System and Sample Suitability Assessment for Potency Methods

Ruojia Li

Bristol-Myers Squibb

CASSS Bioassays

April 16-17, 2018

Bristol-Myers Squibb

Outline

- Overview of system and sample suitability assessment for potency methods
- Suitability parameters and acceptance criteria considerations
- Examples

Introduction

System suitability

- Apply to reference standard and control sample
- Assess the validity of an assay

Sample suitability

- Apply to test sample
- Assess the validity of the sample potency estimate

Why Do We Need System and Sample Suitability Assessment?

Ensure the quality of potency assay results

- Biologically meaningful
- Good data quality (acceptable dose-response curve fit, similarity) which ensure reliable potency estimation

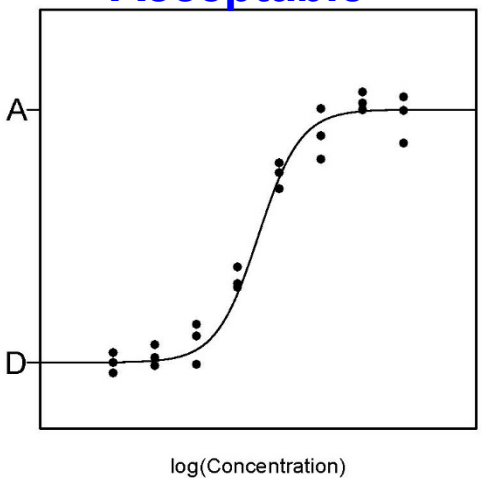
What could happen when the suitability was not assessed properly

- Unreliable or meaningless potency estimation
  - Due to poor curve fit
  - Due to violation of inherent assumption for potency calculation (similarity between reference and sample)
- Unreasonable high assay/sample failures

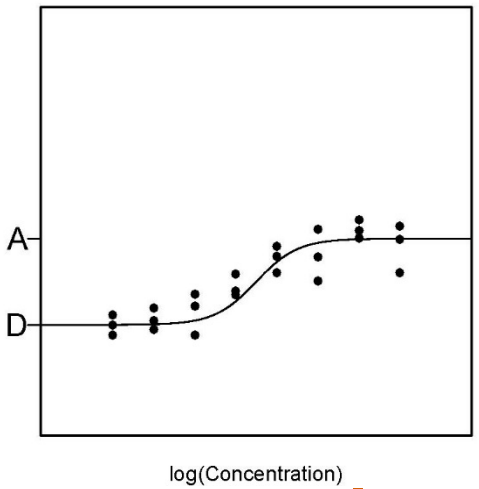
Examples of Suitability Parameters

- **Parameters to ensure acceptable signal and noise**
  - **Sufficient signal: A > xx**
  - **Under controlled noise: D < xx**
  - **Sufficient signal to noise separation: A/D, A-D**

Acceptable



Undesirable

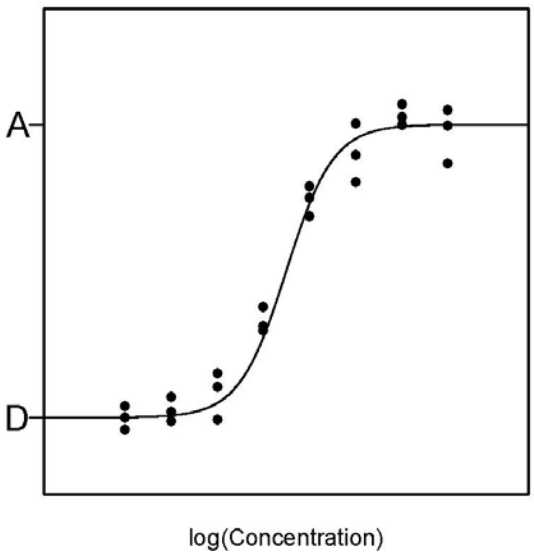


- **Parameter to ensure proper control sample result**
  - **Potency of control falls within expected range**

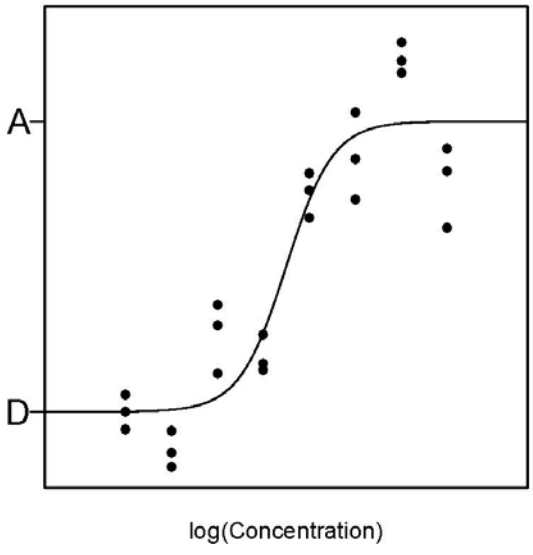
Bristol-Myers Squibb

Examples of Suitability Parameters

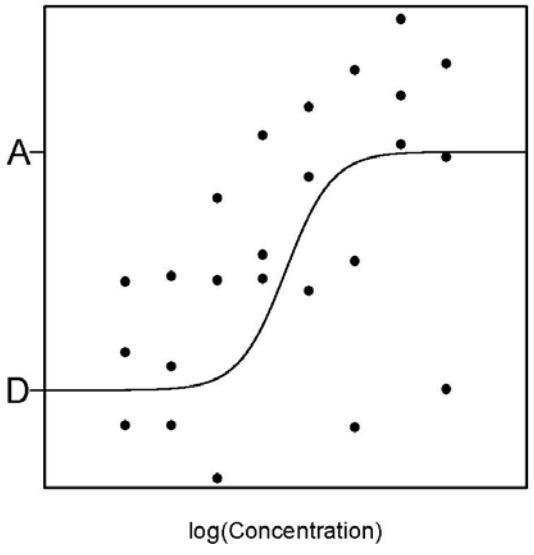
- **Parameters for quality of dose-response curve fit**
  - **Goodness of fit: Lack-of-fit (LOF) measurements (sum of squares, P-value, etc.), R^2**
  - **Precision: Residual mean squared error, %CV of replicated response, confidence interval for potency estimate**



Acceptable



Undesirable (Lack of fit)

