

# Statistical Recommendations on Immunogenicity Method Performance Parameters

Viswanath “Devan” Devanarayan, Ph.D.  
Merck & Company

*viswanath\_devanarayan@merck.com*

7<sup>th</sup> Annual IIR Immunogenicity for BioTherapeutics  
San Francisco, May 23-24, 2006

# Acknowledgements

## AAPS Immunogenicity White-Paper Groups on Ligand Binding & NAb assays:

- **Authors of** “Recommendations for the Design of Immunoassays used in the Detection of Antibodies against Biological Products”, *Journal of Immunological Methods*, Vol. 289 (1-2), pages 1-16, 2004
- **plus authors of the ongoing white papers on immunoassay validation, NAb validation and immunogenicity testing strategy.**
  - *These white paper efforts are driven by the Ligand Binding Focus Group of the Biotec Section of the AAPS.*

## Thanks also to:

- **Dr. Wendell Smith, B2S Consulting**
- **Dr. Ronald Bowsher, Linco Diagnostics & B2S**

# Immunogenicity Methods

## Strategy and Method Characterization

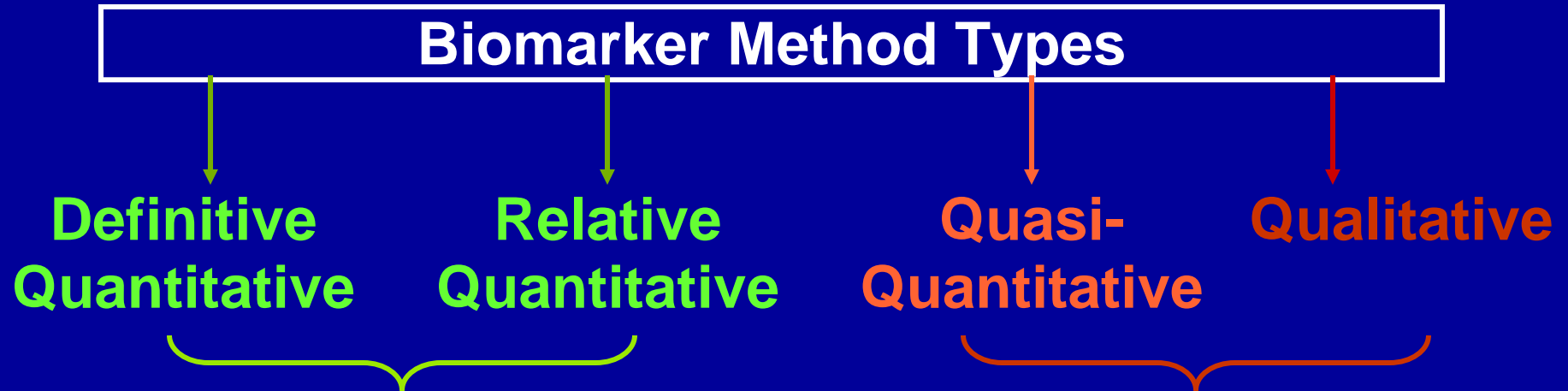
1. Screening (Identification of Positive Samples)
  - Minimum Required Dilution, Cut Point, Sensitivity, ...
2. Study-Drug Specificity & Drug Interference
3. Quantitation of “Confirmed” Positive Samples
  - Titer estimation

Precision

Selection of positive controls

# Types of Biomarker Methods

(Lee et al., *Fit-for-Purpose Method Development & Validation for Successful Biomarker Measurement, Pharm. Research, 2006*)



**Biomarker levels are calibrated from a “reference standard”.**

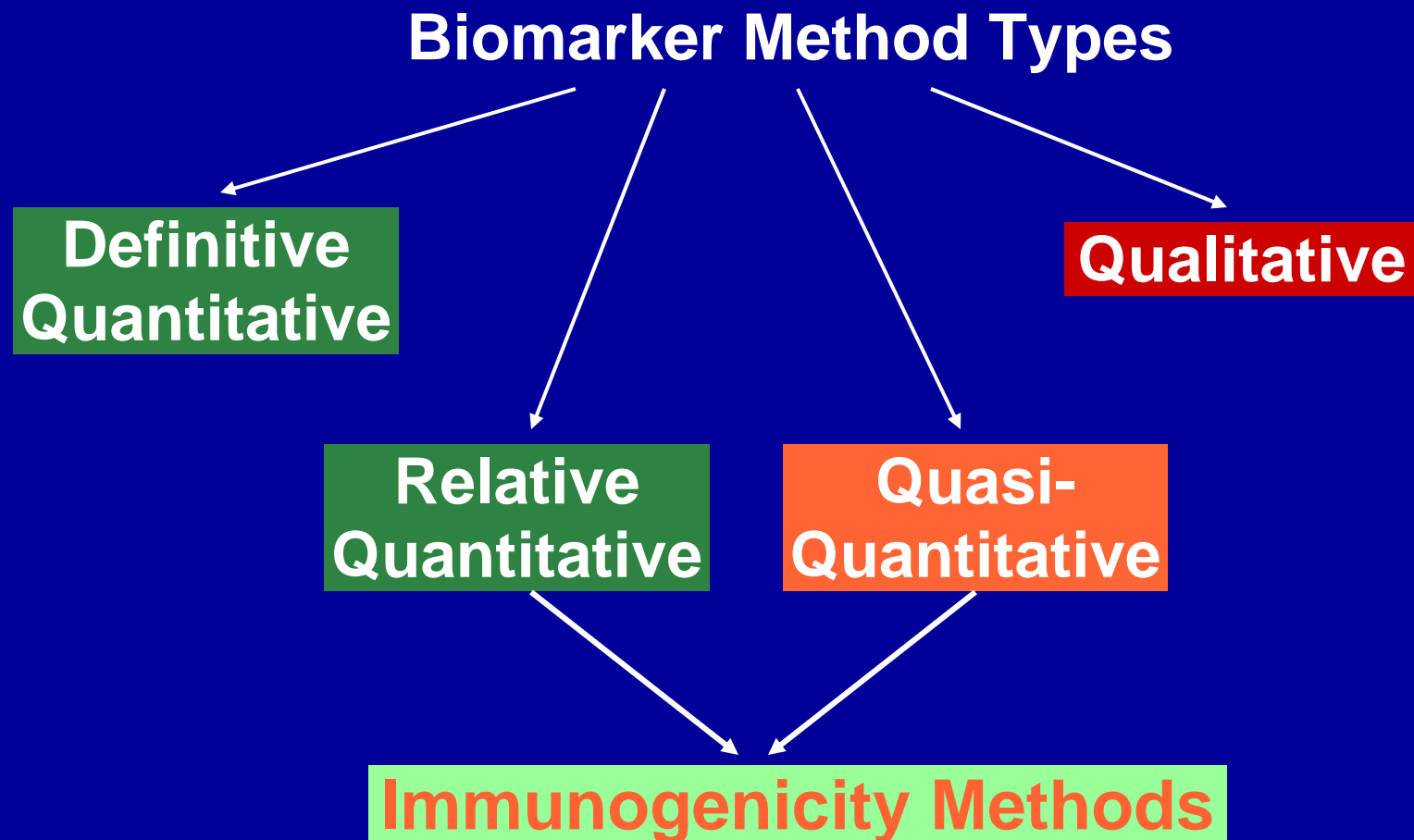
- Reference standard is representative of test samples.
  - Parallelism
- Fit-for-purpose biomarker assay white paper
  - *Lee, et al (2006), Pharm Research*

