New Paradigm for ADA Cut Point Determination

Charles Tan, PhD

Non Clinical Statistics Conference

October 4, 2018

Early Clinical Development

Pfizer WORLDWIDE RESEARCH & DEVELOPMENT

SCIENCE FOR LIFE-CHANGING IMPACT

Outlines

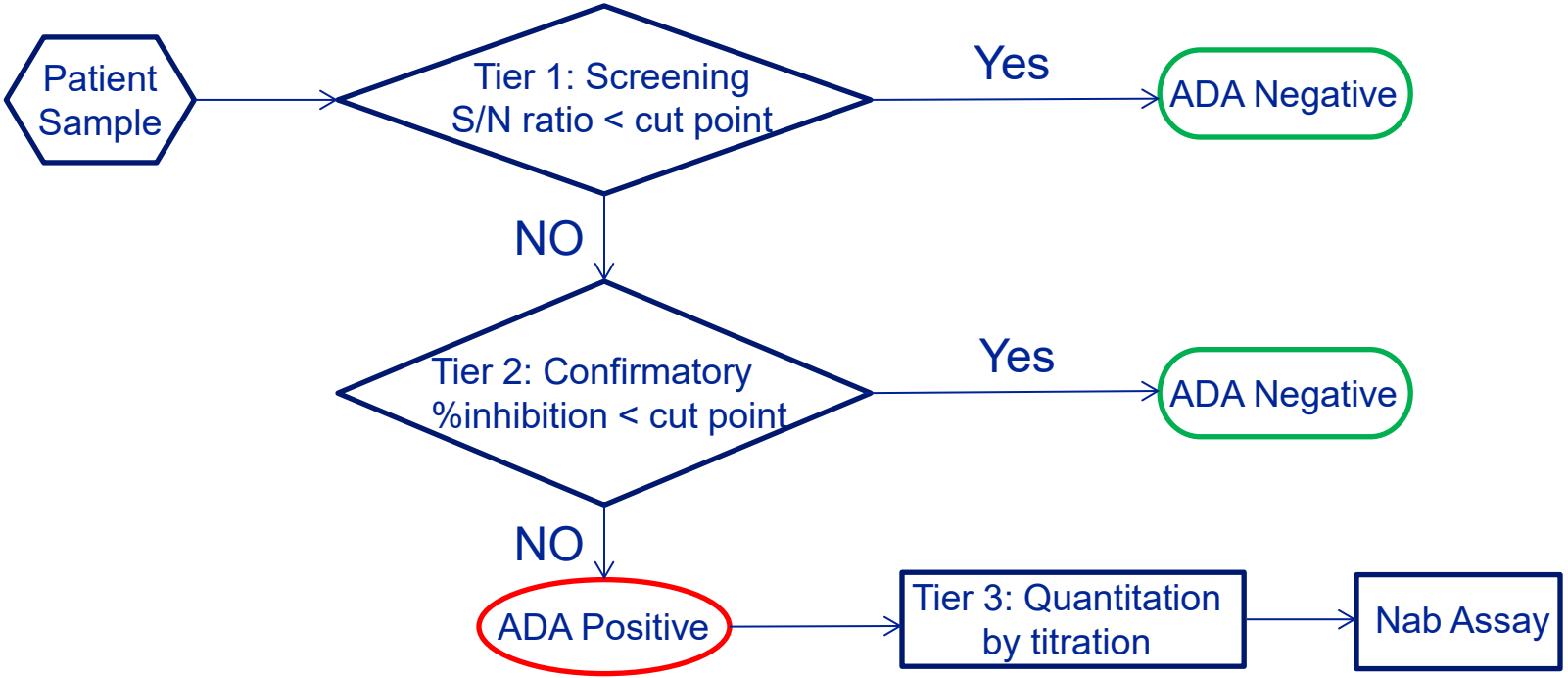
- Introduction
- Proposed new paradigm
- Examples
  - Identify putative positives (pre-existing reactivity)
  - Derive and assess cut points
- Summary
- Q&A

ADA: what and why

- Anti-Drug Antibody (ADA) response
- Most biological products elicit some level of ADA response
  - Size of the molecule, route of admin, foreignness
- A major safety and efficacy concern for regulatory agencies, drug manufacturers, clinicians, and patients

- Determination of assay cut points is not standardized within Pfizer or industry
- Assay cut points and resulting immunogenicity rates have come under increased regulatory and internal Pfizer scrutiny

ADA: tiered approach



Where the industry is at

- FDA guidance:
  - Tier 1 Screening cut point: 5% false positive rate
  - Tier 2 Confirmatory cut point: 1% false positive rate
- Typical stat approach:
  - Look at tier 1 and 2 data separately
  - Removal of outliers: box plot
  - Parametric methods: normality test, transformation
- Central scientific/clinical challenges:
  - Heterogeneity of the naive populations
  - Clinical relevant ADA level