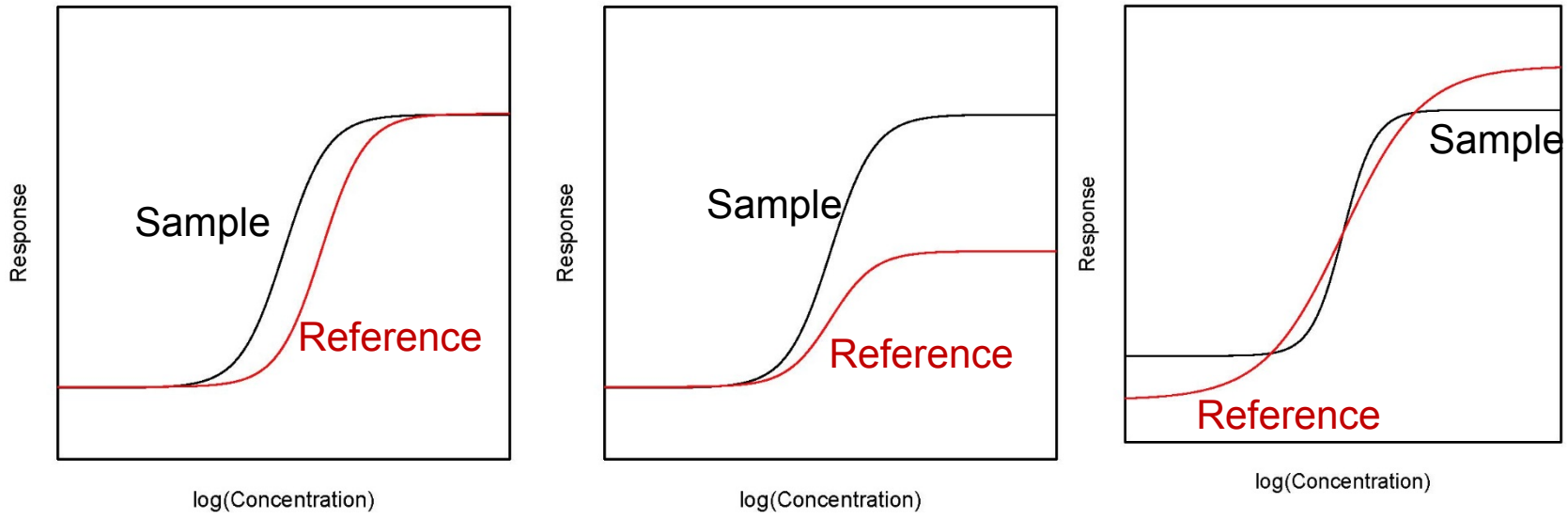


Undesirable (Poor precision)

Bristol-Myers Squibb

Examples of Suitability Parameters

- **Parameters for similarity between reference and sample curves**
  - **Ratio or difference between reference and sample curve parameters, non-parallelism sum of squares**



Acceptable **Undesirable** **Undesirable**

Selection of Suitability Parameters

General principle

Use parameters with low correlations to provide meaningful assessment for quality of assay results

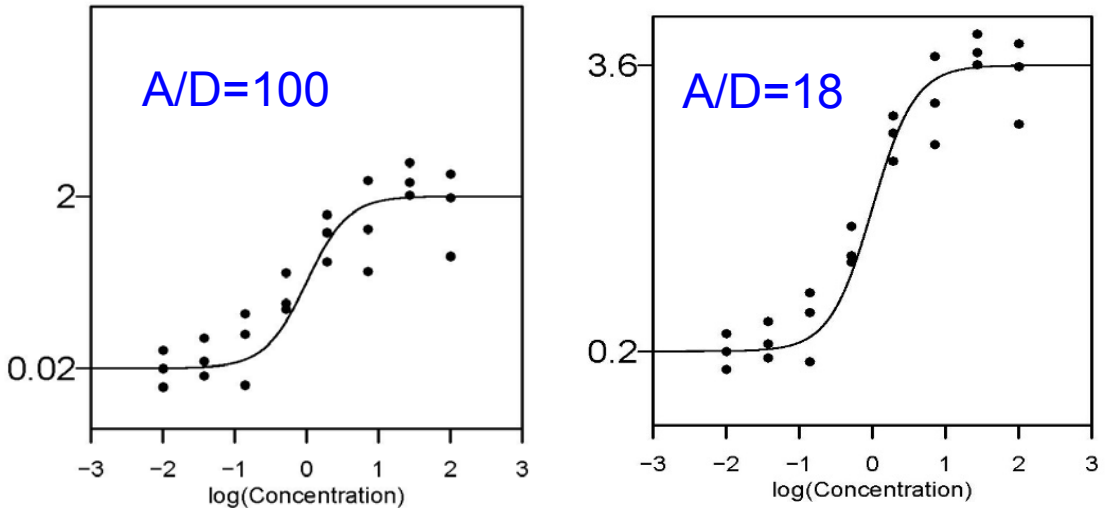
Considerations for suitability parameter selection

- **What is the intended use of the parameter? Is it directly related to the quality of assay and/or potency estimation?**
  - **Many curve fitting parameters may be routinely monitored. However, not all should be applied for suitability assessment**
- **Does the parameter provide meaningful assessment for the intended purpose?**
  - **Suitable for the type of assay**
  - **Effectively reject undesirable assay/sample and retain acceptable assay/sample**
- **Do other parameters provide similar assessment?**
  - **Avoid redundancy**

Example: Does signal to noise ratio (A/D) properly ensure meaningful dose-response curve?

- Reference curve signal to noise ratio (A/D) is a commonly used system suitability parameter
- A/D provides meaningful control for many methods
- However, caution should be taken to avoid arbitrary or insufficient assessment.

Example 1: D values (noise) are expected to be nearly 0

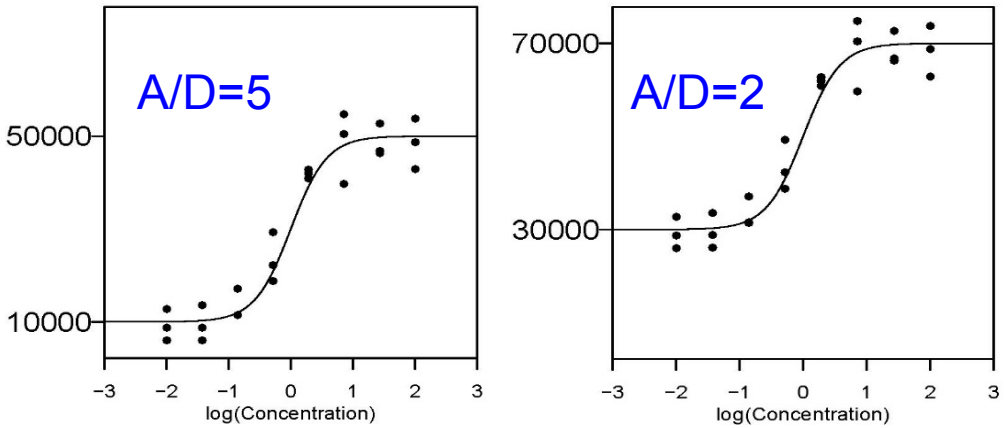


- Small changes in D have big impact on A/D.
- Alternative parameters (e.g., A-D) should be considered.

