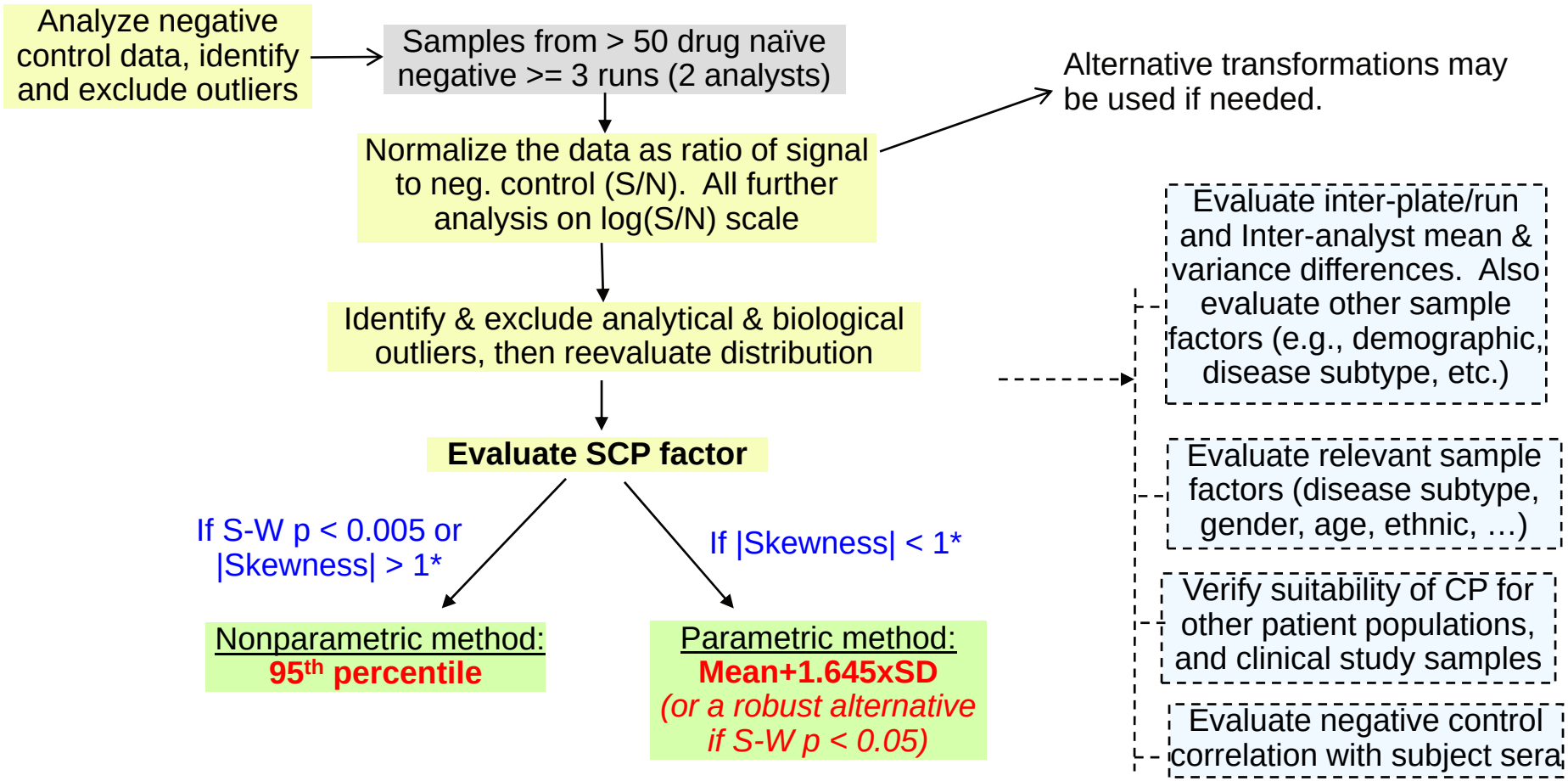
Screening CP Evaluation, *Shankar et al, 2008*



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

Screening CP Evaluation

- a simpler flow-scheme that works in most cases



** Skewness test can be used instead; Brown & Hettmansperger, JASA, 1996*

R code: Cummins & Devanarayan (1998): https://biostat-lists.wustl.edu/sympa/arc/s-news/1998-11/msg00026.html

Typical considerations for a cut-point experiment

~ 50 drug naïve ADA negative subjects

- Preferably from the target disease population if available.
- Should represent relevant demographic subgroups (gender, race, age, etc.).
- Multiple disease subtypes can be included, to investigate common or separate cut-points (e.g., n>20 for each cancer type).