Proposal: Tier 1/Tier 2 cut point determination

- **Step 1: data preparation**
  - Remove analytical failures: exclude the pair of wells with CV > 20%
  - Remove putative positives (pre-existing antibodies in naïve samples): exclude the samples with S/N ratio > threshold_1 AND %inhibition > threshold_2
    - The thresholds would be adjusted by each assay's characteristic relationship between S/N ratio and %inhibition
- **Step 2: calculate cut points**
  - Nonparametric (distribution-free) point estimator
  - Target 95 percentile for Tier 1 (S/N ratio), and 99 percentile for Tier 2 (%inhibition)
- **Step 3: evaluate the cut points**
  - A graphic method to compare the cut points from Step 2 to
  - "Minimum" cut point implied by the precision of the assay, and
  - "Maximum" cut point implied by the positive controls at 100 ng/mL
- **Step 4: done or back to Step 2 with alternatives**
  - If the calculated cut point is closer to the "minimum" cut point: DONE
  - If the calculated cut point is too close to the "maximum" cut point: Go back to Step 2 with alternative approaches
    - Alternative 1: use lower confidence bound, instead of point estimator
    - Alternative 2: seek statistician's direct inputs

FDA Draft Guidance April 2016

C. Sensitivity

1. Assay Sensitivity

The sponsor should determine the sensitivity of the assay to have confidence when reporting immunogenicity rates. Assay sensitivity represents the lowest concentration at which the antibody preparation consistently produces either a positive result or readout equal to the cut point determined for that particular assay.[14] FDA recommends that screening and confirmatory ADA assays achieve a sensitivity of at least 100 nanograms per milliliter (ng/mL). Although traditionally FDA has recommended sensitivity of at least 250–500 ng/mL, recent data suggest that concentrations as low as 100 ng/mL may be associated with clinical events (Plotkin 2010; Zhou, Hoofring, et al. 2013). However, it is understood that neutralization assays may not always achieve that level of sensitivity.

Three Innovations

- Identify putative positives (pre-existing reactivity) among naïve populations (in Step 1)
  - Make exclusion/inclusion decision at sample level
  - Make sample exclusion/inclusion using Tier 1 ratio and Tier 2 %inhibition simultaneously
  - Do NOT use boxplot to exclude/include individual measurement one tier at a time.
- Nonparametric methods (Step 2)
  - Can handle mixture distributions
  - Do NOT rely on distributional assumptions
- Graphic assessment of cut point decisions (Step 3)
  - Focus the decisions on scientific / regulatory merit
  - Visualize the decisions' context
  - Move the decisions away from driving the assay to its limits

Programs selected for testing proposal

- The ADA cut point has been calculated using the new approach for the Tier 1 and the Tier 2 ADA validation data of the following programs:
  - Example 1: mAb; very clean ADA data set
  - Example 2: ADC with pre-existing Abs
  - Example 3: a Recombinant Mimetic of Pooled Human IVIG; lots of pre-existing antibodies

Identify putative positives

- Putative positives: naïve samples who behave the same way that real ADA positive samples would have behaved
  - Tier 1: high ratio, AND
  - Tier 2: high %inhibition
- They have to be removed
  - Otherwise, they inflate the cut points, regardless statistical approaches
- How to identify them: how high is “too high”
  - Criterion based on each assay’s unique characteristic

Assay Characteristic Curve

- N: the Negative Control response on each plate
- S_1: the binding response (tier 1 raw data)
- S_2: the inhibited response (tier 2 raw data)

- Tier 1: Ratio = S_1 / N
- Tier 2: %inhibition = $(1 – S_2/S_1)*100\% = (1 – S_2/N / S_1/N)*100\%$

- Assay Characteristic Constant: h = median of (S_2 / N)
  - A feature of the assay, not individual samples
  - Example 1: h = 0.9532
  - Example 2: h = 0.9152
  - Example 3: h = 0.6257

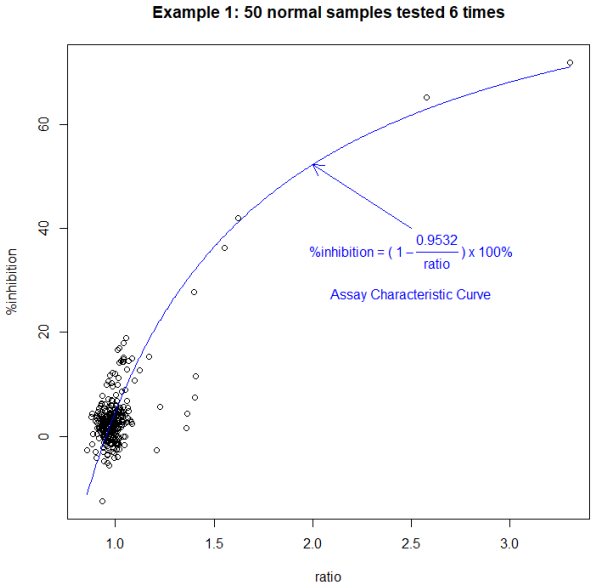
- Assay Characteristic Curve: %inhibition = (1 – h/ratio) *100%
  - Example 1: %inhibition = (1 – 0.9532/ratio) *100%
  - Example 2: %inhibition = (1 – 0.9152/ratio) *100%
  - Example 3: %inhibition = (1 – 0.6257/ratio) *100%