Bristol-Myers Squibb

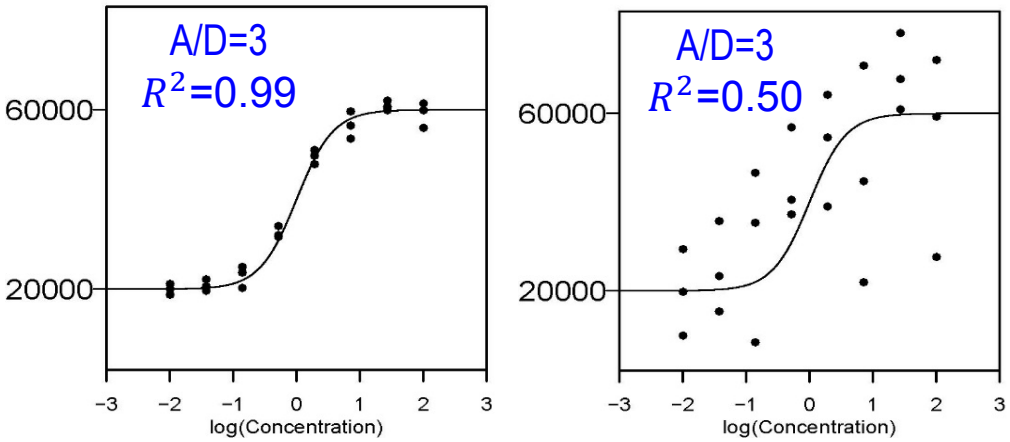
Example: Does signal to noise ratio (A/D) properly ensure meaningful dose-response curve?

Example 2: A and D values vary significantly from assay to assay



- A/D heavily rely on the absolute readouts and may become an arbitrary measurement.
- Alternative parameters (e.g., A-D) should be considered.

Example 3: Assays with same A/D but different variability



- Same A/D doesn't mean same quality of curve fit.
- A/D should be coupled with other parameters (e.g., R^2) to provide meaningful control)

Bristol-Myers Squibb

Determination of Acceptable Range for the Selected Suitability Parameters

General principle:

Use representative data and proper evaluation to determine the suitability acceptance criteria

Considerations for suitability criteria determination

- Representative data set
  - Data generated under final assay condition
  - Consider typical sources of variation
- Distribution of the suitability results
- Examination of extreme results
- Impact on potency estimation
- Intended purpose of the method / Phase of study / Method knowledge and experience

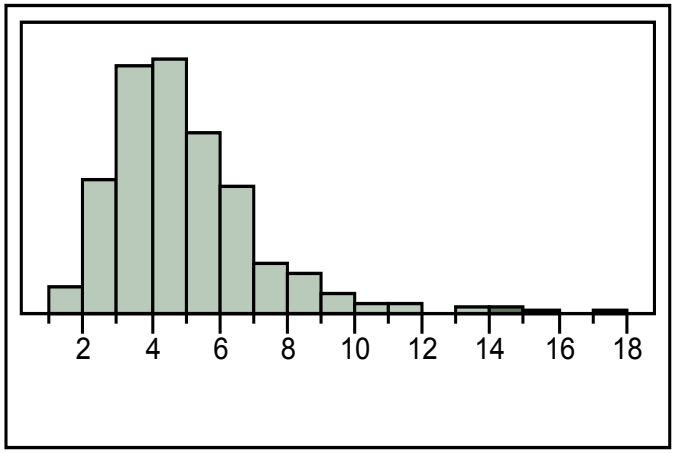
Example: Determination of Suitability Parameter Acceptable Range

Parameter: Relative LOF error

- A measure of lack-of-fit
- High value indicate inadequate model fit

Evaluation of data distribution

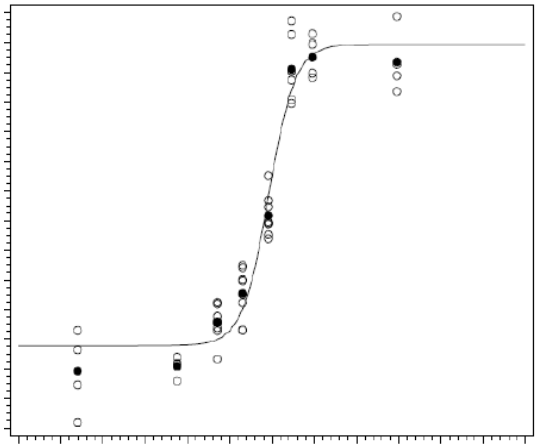
- Histogram of relative LOF results generated during method validation



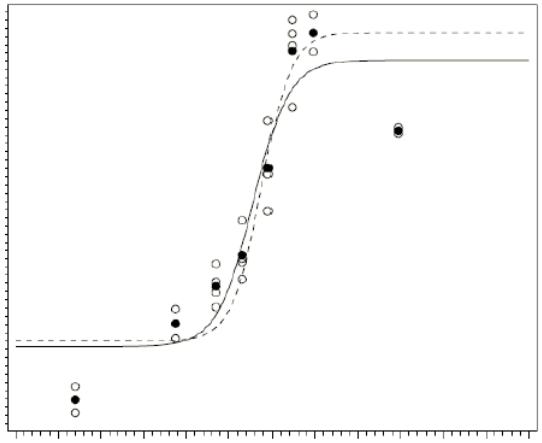
Example: Determination of Suitability Parameter Acceptable Range

Examination of extreme results

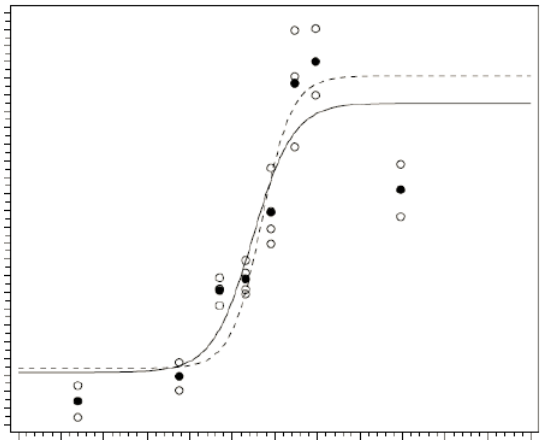
Examples:



Acceptable fit
(Relative LOF error =6.0%)



undesirable fit
(Relative LOF error =11.5%)



Poor fit
(Relative LOF error =14.6%)

Determined suitability criteria:

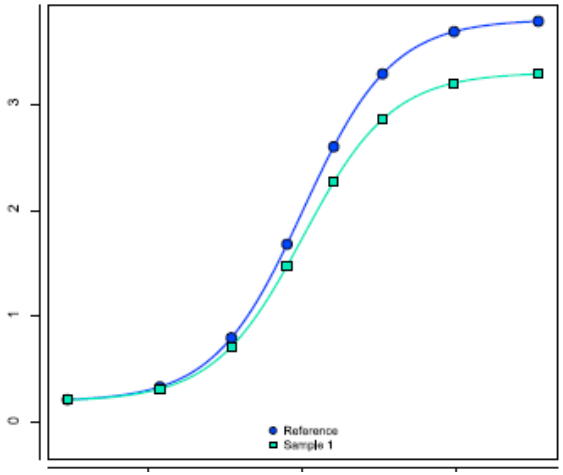
Relative LOF error ≤ 10.0%

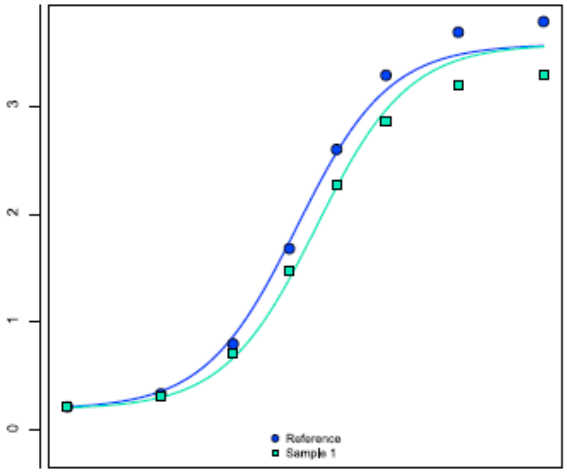
Example: Consideration of Impact on Potency Estimation