Test these subject samples in 6 runs, by 2 analysts (3 runs per analyst)

- If multiple analysts will test study samples

Each sample tested in duplicate

- Reportable result: Average of duplicate samples divided by NQC.

Negative QC: 2-3 reportable results/plate, each in duplicate, and located in different parts of the plate.

Low QC and High QC: >=2 reps/plate, each in duplicate

Three plates used per run for testing these samples.

Include drug-spiked samples as well for confirmatory cut-point. Sensitivity, Precision, Confirmation CP, Titration CP, etc., can all be evaluated from this expt.

Balanced design (Shankar et al., 2008)

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 and every plate, by both analysts.

*Statistical modeling approach for **outlier** evaluation*

1. Fit a mixed-effects model on the normalized response.
   - Random effects: Subjects nested within Subject Groups, Run number nested within Analyst, and Plate ID.
   - Fixed effects: Subject Groups, Analyst, Plate testing order and the interaction of Analyst and Plate testing order (+ gender, disease types, demographics, …, 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 to identify outliers from the conditional residuals → These are ***Analytical** outliers*.
4. Iterate steps 1-3 until all analytical outliers are removed

*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. Use the "outlier box-plot" criteria to identify & exclude outliers from subject BLUPs. → These are ***Biological** outliers*.

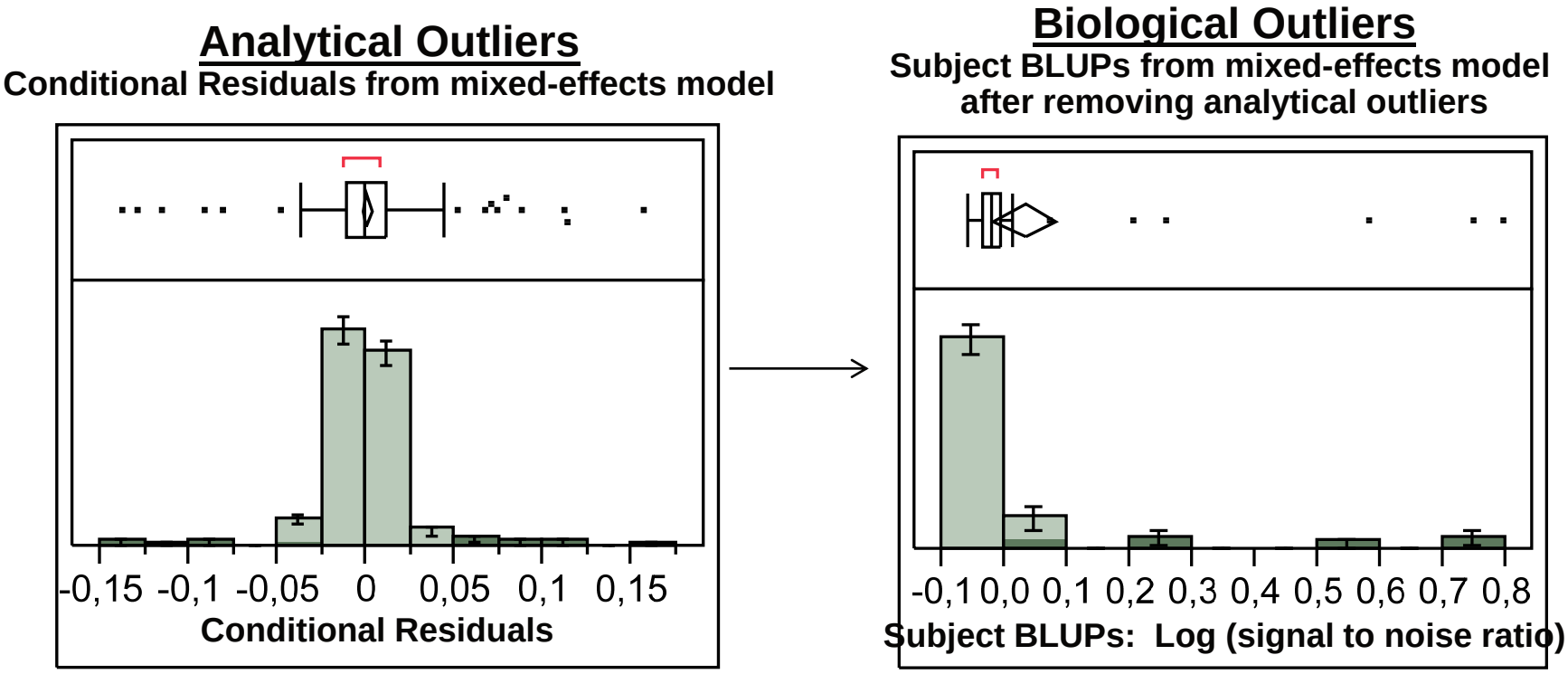
7. Iterate steps 5-6 until all biological outliers are removed.