**Biomarker levels represented only by assay signal.**

- Reference material not available or
- Not representative of test samples.
  - Non-Parallelism

# Immunogenicity Methods

## Relative or Quasi Quantitative???



# Immunogenicity Methods

## Relative Quantitative: Calibration Approach

Reference standards (purified Ab preparation) used.

Samples are identified as +/- based on the "calibration cut-point" (using calibration curve).

Positives are quantified using calibration curve.

# Immunogenicity Methods

## Relative Quantitative: Calibration Approach

**Reference standards usually do not accurately reflect Ab affinities/proportions in patient samples.**

→ Lack of similarity between reference standard & study samples

→ Lack of similarity between study samples.

**So we consider Immunogenicity Methods as Quasi-Quantitative & propose the Dilution-based approach.**

# Immunogenicity Methods

## Quasi Quantitative: Dilution Approach

Reference standards not used for screening study samples.

Samples are identified +/- based on signal cut-point.

Positive samples are quantified using dilution levels (titers).

# Minimum Required Dilution (MRD)

**Definition:** Minimum dilution of a sample necessary for the detection of analyte in biological matrix with least interference.

- Endogenous levels are significant in most biomarker applications

**Data:** Requires the dilution series of > 10 drug-naïve samples, preferably from similar patient population.

- Need > 4 replicates of assay diluent samples.

**Method:** Determine the smallest dilution at which the signal is not significantly greater than the assay diluent signal, without compromising on sensitivity & cut point.

- Need to balance statistical and assay considerations.

# Cut Point

## Definition & Purpose

**Definition:** Lowest assay signal above which the sample is considered to be positive.

**Purpose:** Defines the positive/negative criteria for the dilution approach.

- Also used for determining the calibration cut point when using the calibration approach instead of the dilution approach.

# Cut Point (validation phase)

## Data

Generate data from  $> 3$  independent runs of  $> 30$  samples in duplicate at pre-determined MRD.

- preferably from untreated patient population.

Set aside some samples to serve as negative control

- Adequate amount should be set aside for future assay runs.

Analyze the distribution of the data.

- Identify potential positives (high outliers) using inter-quartile range from box-plots of untransformed data. Confirm the positives as appropriate.
- If the distribution of the data is not symmetric, use appropriate data transformation or use robust methods.

# Cut Point (validation phase)

## Methods

### 1. Cut Point = Mean + 1.645\*SD

- 1.645 is the std. normal distribution threshold for 5% false positive rate.
- Standard deviation (SD) is determined from variance-component analysis
- Mean and SD can be determined based on transformed data, if needed.

### 2. Median & Median Absolute Deviation (MAD) can be instead of Mean & SD, if distribution is not symmetric.

- $MAD = \text{Median of } |Data - \text{Median(Data)}|$
- Replace SD by  $1.4826 * MAD$ .

### 3. Upper 95<sup>th</sup> percentile of the data.

# Cut Point (in-study/production phase)

## Methods

Cut Point for each assay run during in-study phase

= Mean or Median of the Negative Control from each assay +  $1.645 \cdot SD$

Or = Pre-dose sample level +  $1.645 \cdot SD$ , for e.g., if drug naïve patient samples are highly heterogeneous (say,  $CV > 20\%$ )

If the variability is constant across validation runs,

- $SD = SD$  from the validation phase
- ANOVA (Bartlett or Levine test): for testing equality of variances.

Otherwise,  $SD$  is estimated from each study run

- need more replicates of neg. control.

# Sensitivity

Need to show that the assay is sensitive enough for intended use.

- Since the assay is intended for Ab detection, sensitivity should be determined in mass units of Ab, even if dilution approach is used for screening samples.