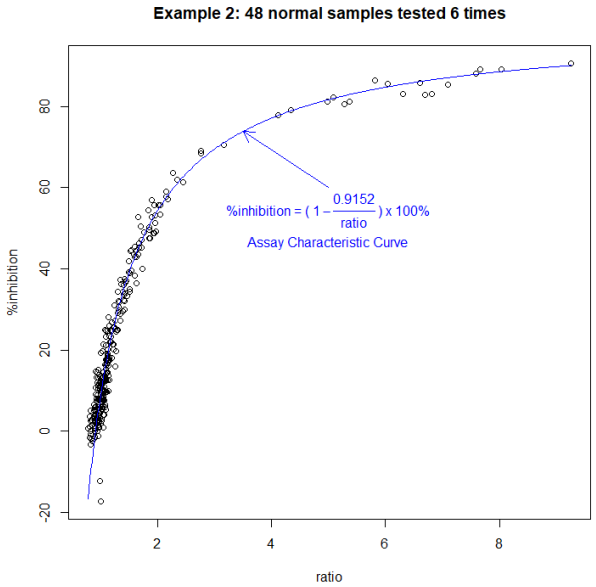
An actual plate

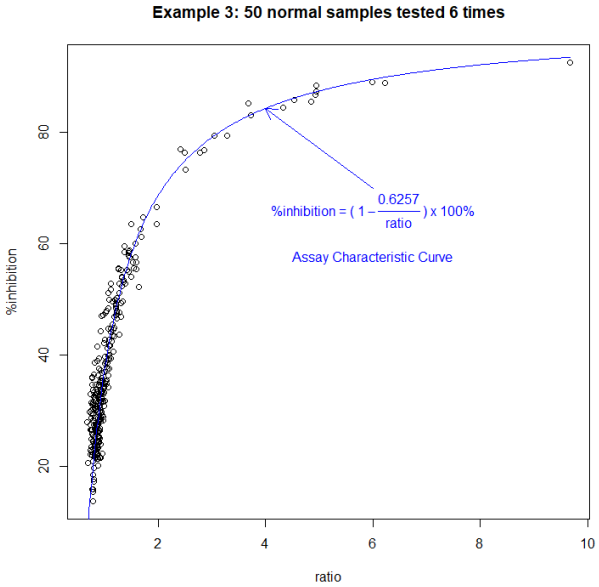
	PC Titer												Sample Treatment
	1	2	3	4	5	6	7	8	9	10	11	12	
A	4415	4311	147	145	142	145	145	148	142	144	152	151	No INH
B	1636	1656	145	144	139	137	144	145	138	137	148	148	INH
C	611	627	146	142	137	137	152	157	144	142	148	147	No INH
D	297	296	146	146	136	133	146	146	143	138	148	147	INH
E	188	186	2929	2965	135	141	143	139	142	138	3489	3470	No INH
F	157	148	135	132	134	135	139	133	134	137	146	141	INH
G	141	136	261	266	139	139	149	143	141	144	277	271	No INH
H	139	129	131	130	134	133	136	136	137	137	136	136	INH

	PC Titer												Sample Treatment
	1	2	3	4	5	6	7	8	9	10	11	12	
A	4363		145		144		147		143		152		No INH
B	1646				138		145		138		148		INH
C	619				137		155		143		148		No INH
D	296.5				135		146		141		148		INH
E	187		2947		138		141		140		3480		No INH
F	152.5		134		135		136		136		144		INH
G	138.5		264		139		146		143		274		No INH
H	134		131		134		136		137		136		INH

Examples of Assay Characteristic Curves







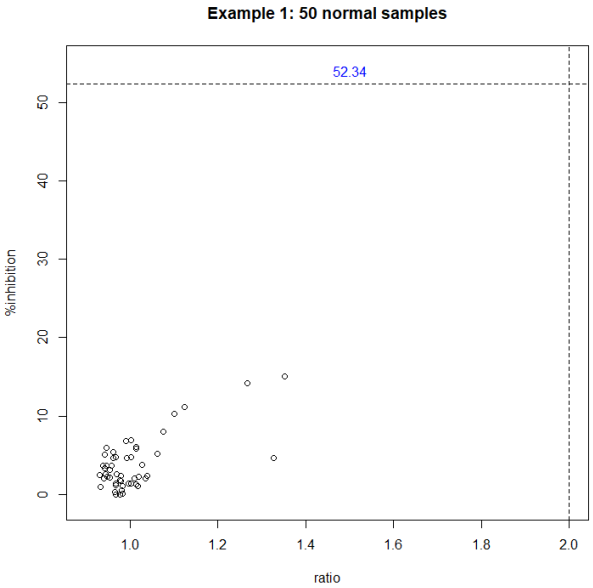
Tier 1 ratio and Tier 2 %inhibition are intrinsically related. This relationship is well described by the Assay Characteristic Curve with a single constant. This parameter is a feature of the assay, not individual samples.