8. Refit the model without these outliers. Obtain **marginal** residuals.
   - Difference between the observed & *mean* predicted values; *reflects both subject random effect & measurement error.*

9. Assess normality & symmetry of these marginal residuals
   – Shapiro-Wilk test & Skewness Test (Brown & Hettmansperger, JASA, 1996)

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

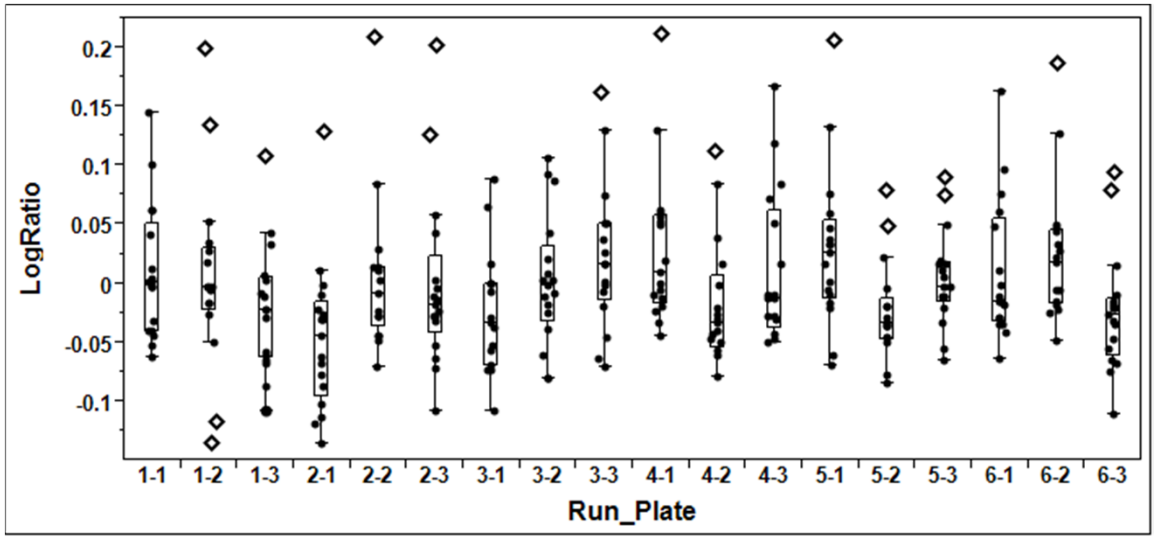
Illustration of outlier evaluation with statistical modeling approach



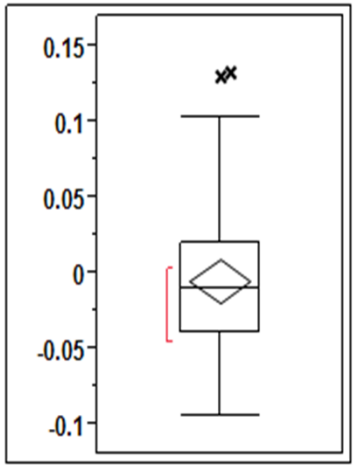
- Mixed effects model (as described earlier) is fit on the ratio of individual samples to the negative control. Conditional Residuals are evaluated.
- Analytical outliers are first identified and excluded. After iteratively excluding all the analytical outliers, the biological outliers are evaluated.

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

Analytical Outliers



Biological Outliers



Subject averaged data

- Identify & exclude analytical outliers (AO) from each assay run/plate separately. Iterate until no more AO.
- Then identify & exclude biological outliers (BO) by evaluating the distribution of subject averaged data. Iterate until no more BO.
- Then verify 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).*

SD evaluation for screening cut-point

SD should incorporate all variance components (total variation) expected during sample testing. Can be calculated easily using Excel.