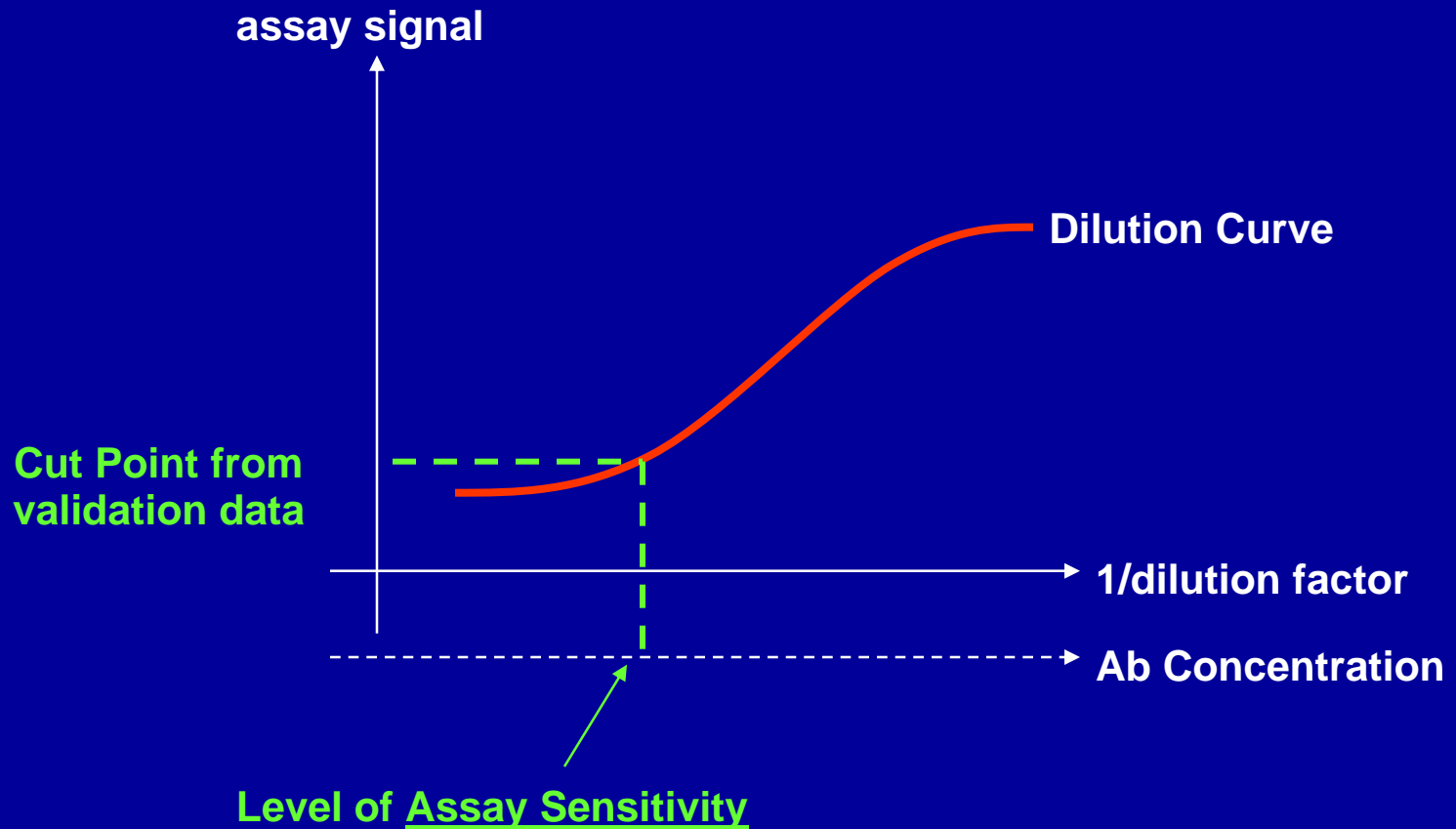
Run a dilution series of a suitable preparation of Antibodies specific to the product (“reference standard”).

- Spiked in same matrix used in screening, e.g. negative control serum
- Run these at predetermined minimal dilution (if MRD = 1:50, all diluted samples should be at 2% serum).

Estimate the dilution where the assay signal equals the cut point.

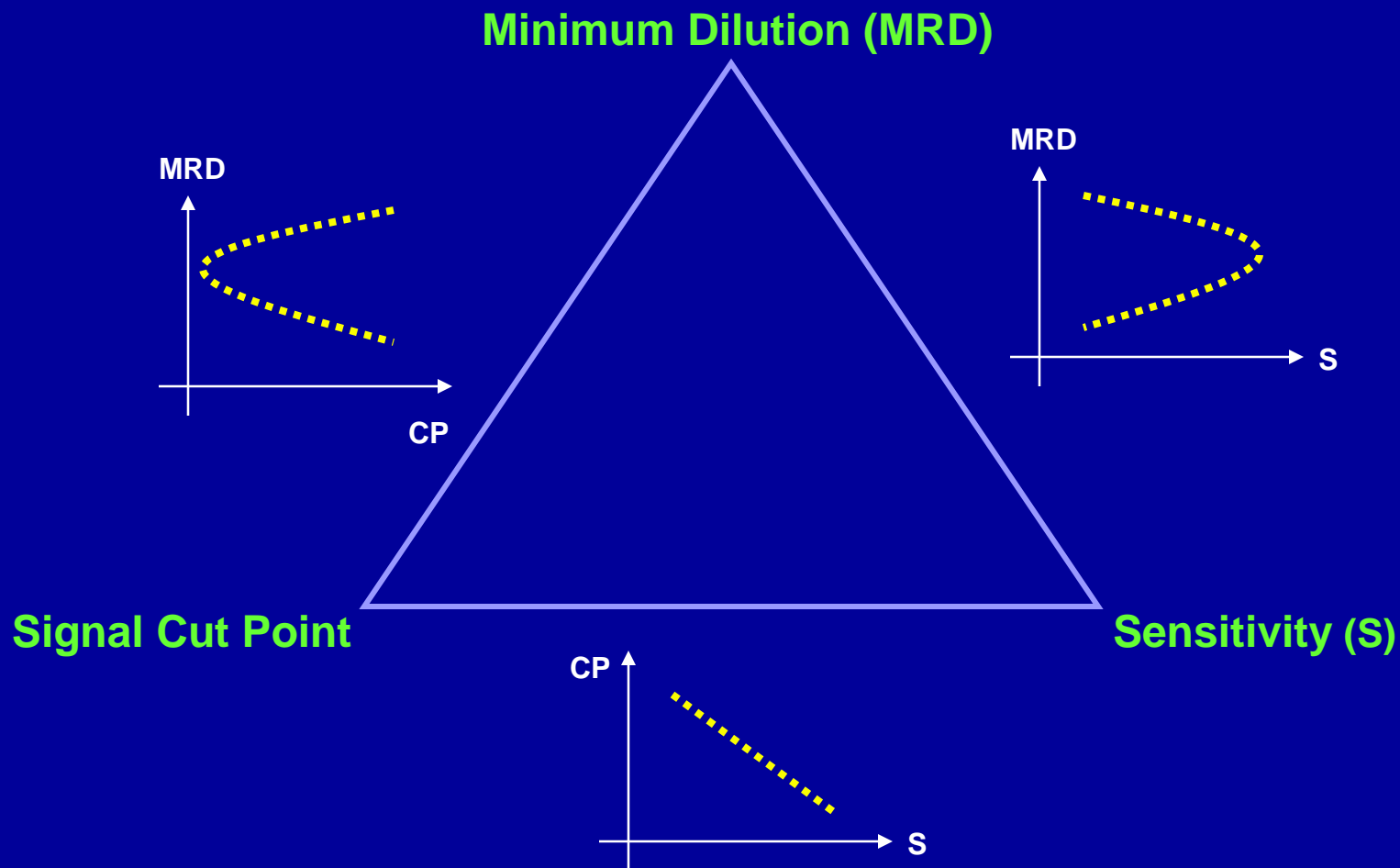
- Use interpolation from a regression curve fit.
- Convert this dilution back to Ab mass units for reporting the assay sensitivity.