The acceptance range of the suitability parameter may also be informed by the impact on potency estimation

Example: Impact of non-parallelism

Assay 1





Unrestricted curves

Upper asymptote: A_{ref}=3.8, A_{sample}=3.3

Lower asymptote: D_{ref}=D_{sample}=0.2

Inflection point: C_{ref}=C_{sample}=1

Hill's slope factor: B_{ref}=B_{sample}=1

Restricted curves

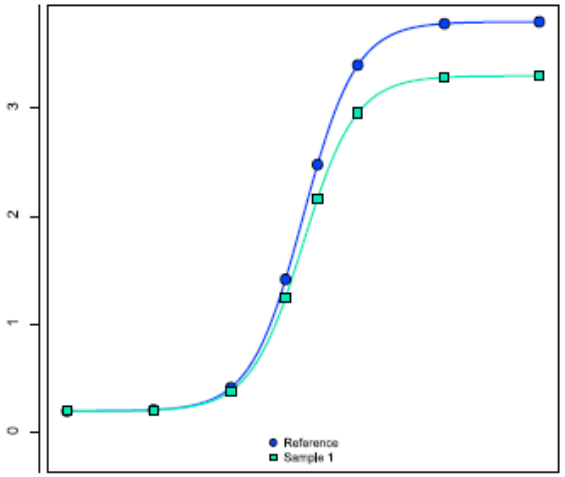
Estimated sample potency =65% (significantly impacted by the deviation at upper asymptote)

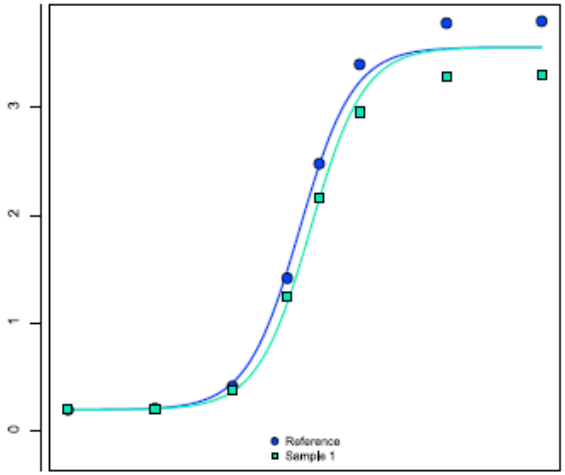
Bristol-Myers Squibb

Example: Consideration of Impact on Potency Estimation

Example: Impact of non-parallelism

Assay 2





Unrestricted curves

Upper asymptote: A_{ref}=3.8, A_{sample}=3.3

Lower asymptote: D_{ref}=D_{sample}=0.2

Inflection point: C_{ref}=C_{sample}=1

Hill's slope factor: B_{ref}=B_{sample}=3

Restricted curves

Estimated sample potency =87%

(same level of the deviation at upper asymptote has less impact on potency estimation due to steeper slope)

All the curve parameters are the same as assay 1 except for B factor

Bristol-Myers Squibb

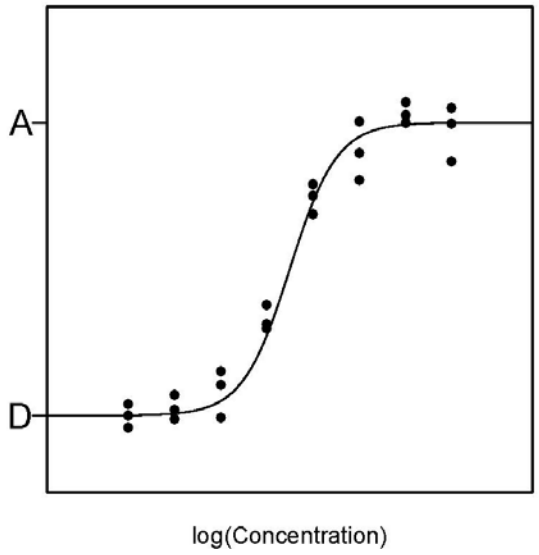
Example: Consideration of Impact on Potency Estimation

- Same level of non-parallelism have different impact on potency estimation for assays with different B factor
  - Assays with smaller B are more sensitive to non-parallelism and require tighter control of the non-parallelism parameters
  - Assays with steeper slope are less sensitive and can tolerate higher level of non-parallelism
- The impact on potency estimation should be taken into consideration when setting acceptance range for suitability parameter

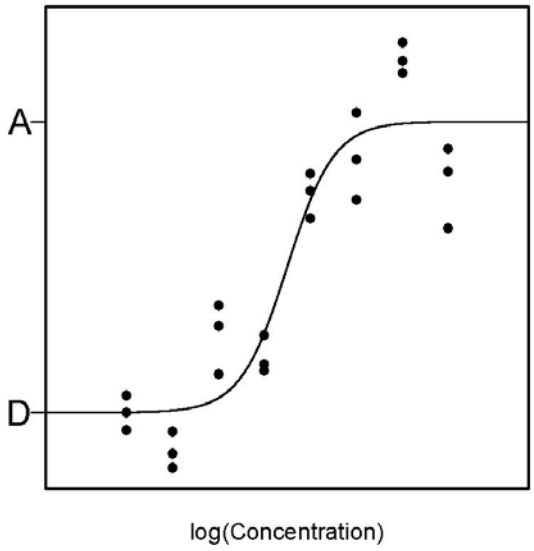
Example: Different Suitability Parameters for Lack-of-Fit Assessment

Lack-of-Fit (LOF) assessment:

Assess the adequacy of the dose-response model. Measure the closeness of the fitted curve to the observed data.