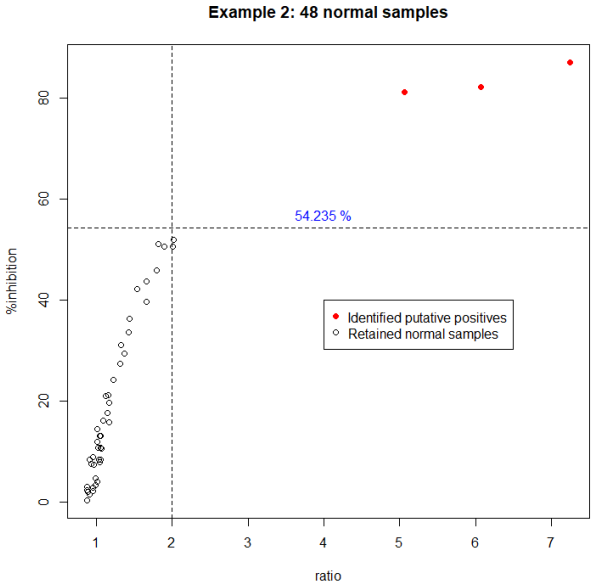
Putative Positives Criteria

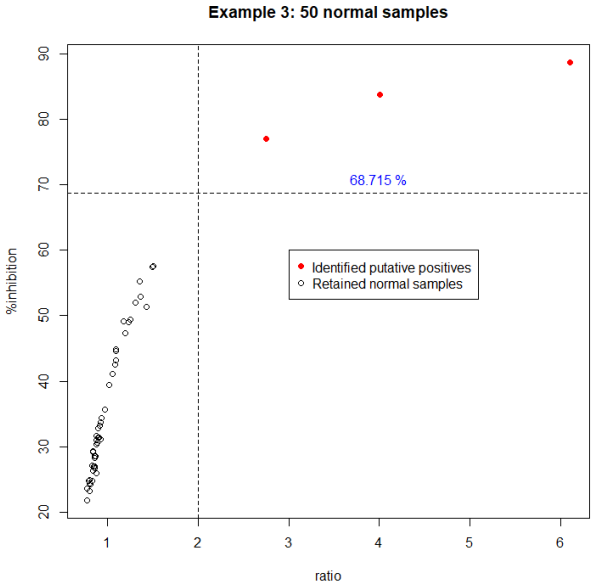
- A sample is removed if
  - The mean ratio > 2 and
  - The mean %inhibition > (1 – h/2) *100%
    - Example 1: (1-0.9532/2)*100%=52.3%
    - Example 2: (1-0.9152/2)*100%=54.2%
    - Example 3: (1-0.6257/2)*100%=68.7%

- Justification
  - High ratio and high %inhibition: indistinguishable from positives
  - Ideal case h = 1: %inhibition = (1 – 1/ratio) *100%
    - ratio = 2 ⇔ %inhibition = 50%

Putative positives: Ex1 (0), Ex2 (3), Ex3 (3)





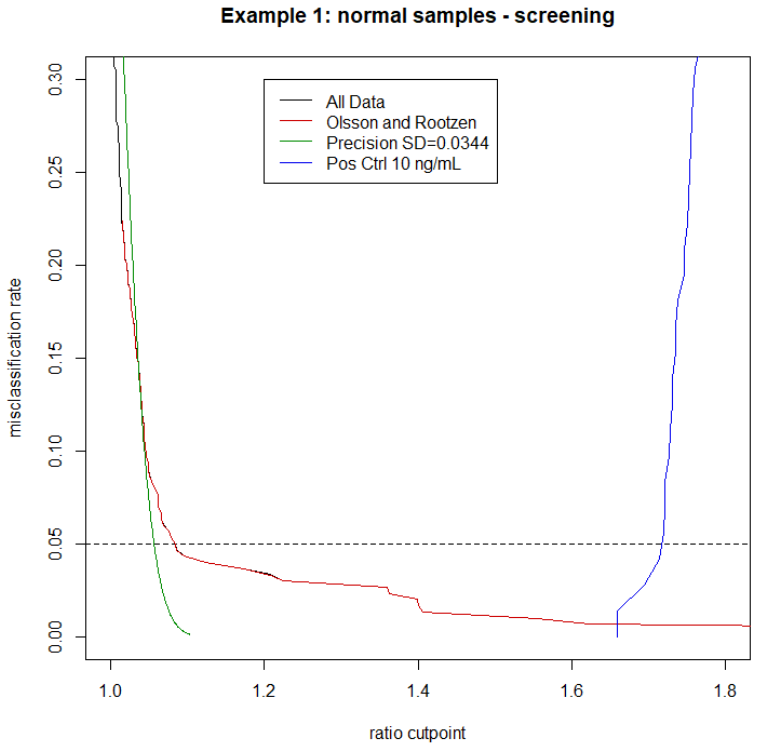


Each dot is a sample: sample level mean %inhibition versus mean ratio

Nonparametric methods for cut point calculation

- Our preferred method for nonparametric cut point calculation is the point estimator of percentiles per Olsson and Rootzen (1996)
  - Recognize the fact that 50 samples were tested 6 times each, not 300 independent samples
- If there are push backs, we can propose alternative cut points based on the lower confidence bound of the percentiles as recommended in the draft FDA guidance