# Illustration



# MRD, Signal Cut Point & Sensitivity

## Inter-dependence



# MRD, Cut Point & Sensitivity

## Estimation – Sequential vs. Simultaneous

### Sequential estimation:

- Determine MRD first, then cut point and sensitivity at this predetermined MRD.
- If the cut point or sensitivity is not satisfactory, determine MRD again, and re-compute the cut point and sensitivity.
- Keep iterating until a satisfactory solution is obtained.

### Simultaneous estimation:

- Determine the cut point, sensitivity & % background reduction for different levels of MRD.
- Determine the optimal MRD based on the above data.

# MRD, Cut Point & Sensitivity

## Simultaneous Estimation

Set MRD at a few levels, say, 1:25, 1:50, 1:100.

- Determine the following at each MRD level
  1. % Background Reduction (BR)
  2. Cut Point (CP) & Sensitivity (S)

Determine the MRD that optimizes both.

MRD ->	1:25	1:50	1:100
% BR	60%	70%	80%
CP	0.14	0.1	0.2
S (ng/ml)	500	300	600

### Optimal MRD

Achieves significant background-reduction, while ensuring optimal sensitivity

# Study Drug Interference

## “Verification” of Negatives

Negatives are conclusive only if the samples are “drug-free”.

Immunogenicity testing may be done before washout ends.

Low Positives undetectable due to drug interference.

In these situations, ADA-negative results should be considered “inconclusive” until PK assessments are done.

Negative study samples with PK drug levels  $>$  Drug Tolerance Limit should be considered as “**Inconclusive**”.

**Need to determine the Drug Tolerance Limit.**

# Study Drug Interference

## Determination of Drug Tolerance Limit

Prepare 10-20 low positive “mock” samples by spiking it into individual drug-naïve matrix samples.

Spike increasing levels of study drug in all these samples to generate drug dose-response profiles.

- Preferably in duplicates over at least 2 days

Fit appropriate regression model for these data (e.g., 4PL).

**Drug Tolerance Limit = Interpolated avg. drug level that correspond to the assay signal cut point.**

# Study Drug Specificity

## “Confirmation” of Positives

Positive samples should be tested further to determine whether the Ab is specific to the study drug.

Spike with low and high study drug concentrations.

- A popular approach is to set arbitrary thresholds such as 50% inhibition that are fairly subjective.
- If the inhibition is above the threshold, then the positive sample is study drug-specific.
- **Need a more objective way to determine this threshold (next slide). We can call this “confirmation cut point”.**

# Study Drug Specificity

## Determination of “Confirmation Cut Point”

Prepare 10-20 low positive “mock” samples by spiking it into individual drug-naïve matrix samples.

Spike study drug in these samples at the drug tolerance limit, plus > two concentrations above the drug tolerance level (say, low & high). *Or simply use all the data from the drug tolerance experiment described earlier.*

Prepare in duplicates, run these over at least two days.

Determine % inhibition for all the samples (drug spiked vs. unspiked)

Compute overall SD (including both intra & inter-day) using variance component analysis (ANOVA)

**Specificity Confirmation Cut Point = Mean % inhibition at the drug tolerance limit + 1.645\*SD**

# Quantitation of Positive Samples

## Definition

If a positive sample is confirmed, run a series of dilutions

- Define Titer =  $\log_2(\text{dilution factor})$
- EXAMPLE:

<b>Dilution</b>	<b>1:1</b>	<b>1:2</b>	<b>1:4</b>	<b>1:8</b>
<b>Dilution Factor</b>	<b>1</b>	<b>2</b>	<b>4</b>	<b>8</b>
<b>Titer</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>

- NB: Use  $\log_{10}$  to define Titer, if much higher dilutions are used.
- **Ab level in a positive sample ~ Titer of the sample at which the response equals the cut-point.**