After excluding all the outliers, calculate SD of all the data from the validation experiment.

- Suppose 50 subjects were tested in 6 runs (300 samples).
- If there were 20 outlier samples, then calculate SD of the remaining 280 samples directly from Excel using STDEV function.

In most cases, this is *quite similar to the more rigorous calculation of variance components from random-effects ANOVA*.

- Random-effects ANOVA helps understand the relative contribution of different variance components.

Use MAD instead of SD and Median instead of Mean, if tails are long after outlier exclusion (see flow-scheme).

Confirmatory cut-point (CCP) – Evolution

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

Shankar et al (2008) published an experimental approach to evaluate CCP based on biological & analytical variability.

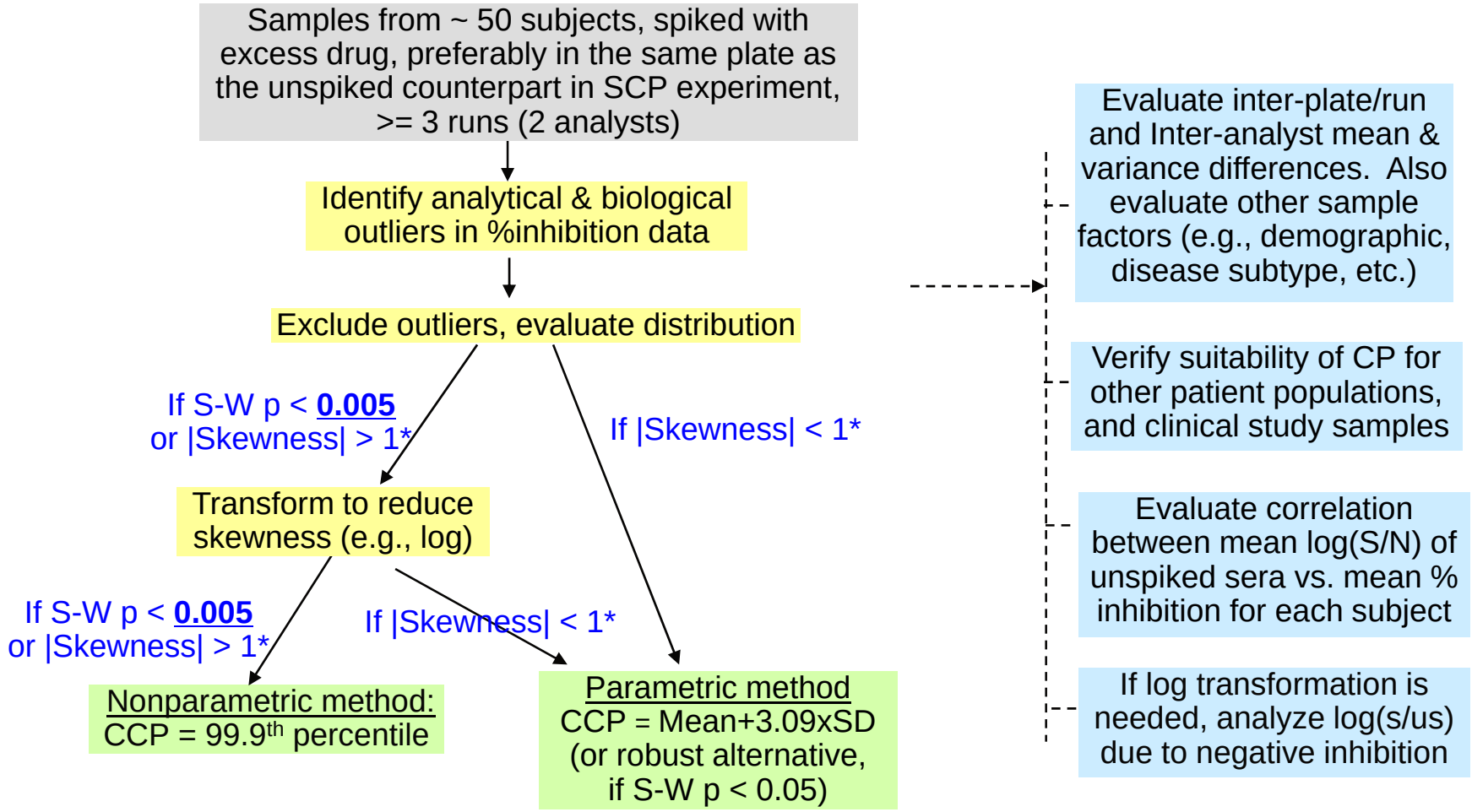
Alternative approaches were discussed for several months by a focus group of industry and FDA scientists.