Acceptable fit



Undesirable fit

Example: Different Suitability Parameters for Lack-of-Fit Assessment

LOF P-value

$$F\ ratio = \frac{SS_{LOF}/DF_{LOF}}{SS_{PE}/DF_{PE}} = \frac{\sum_{i,j}(\bar{y}_i - \hat{y}_i)^2/DF_{LOF}}{\sum_{i,j}(y_{i,j} - \bar{y}_i)^2/DF_{PE}}$$

- Based on ANOVA F test
- Conclude lack of fit if P-value is significant (e.g., < 0.05)
- Compare LOF error to pure error
  - SS_{LOF}: Measures overall LOF error (difference between local average $\bar{y}_i$ and fitted value $\hat{y}_i$)
  - SS_{PE}: Measures overall pure error (difference between individual observation $y_{i,j}$ and local average $\bar{y}_i$)

LOF sum of squares

$$SS_{LOF} = \sum_{i,j}(\bar{y}_i - \hat{y}_i)^2$$

- Directly measures LOF error without comparing to pure error
- Conclude lack of fit if SS_{LOF} is large

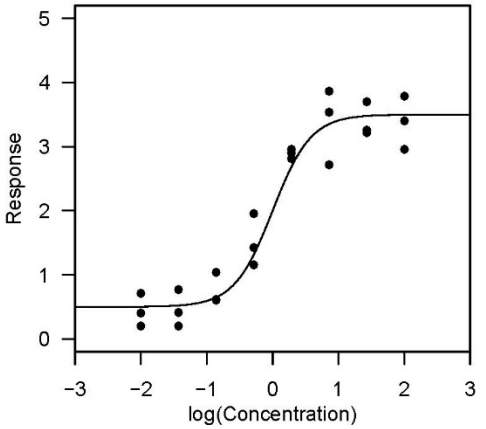
Relative LOF error

$$\frac{\sqrt{SS_{LOF}/N}}{\boldsymbol{A}_{ref} - \boldsymbol{D}_{ref}} \times 100\%$$

- LOF error normalized against reference A-D

Example: Different Suitability Parameters for Lack-of-Fit Assessment – LOF P-value vs. SS_{LOF}

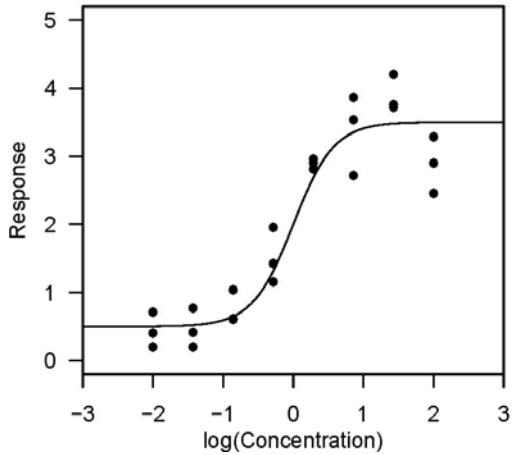
Example 1: LOF P-value successfully conclude good vs. poor fit



SS_{LOF}=0.04 (pass)

SS_{PE}=1.9

LOF P-value = 0.99 (pass)

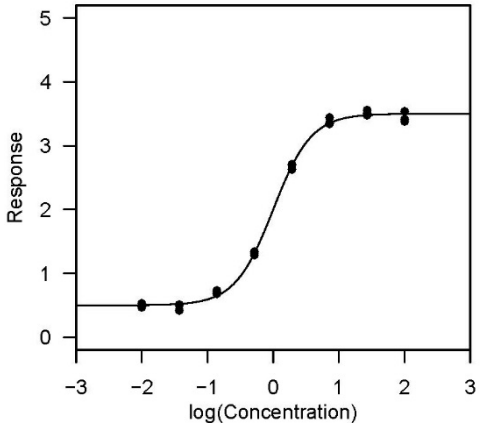


SS_{LOF}=1.6 (fail)

SS_{PE}=1.9

LOF P-value = 0.04 (fail)

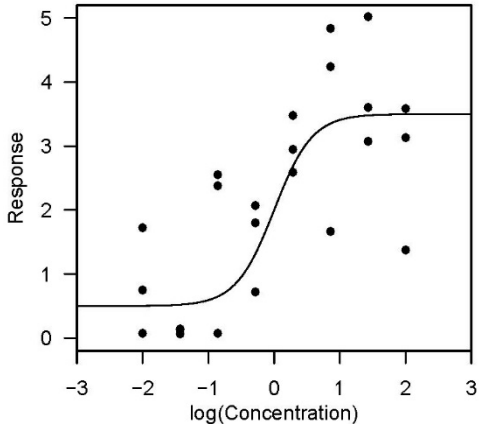
Example 2: LOF P-value unable to properly conclude good vs. poor fit



SS_{LOF}=0.03 (pass)

SS_{PE}=0.03

LOF P-value = 0.03 (fail)



SS_{LOF}=5.5 (fail)

SS_{PE}=17.0

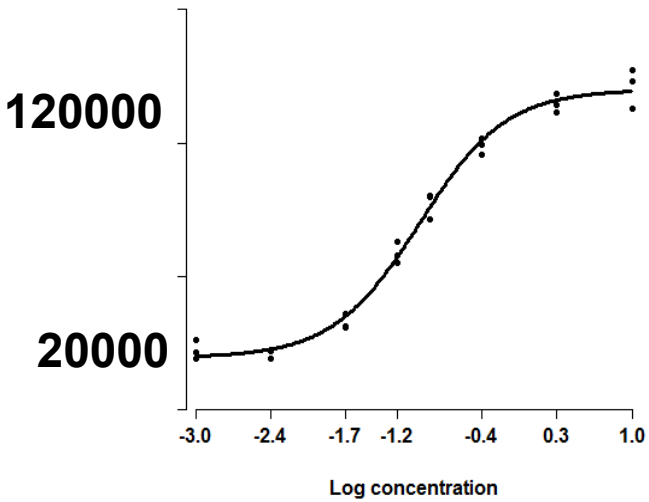
LOF P-value = 0.32 (pass)

In both examples, SS_{LOF} works properly

Example: Different Suitability Parameters for Lack-of-Fit Assessment - SS_{LOF} vs. Relative LOF Error

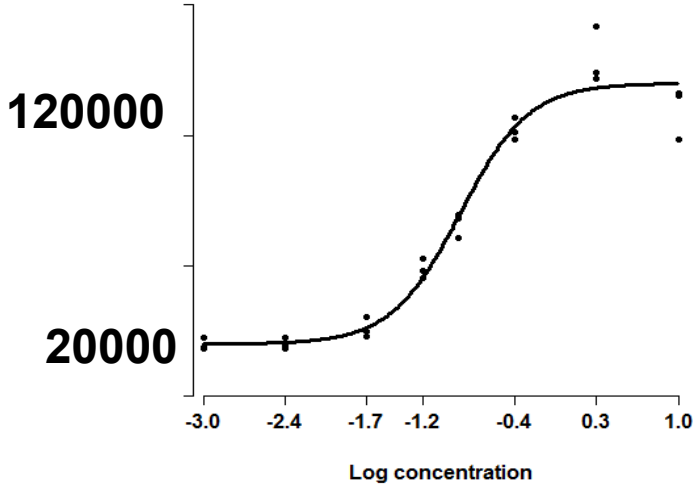
Example 3: LOF sum of squares successfully conclude good vs. poor fit when comparing curves from same instrument with high readouts

Curve 1
Acceptable fit
Instrument A (High readouts)



$SS_{LOF} = 7.2 \times 10^7$ (low)
Relative LOF error = 1.7% (low)

Curve 2
Undesirable fit
Instrument A (High readouts)