# Quantitation of Positive Samples

## Methods

Ab level in a positive sample

~ Titer of the sample at which the response equals the cut-point

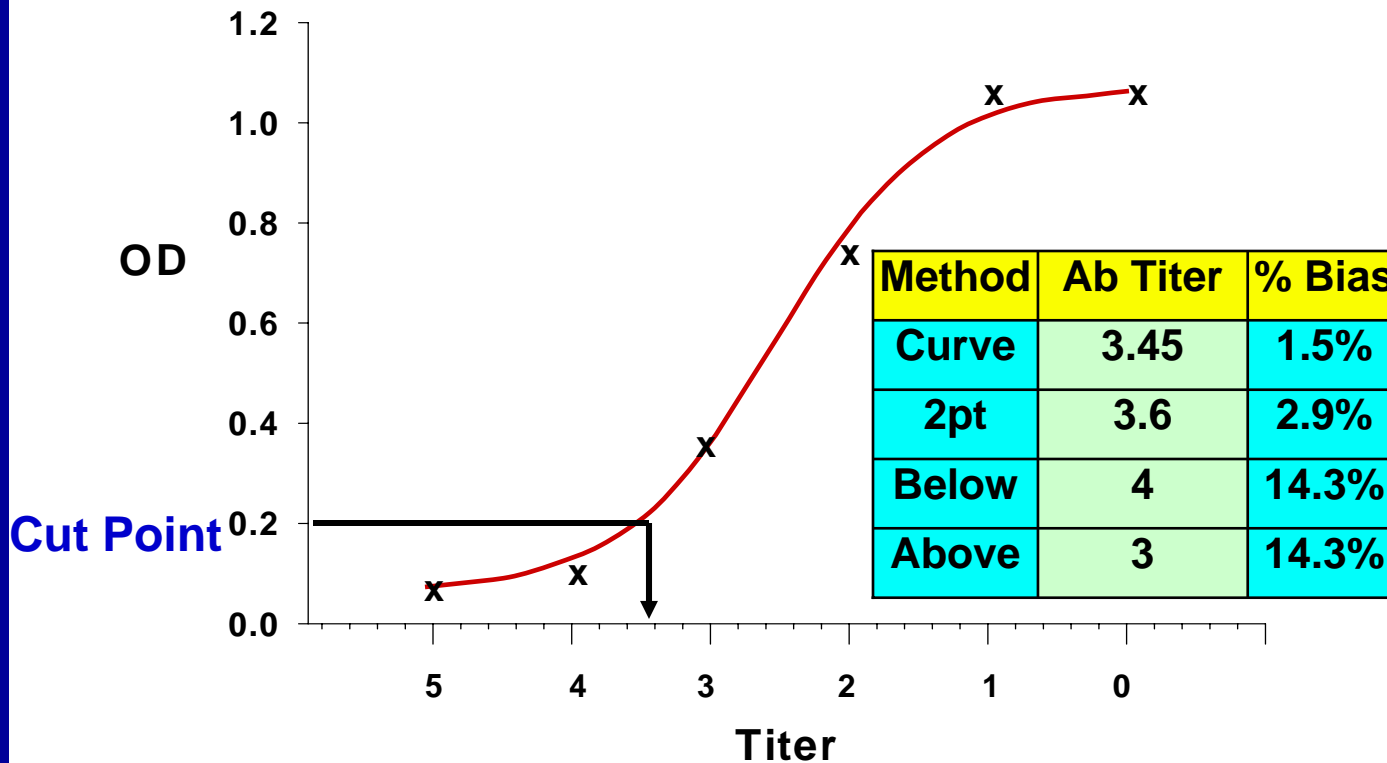
Commonly used methods:

1. **Interpolate from a curve fit (preferred!)**
2. **Interpolate from the two points flanking the cut point (sensitive to outliers).**
3. **Titer result corresponding to the response just below the cut point (biased).**
4. **Titer result corresponding to the response just above the cut point (biased).**

# Quantitation of Positive Samples

## Illustration

Titration/Dilution Profile of a Positive Sample



# Low Positive Control

## How to select the right concentration?

An objective approach:

Run a dilution series of 10-20 high positive control samples (affinity purified standard).

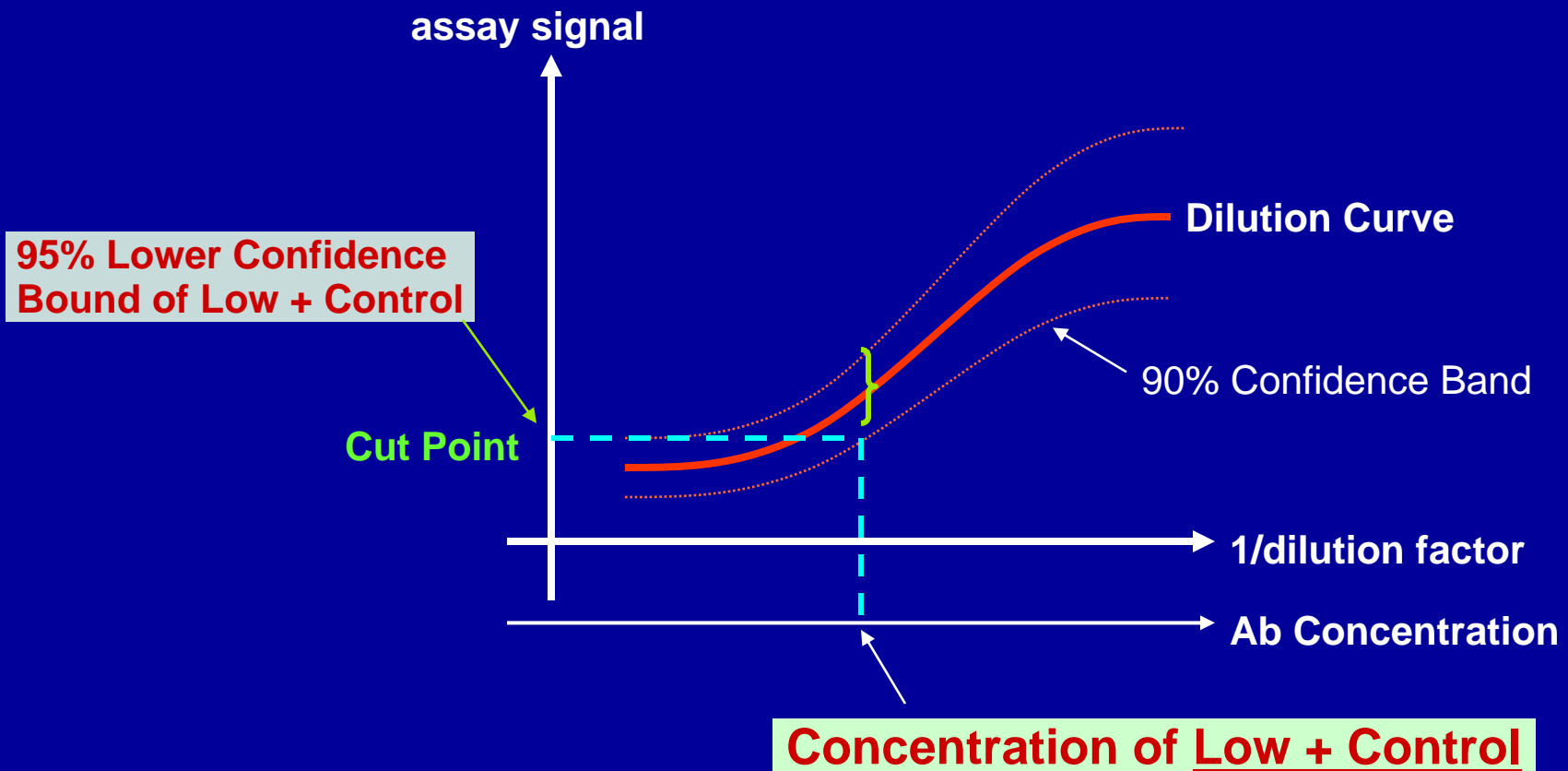
Fit regression curve and estimate the 90% confidence band around the curve.