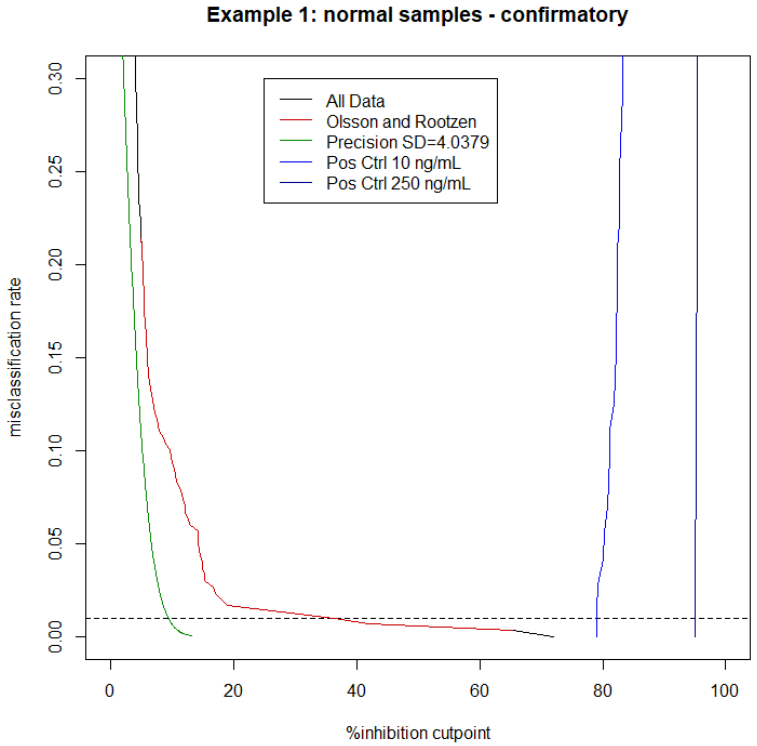
Example 1 Tier 1 cut point at 5% FPR



Cut point (S_1/N) based on proposal: 1.084
Cut point (S_1/N) reported by CRO: 1.05

Conclusion: Ex1 Tier 1 ratio data are very clean. The proposed cut point is further away from minimum cut point based on measurement precision alone than other approaches. Yet the proposed cut point is well below the maximum cut point suggested by the 10 ng/mL positive control.

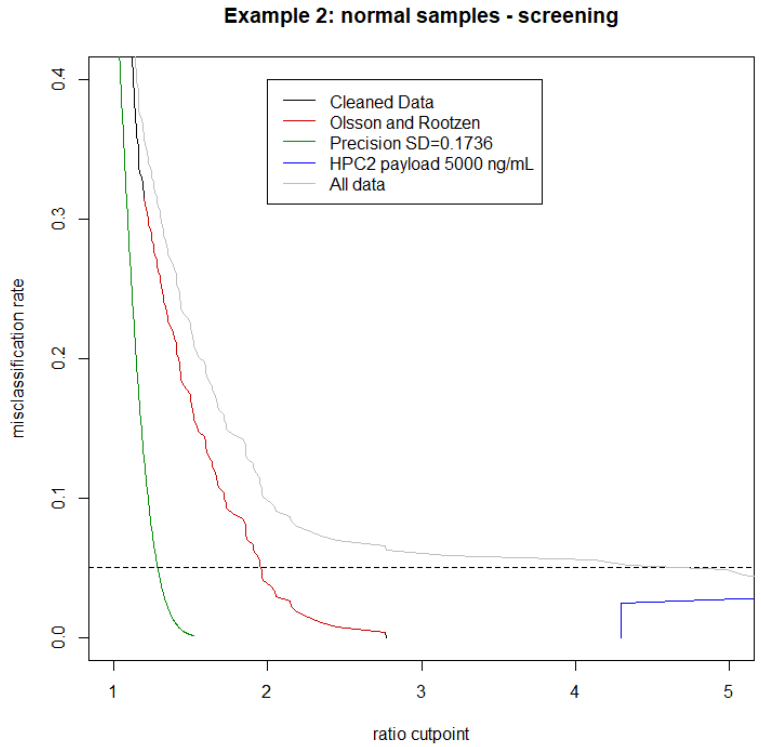
Example 1 Tier 2 cut point at 1% FPR



%inhibition cut point based on proposal: 36.4%
%inhibition cut point reported by CRO: 8.02%

Conclusion: Ex1 Tier 2 data have a heavier tail than common parametric models suggest. Our nonparametric method produces a cut point significant higher than cut points produced by other approaches, yet is still well below the maximum cut point suggested by the 10 ng/mL positive control.

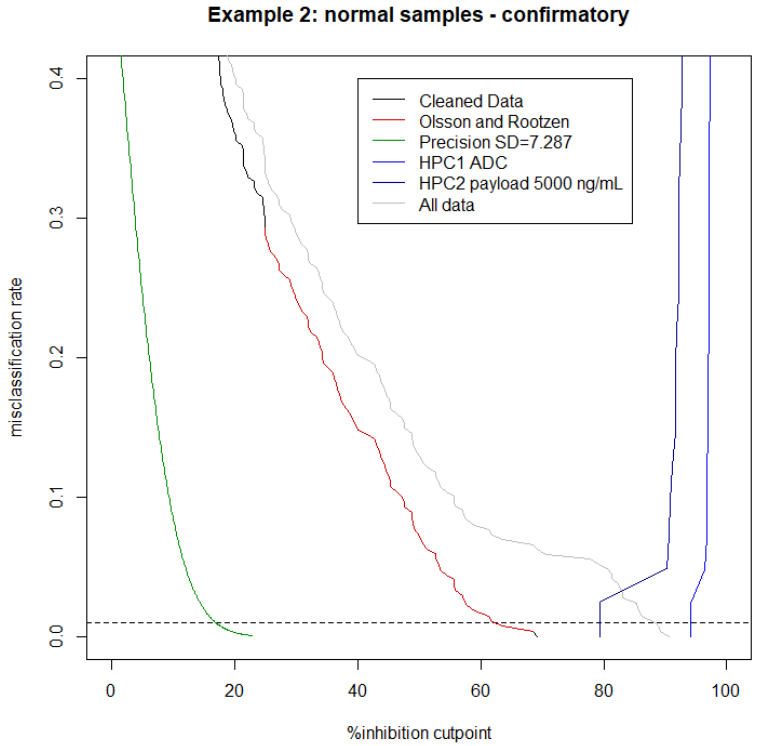
Example 2 Tier 1 cut point at 5% FPR



Cut point (S_1/N) based on proposal: 1.95
Cut point (S_1/N) reported by CRO: 1.42

Conclusion: Removal of putative positives made a significant impact. Our proposed cut point is further away from minimum cut point based on measurement precision alone than other approaches. Yet the proposed cut point is well below the maximum cut point suggested by the payload positive control.