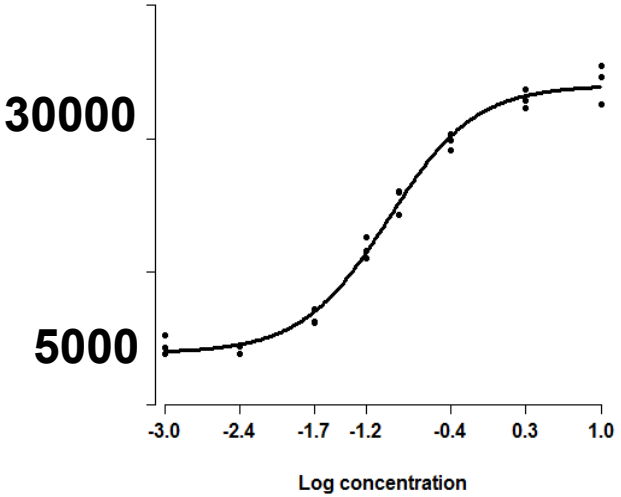


$SS_{LOF} = 7.0 \times 10^8$ (high)
Relative LOF error = 5.4% (high)

Example: Different Suitability Parameters for Lack-of-Fit Assessment - SS_{LOF} vs. Relative LOF Error

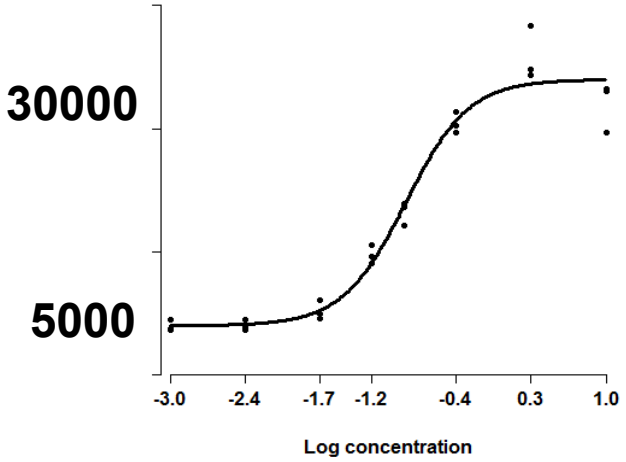
Example 4: LOF sum of squares successfully conclude good vs. poor fit when comparing curves from same instrument with low readouts

Curve 3
Acceptable fit
Instrument B (low readouts)



$SS_{LOF} = 4.5 \times 10^6$ (low)
Relative LOF error = 1.7% (low)

Curve 4
Undesirable fit
Instrument B (low readouts)

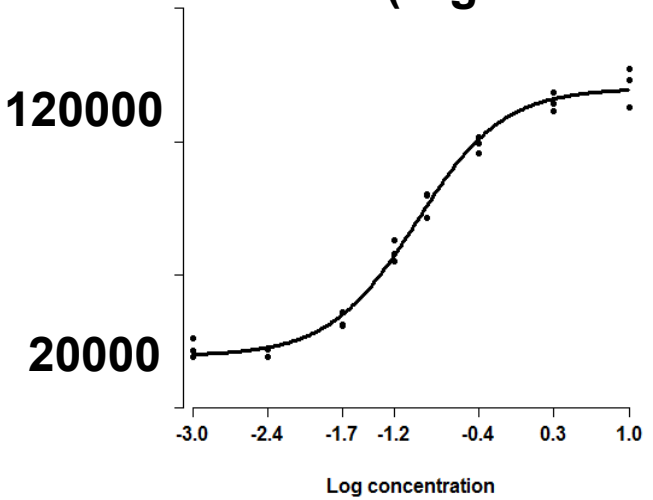


$SS_{LOF} = 4.4 \times 10^7$ (high)
Relative LOF error = 5.4% (high)

Example: Different Suitability Parameters for Lack-of-Fit Assessment - SS_{LOF} vs. Relative LOF Error

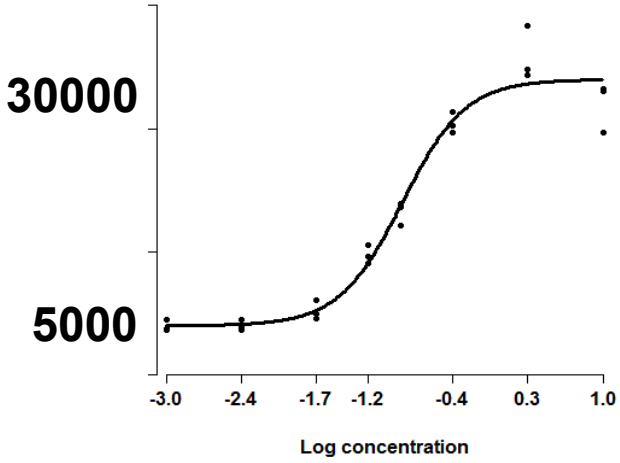
Example 5: LOF sum of squares doesn't work properly when comparing curves between instruments with different readouts

Curve 1
Acceptable fit
Instrument A (High readouts)



$SS_{LOF} = 7.2 \times 10^7$ (high)
Relative LOF error = 1.7% (low)

Curve 4
Undesirable fit
Instrument B (low readouts)



$SS_{LOF} = 4.4 \times 10^7$ (low)
Relative LOF error = 5.4% (high)

Relative LOF error still works properly

Bristol-Myers Squibb

Example: Different Suitability Parameters for Lack-of-Fit Assessment

LOF P-value

- Works properly when the level of pure error are consistent from assay to assay
- May over-sensitively reject precise data and propensity to retain undesirable noisy data

LOF sum of squares

- Overcomes the limitation of LOF P-value since not impacted by the pure error
- Requires A-D to be consistent across labs /instruments /analysts in order to provide meaningful assessment

Relative LOF error

- Independent of the magnitude of the response readings. Therefore more robust than LOF sum of squares when A-D vary from assay to assay

Summary: Common Types of Parameters for Suitability Assessment

- Signal to noise
  - E.g., A, D, A/D, A-D
- Potency of the control sample
- Quality of dose-response curve fit
  - Goodness of fit
  - Precision
- Similarity between reference and sample

Summary: Considerations for Suitability Parameter and Criteria Determination

- Be aware of intended use and limitations of the parameters. Carefully select suitable parameters.
- Set acceptance range based on representative data set and thorough evaluation
  - Data distribution / examination of extreme results
  - Impact on potency estimation
  - Phase of study / experience

General Conclusion

- Suitability assessment is a integral part of the potency methods
- Proper suitability assessment ensure scientifical meaningfulness and good data quality which produce reliable potency results
- It is critical to carefully determine system and sample suitability parameters and acceptance criteria that are suitable for intended purpose

Acknowledgement

Bioassay Center of Excellence

Marcel Zocher

Jeff Glenn

Cassie Liu

Weiguo Cai

Isam Qahwash

Scott Umlauf

Back-Up Slides

Example: Different Suitability Parameters for Lack-of-Fit Assessment

Case study

- The LOF P-value criterion of a ELISA method was replaced by LOF sum of squares criterion
- Summary of retrospective analysis results demonstrate that the new criterion (LOF sum of squares) more efficiently reject undesirable results and retain acceptable results

Case Study: Method Performance Comparison: LOF P-Value vs. LOF Sum of Squares

Data Set		Mean of LOF Sum of Squares	Mean of pure Error Sum of Squares	Mean of QC Potency (%)	SD of QC Potency (%)
Overall (N=321)*		0.11	0.52	101.1	7.2
LOF P-value (Old)	Pass (N=268)	0.10	0.59	101.3	7.5
	Fail (N=53)	0.19	0.15	100.4	5.4
LOF sum of squares (New)	Pass (N=303)	0.09	0.46	100.9	7.1
	Assays that pass LOF sum of squares & failed P-value (N=46)	0.12	0.11	100.2	5.5
	Fail (N=18)	0.57	1.46	104.8	8.0

* Assays failed other suitability criteria (e.g., non-parallelism) were excluded

Summary of Case Study

- The LOF P-value criterion causes higher failure rate.
- The assays that failed the LOF P-value criterion have better accuracy and precision than the assays that passed the LOF P-value.
- The LOF sum of squares criterion can more effectively invalidate assays with poor fit and retain assays with precise fit.