**The low + control should be at the concentration where the 90% lower confidence limit of the assay response equals the cut point.**

⇒ 5% of the low + control samples will fall below the cut point.

# Low Positive Control

## How to select the right concentration?



Need to apply appropriate weighting when fitting these curves and estimating Confidence bands.

# Precision

## Initial Screen (+/-)

**Data:** Run neg. control, low & high positive controls

- > 3 runs, > 20 samples per run

Titrate the low + and high + control samples

- Generate dilution curve for each low + and high + control sample.

**Analysis:** Obtain the following four types of information

1. Assess the precision of assay signal from neg. control and low + control samples.
  - Intra-run %CV, Inter-run %CV
  - Not too meaningful, but gives a rough idea about the variability.

# Precision

## Initial Screen (+/-)

### 2. Determine false + rate using neg. control samples.

#### Procedure:

- If you have (say) 40 negative control samples, randomly create two groups of 20 samples each.
  1. Determine the cut point based on the 1<sup>st</sup> group of samples.
  2. Using this cut point, determine the false positive rate in 2<sup>nd</sup> group.
- Again randomly create two groups of 20 samples each from the same 40 samples, and determine false positive rate. Repeat this, ~ 10 times.
- Determine the average false + rate from all the random sets.

### 3. Determine false neg. rate using low + control.

- Just compare the response of low + controls to the assay cut point to determine the false negative rate.

# Precision

## Initial Screen (+/-)

### Comments:

- Estimates of False positive rate from the negative control samples as described here should be close to the 5% theoretical error rate as defined in the cut-point calculations.
  - If this is not close to 5%, this should trigger the relevance of the cut point calculated, issues with the samples used for the calculations, etc.
- Similarly the false negative rate of low positive control should be close to the theoretical error rate as defined by the way the concentration chosen for the control.

# Precision

## Positive sample quantification

4. Determine the MSD of Ab Titer results in positive controls.

- **MSD (Minimum Significant Difference)** = Smallest difference in any two Titer results that can be considered as “real”.

- **Method:**

- Estimate Ab Titers, interpolating from the dilution curves of all the positive controls

- $MSD = 2 * \sqrt{2} * SD$ , where the standard deviation (SD) comes from the analysis of variance (ANOVA) of the Ab Titers.

- **Example:** MSD = 3 implies that samples with Ab levels within 3 titers of each other are not statistically different.

- If Ab Titers for two positive samples are 4 and 6.5 respectively, then these two samples are not considered to be different.

- If a sample with Ab Titer of 4 is tested again, its Titer can be anywhere from 1 to 7.

# Summary

Key elements of immunogenicity methods such as

- Determination of MRD, Cut Point, Sensitivity, ...
- Study-Drug Specificity
- Quantitation of Positive Samples (Titer estimation)
- Precision of “+/-” determination and Titer results
- Selection of concentration for the positive controls

are governed by a synergy of

**Statistical + Practical + Biological Thinking!**

# BACK-UP SLIDES

# Cut Point

## Outliers

- Outliers that are biologically confirmed as “positives” via drug-specificity testing can be excluded.
- Simple & practical approach for defining statistical outliers using box-plots:
  - **Values that are greater than “75<sup>th</sup> percentile + 1.5\*(75<sup>th</sup> percentile minus 25<sup>th</sup> percentile)” can be defined as statistical outliers and excluded from the analysis when using the Mean and SD approach for defining cut point.**
  - **Per the risk based approach, it is less controversial to remove more “high end” outliers than we should. (i.e., having more false positives is better than having more false negatives).**
- In general, when there are several outliers and/or highly skewed data, use the Median and MAD approach. Or analyze data in the log scale, and take anti-log in the end; note that outliers in the original scale may not be outliers in the log sale.

# Cut Point (in-study/production phase)

## Summary

### Fixed Cut Point:

- Using cut point determined from validation phase for all study runs.
  - Need to show that neg. control signal and the variability are consistent

### Floating Cut Point:

- Using average signal from the neg. control of a study run + 1.645 times the SD that was determined from the validation phase
  - Need to show that the variability of neg. control is similar across runs during the validation phase.