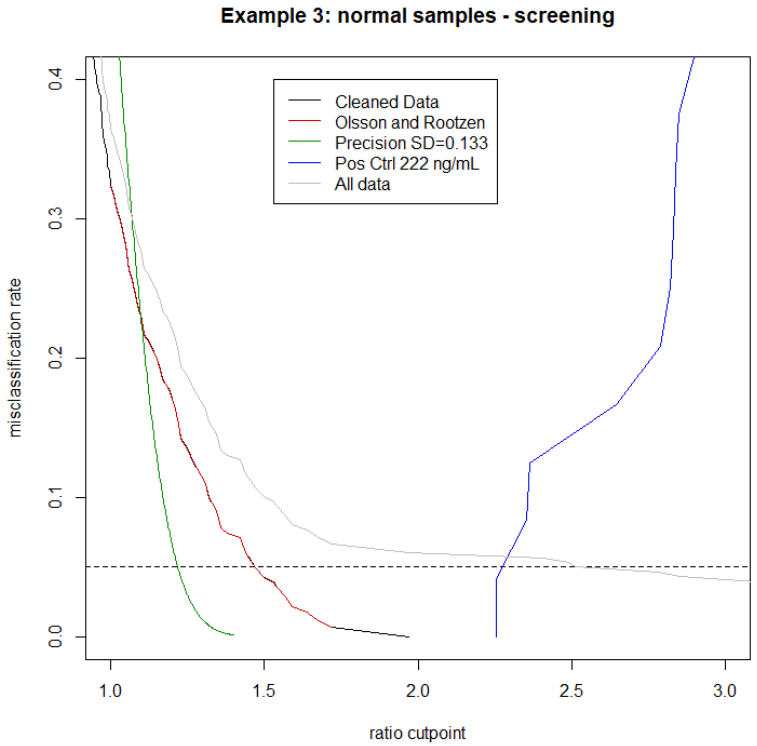
Example 2 Tier 2 cut point at 1% FPR



%inhibition cut point based on proposal: 62.4%
%inhibition cut point reported by CRO: 48.8%

Conclusion: Removal of putative positives made a significant impact. However, the cut point is still high due to the nature of the data. Our proposed cut point is still below the maximum cut point suggested by the positive control.

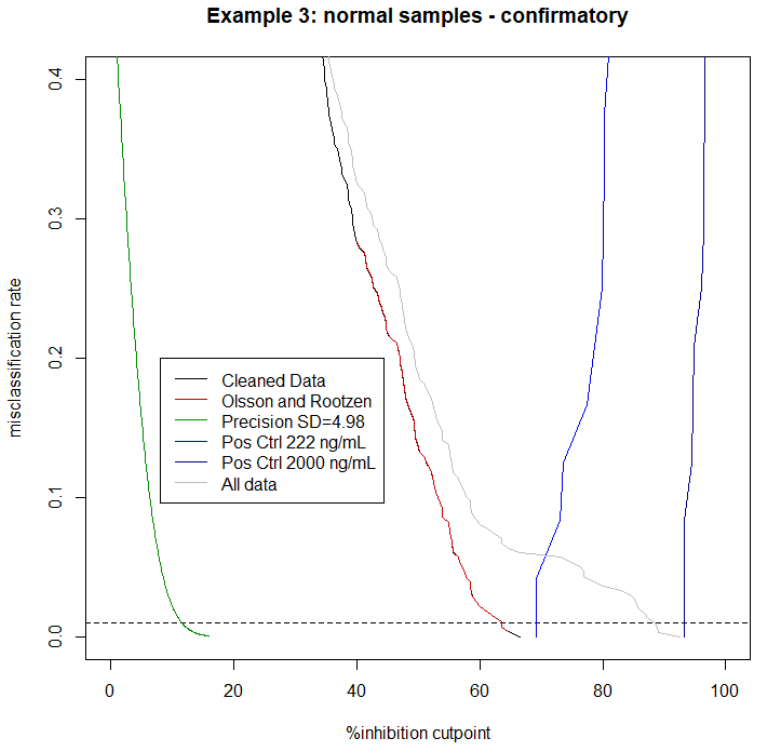
Example 3 Tier 1 cut point at 5% FPR



Cut point (S1/N) based on proposal: 1.47
Cut point (S1/N) reported by CRO: 1.45

Conclusion: Removal of putative positives made a significant impact. Our proposed cut point is more away from minimum cut point based on measurement precision alone than other approaches. Yet the proposed cut point is well below the maximum cut point suggested by the 222 ng/mL positive control.

Example 3 Tier 2 cut point at 1% FPR



%inhibition cut point based on proposal: 63.5%
%inhibition cut point reported by CRO: 63.6%

Conclusion: Removal of putative positives made a significant impact. Our proposed cut point is further away from minimum cut point based on measurement precision alone than other approaches. Yet the proposed cut point is still below the maximum cut point suggested by the 222 ng/mL positive control.