Simulation Study: Impact of Non-Parallelism on Potency Estimation

Objective: study the impacts of non-parallelism on potency estimation.

- **Fix the reference curve**
- **The test curve varies by different combinations of lower and upper asymptote ratios, and slope ratios.**
- **Calculate potency based on restricted model**

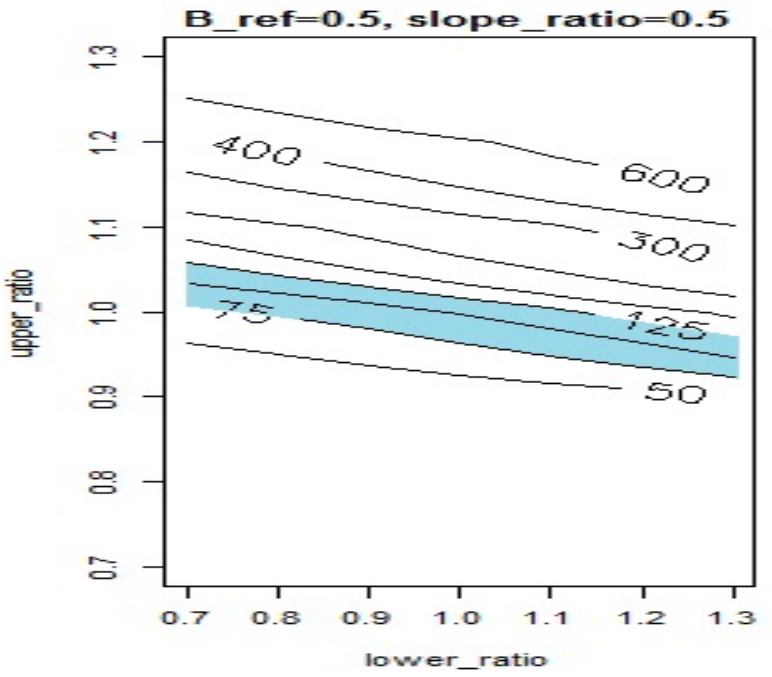
Parameter	Setup
Articles	Reference and test sample
A, C, D values (reference)	A = 0.5, C = 1, D = 3.5
B value (reference)	0.5, 1, 2, 3
Low asymptote ratio (test over reference)	0.7 – 1.3 by 0.1
Upper asymptote ratio (test over reference)	0.7 – 1.3 by 0.1
Slope ratio (test over reference)	0.5 – 2.0 by 0.1

Simulation Study: Impact of Non-Parallelism on Potency Estimation – Contour Graph

Contour graph of potency given reference B and slope ratio

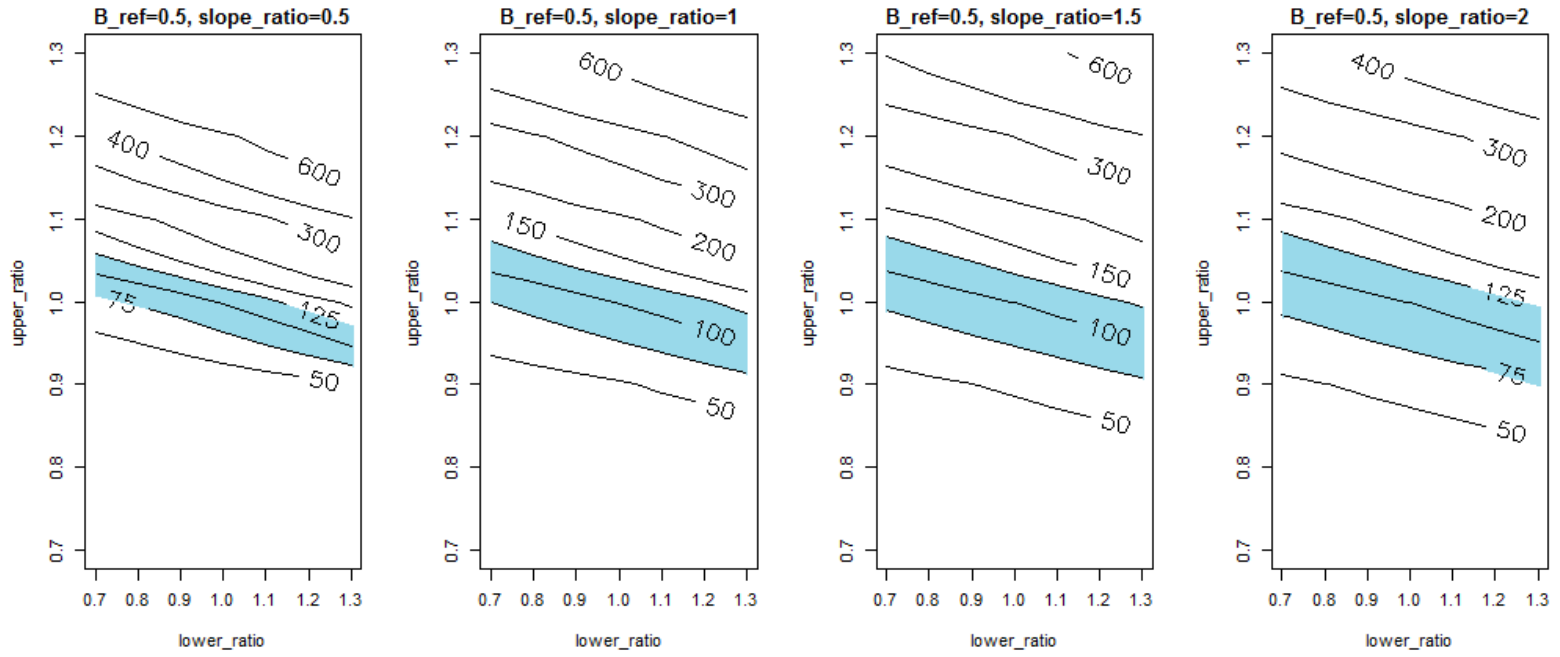
- **Show how the estimates of relative potency changes along the lower and upper asymptote ratios at a given slope ratio.**