- E.g., using individual mock low-positive samples to define a lower 99.9% limit as the CCP.

Conclusions from this discussion (Smith et al, 2011):

- Shankar et al (2008) method can be the default, as it works well in most cases.
- Alternative approaches may be tried in some special cases.

Confirmatory Cut Point (CCP) Evaluation
- a simple flow-scheme that works in most cases



** Skewness test can be used instead; Brown & Hettmansperger, JASA, 1996*

R code by Cummins & Devanarayan (1998): https://biostat-lists.wustl.edu/sympa/arc/s-news/1998-11/msg00026.html

Titration cut-point (TCP)

Screening cut-point may fall on the lower plateau of the positive control dilution curve.

- This will result in highly noisy/variable titers.

In such cases, use a higher cut-point for evaluating titers (Titration CP)

- Using the *same data from the screening cut-point experiment*, calculate TCP = Mean + 3.09xSD or Mean + 6xSD
- 3.09 corresponds to ~ 99.9th percentile.

Titer = MRD for confirmed positives that fall between SCP & TCP.

Other methods based on only negative control may be considered.

- NB: Purpose of Titer CP is fairly simple, so extensive research and additional calculations don't add much value.

Some Practical Considerations

Validity of Negative Control for SCP factor evaluation

Floating SCP or SCP 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 mean of subject sera from each run/plate.
  - Evaluate Slope & Rank Corr. (Need Slope ~ 1, Correlation > 70%).

If this assumption fails, using NC for Floating-CP may not be helpful.

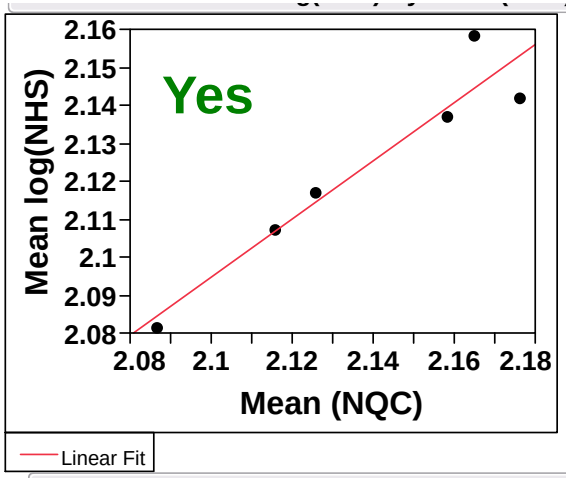
- More likely when analytical variability exceeds biological variability.

Alternatives: New pool, Subject-specific cut-point, other controls (same disease/demographic), ….

Need 3 reportable results of NC, located in different parts of the plate.

Testing the validity of Negative Control (contd.)