Summary

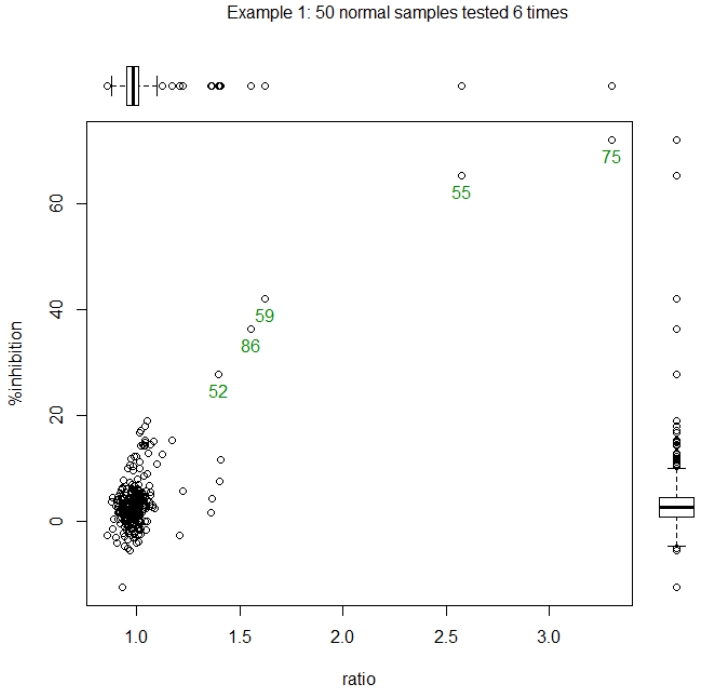
- Proposal compares favorably to
  - Current and common industry practices
  - Alternative methods
- Proposal is
  - Sensible, scientific, and innovative
  - Can be defended on scientific, regulatory, and statistical grounds
  - Allows regulatory discussions to be grounded on practical constrains and scientific merit
- Received positive feedbacks from sponsors
- Considering certain degree of novelty of proposed approach, we are recommending to reach out to regulators in, e.g., a Type C meeting format and/or workshop at FDA

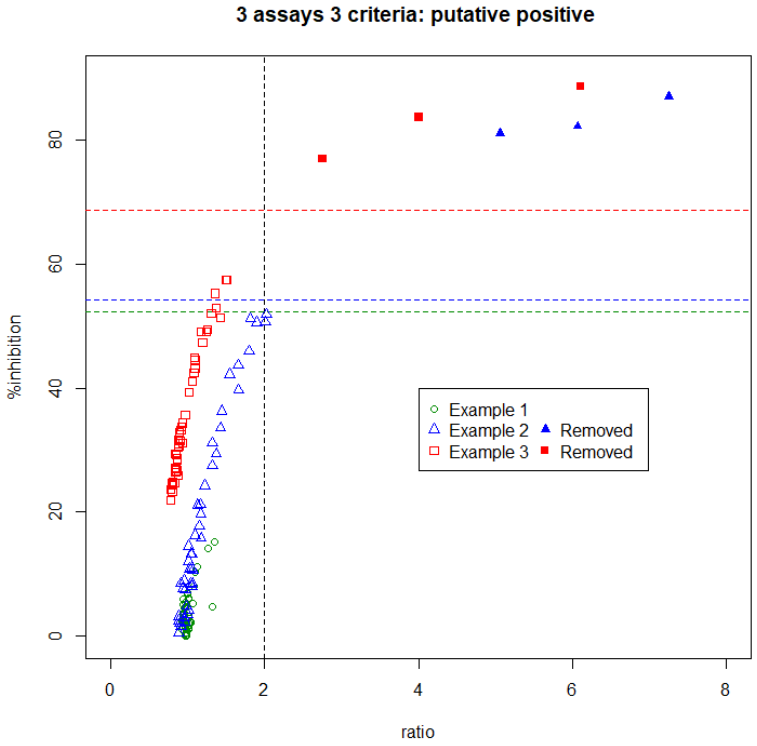
Backups

Why NOT box plot

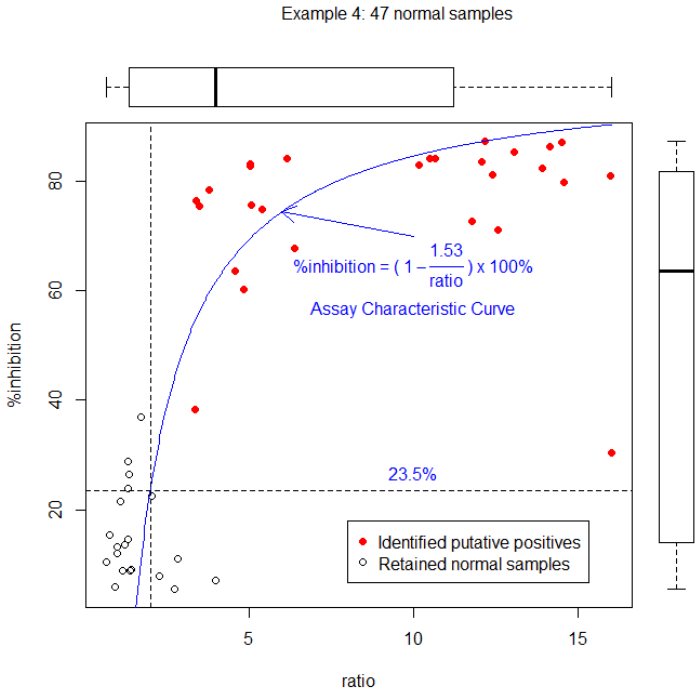
- Box plot was originally developed in the context of exploratory data analysis
  - Graphic tool to display features of data distribution
  - Thick tailed or skewed distributions are supposed to have more data points outside the fence
  - It is well recognized in the original work that data outside the fence are not necessarily “real” outliers
  - It labels potential outliers, but is not a formal outlier detection procedure
  - The main culprit behind “cut points were too tight”
- Our reality: some of the “potential outliers” are the most informative data points in deciding cut points
  - The cut point decision is all about tail behavior
  - The naïve population often is a mixture
  - The putative positive criteria: more science based
    - Look at tier 1 and 2 data together: more coherent biological argument
    - Make sample level exclusion decision: preserve measurement variability