### Patient-Specific Floating Cut Point:

- Using pre-dose sample instead of neg. control.
  - E.g., when there are significant inter-patient differences from differing Ab affinities

# Cut Point (in-study/production phase)

## Summary

### Dynamic Cut Point:

- Using neg. control from each run to obtain the signal and the variability needed for defining the cut point.

### Using ratios with background to define Cut Point:

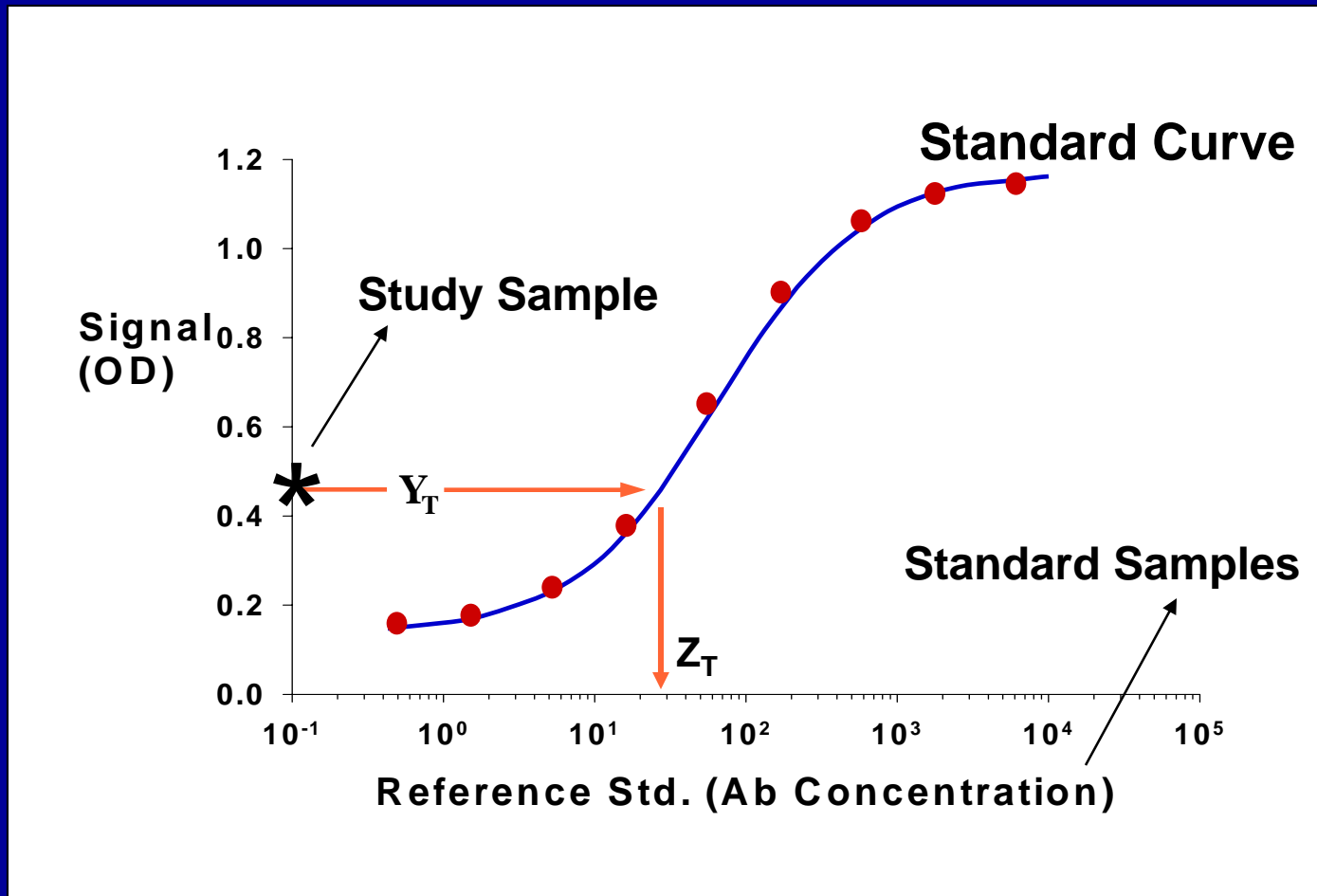
- If this is done, then the analysis should be done with respect to log scale, due to the greater impact the low values can have on the ratios when analyzing in terms of the original scale.

### Overall Recommendation:

- In general, Floating is preferred over Dynamic and Fixed
- Adopting risk-based approach, use whichever is the smallest ☺