Example-1



Linear Fit

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

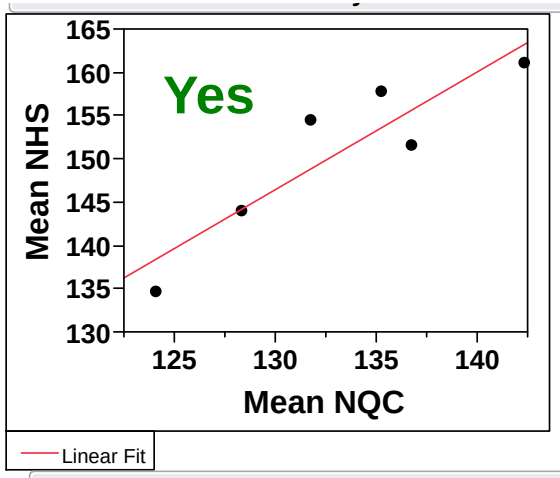
Summary of Fit

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

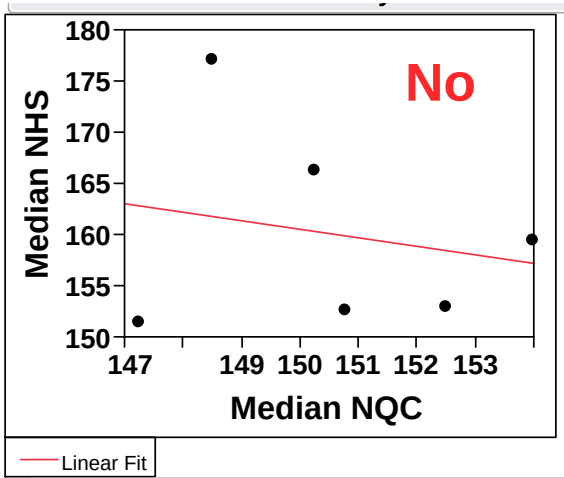
Summary of Fit

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

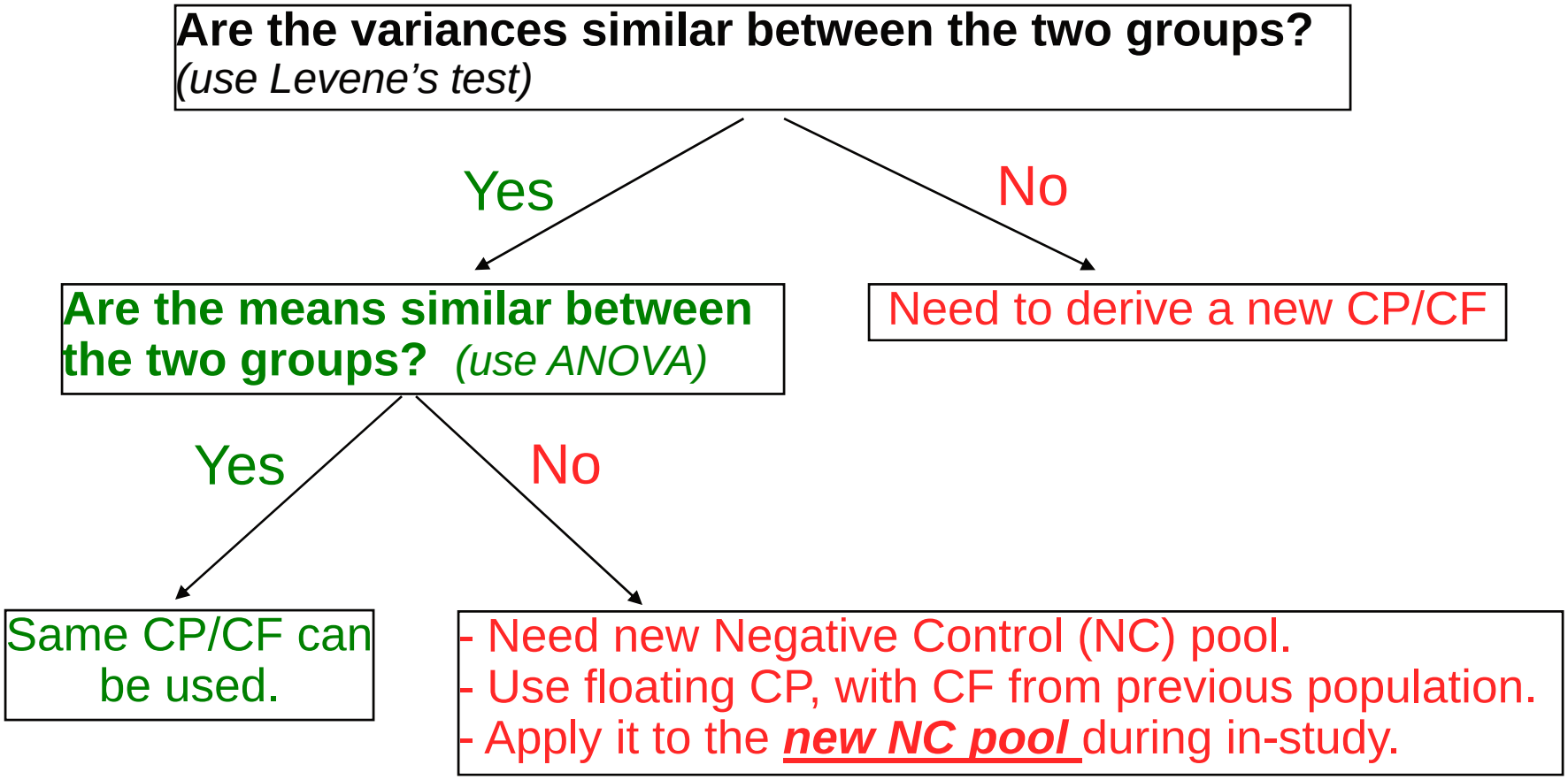
Summary of Fit

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

Can the same CP or CF be used for a different disease/target population?