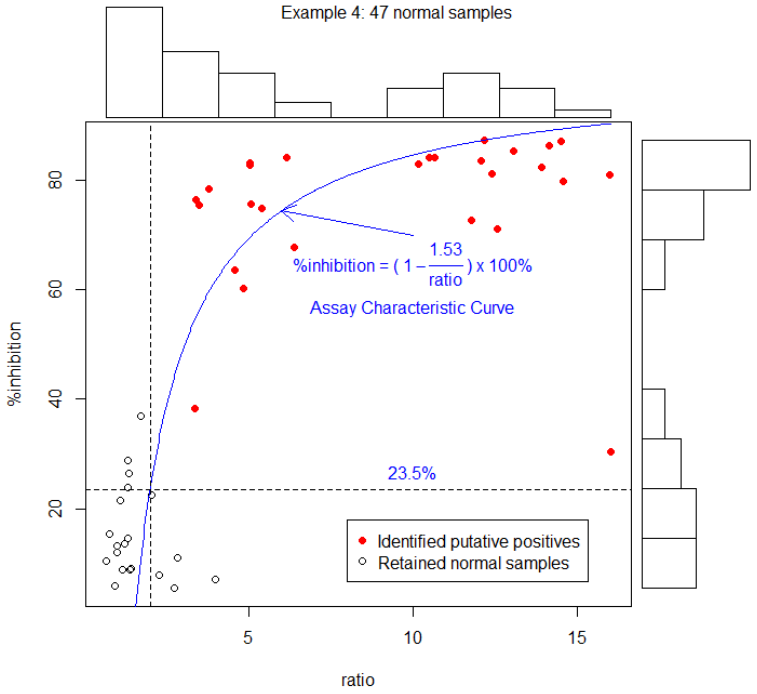
Box plot: clean too much





Box plot: fail to clean

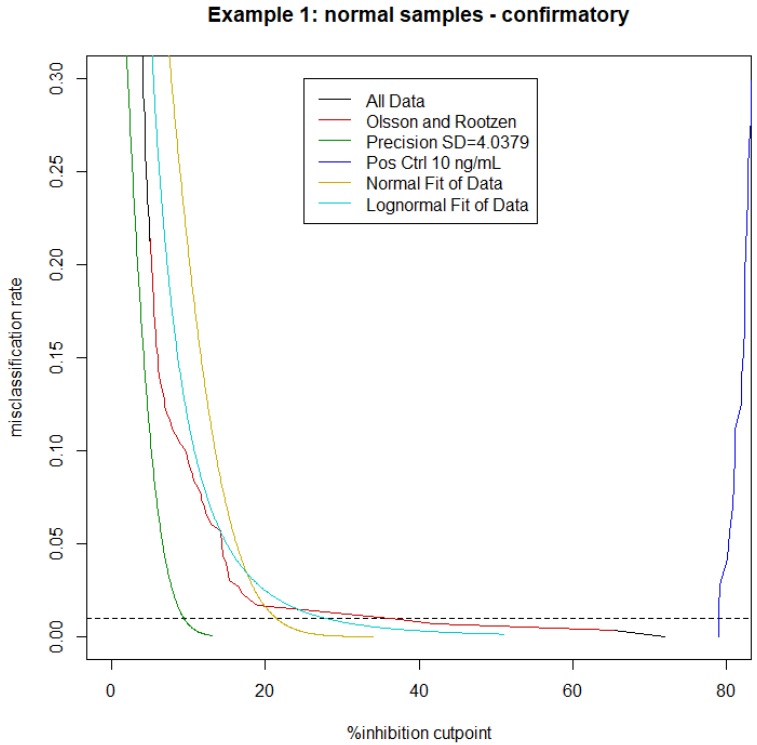




Why NOT parametric methods

- Biological reality: the naïve population often is a mixture
  - The middle doesn't necessarily predict the tail
  - The cut point decisions are about the tail behavior
- Need to identify which parametric distribution
  - Stat tests don't actually confirm distribution, only "fail to reject"
  - Often times, both Normal and Lognormal are rejected
    - Transformation is needed to "normalize" the data
    - The best transformation is very much dataset dependent, loss of easy interpretability
    - The usual process to identify which parametric distribution can quickly become a "rabbit hole":
      - If you chase all the way down, you'd lost in statistical weeds
      - If you curtail the chase, you could end up in bizarre situations
- Nonparametric methods are distribution free
  - Designed to work regardless the underlying distribution
  - Work for mixture distributions

Parametric methods on Example 1 Tier 2



- Compared to the raw data and nonparametric fit
  - Normal fit seriously underestimates the tail
  - (Aggressive) Lognormal fit still underestimates the tail