# Immunogenicity Methods

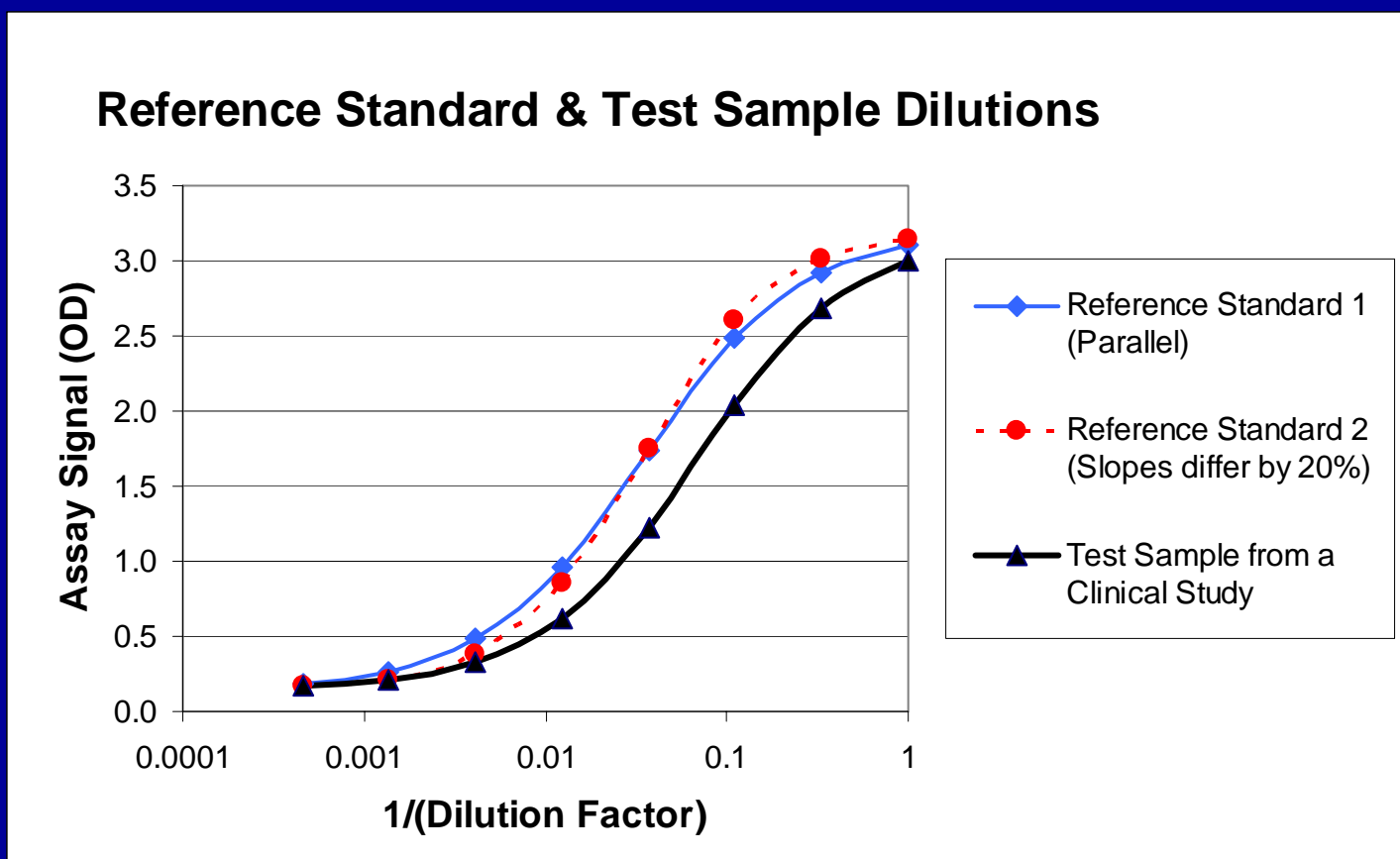
## Relative Quantitative: Calibration Approach



$Y_T$  = Observed Assay signal of Test ( $T$ ) sample

$Z_T$  = Calibrated Biomarker level in test sample =  $F^{-1}(Y_T, \beta)$

# Validity of reference standard

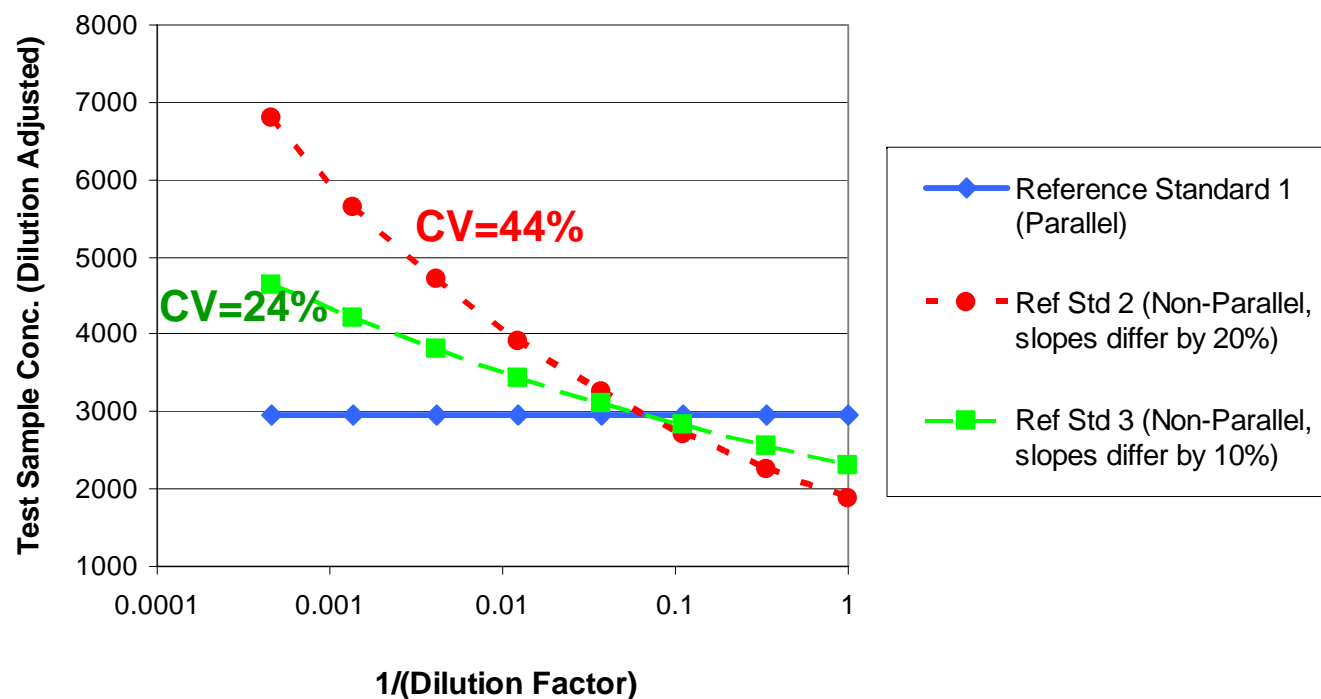


Reference Standard 1: Dilution profile is similar to the test sample

Reference Standard 2: Dilution profile slope differs by 20% from the test sample

# Effect of Invalid Reference Standards

## Dilution Adjusted Concentration versus Dilution



Seemingly minor issues with reference standards can have major impact on the reliability of study sample assessment.