In-study justification of validation cut-points
Why & When?

Why?

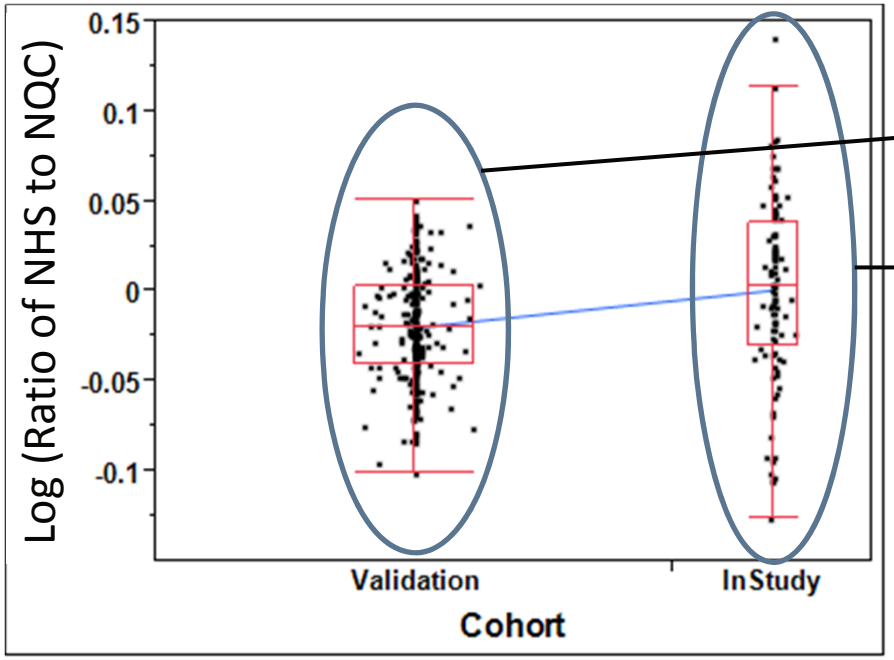
- SCP factor during validation is set to yield ~ 5% false positives. But this is an estimate. As with any estimate, there is variability.
- Based on our Monte-Carlo simulations assuming a typical dataset from a balanced design of Shankar et al (2008), *false positive rate* can vary between 3 to 14% during the in-study sample testing phase.

When?

- Therefore as a general guideline, *>= 15% false positives in clinical baseline samples should trigger follow-up evaluations*.
- Also necessary if the validation and in-study populations are different (e.g., healthy vs. disease, differences in disease & demog.)

In-study justification of validation cut-points (contd.)

Why and When? (example)



- SCP factor from pre-study validation is 1.15.
- Variability of in-study clinical baseline data is significantly higher, $p<0.0001$ (Levene's test).
- Results in > 30% false positives in clinical baseline samples.
- Follow-up evaluation using in-study clinical baseline data is needed.

How to justify pre-study validation cut-points with in-study data, and what are the corrective actions?

In-study justification of validation cut-points (contd.)

How?

Compare the distribution (mean and variance) of validation data versus in-study clinical baseline data.

If the variances are different (Levene's test), need to *derive a new cut-point using the in-study clinical baseline data*.

- However, if clinical baseline data are small (n < 25 subjects from < 2 runs), redo the validation experiment with subjects similar to clinical study population.

If only the means are different, same validation cut-point can be used after redefining the negative control based on disease population (for floating CP).

Apply similar criteria to decide whether to reevaluate the Confirmatory CP using the drug spiked baseline samples.

Summary

- As SCP is usually “Floating”, normalize by NQC, evaluate SCP factor.
  - Assumes NQC drifts with subject sera. This should be verified.
- Multiple populations can be tested in the same balanced design.
- Outlier evaluations are critical (use mixed-effects ANOVA or simpler alternatives).
- SD evaluation should include relevant variance components (inter-analyst, inter-run, intra-run, inter-subject, etc.)
- Titration CP can be defined at higher limits using Screening CP data.
- Cut-points from validation should be justified with in-study clinical baseline data.
- Also, should justify use in other sample types (disease, demog, etc.).