- **Highlight the contours of potency between 75% and 125% (expected potency is 100%).**



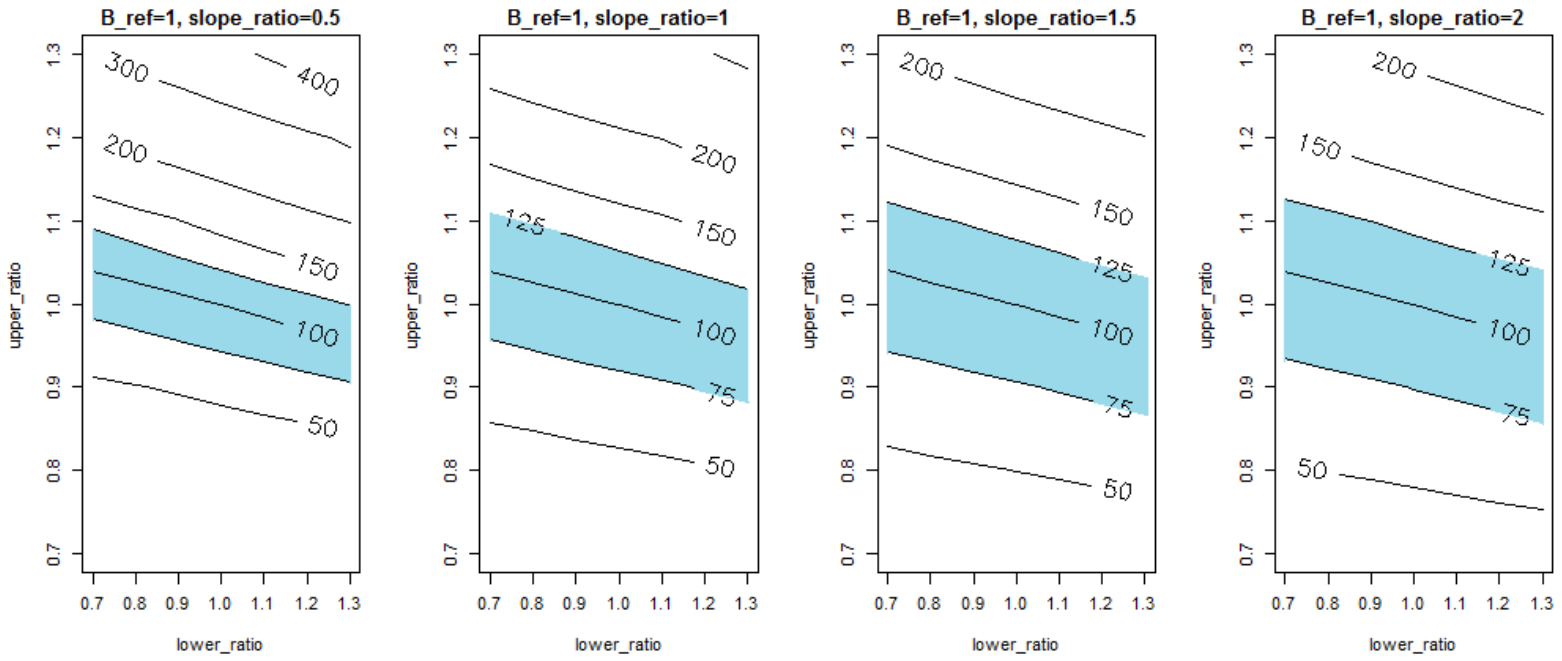
Simulation Study: Impact of Non-Parallelism on Potency Estimation – Results (B=0.5)

Slope ratio is 0.5, 1.0, 1.5, and 2.0



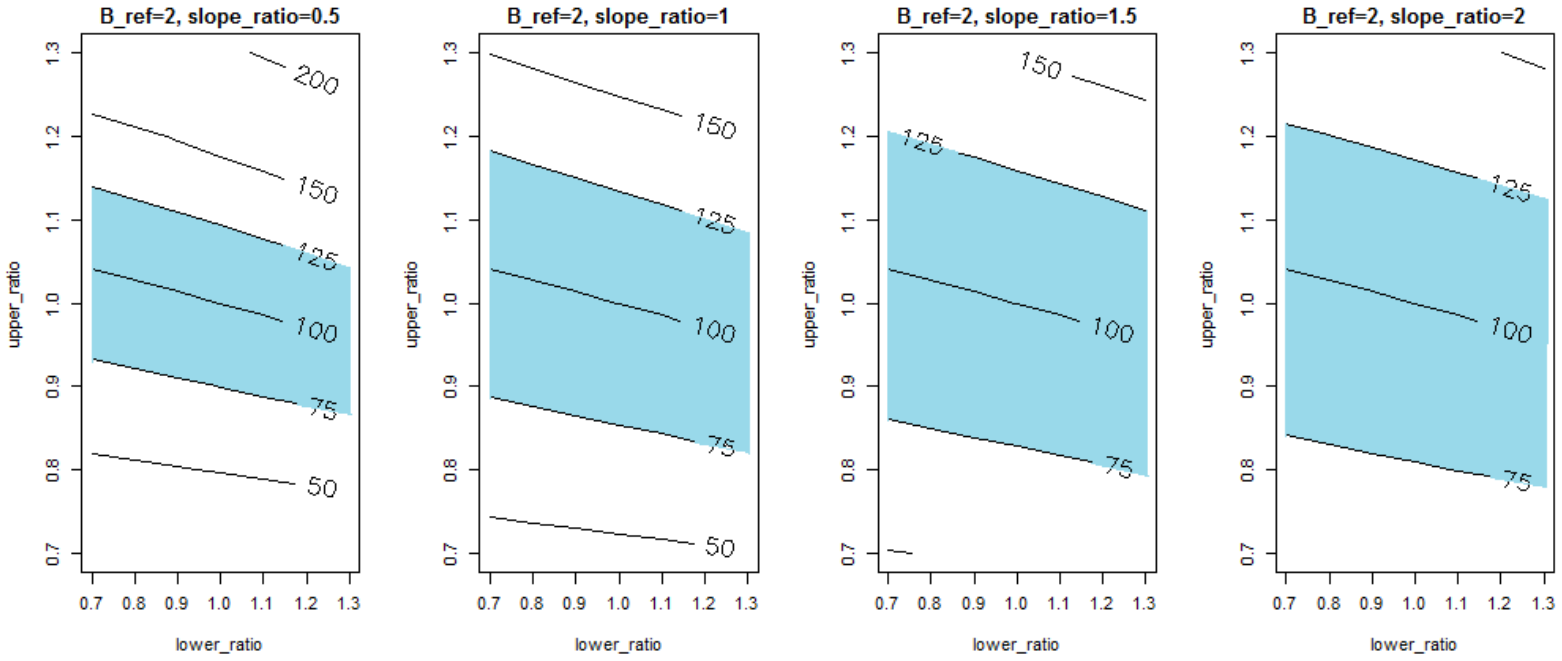
Simulation Study: Impact of Non-Parallelism on Potency Estimation – Results (B=1)

Slope ratio is 0.5, 1.0, 1.5, and 2.0



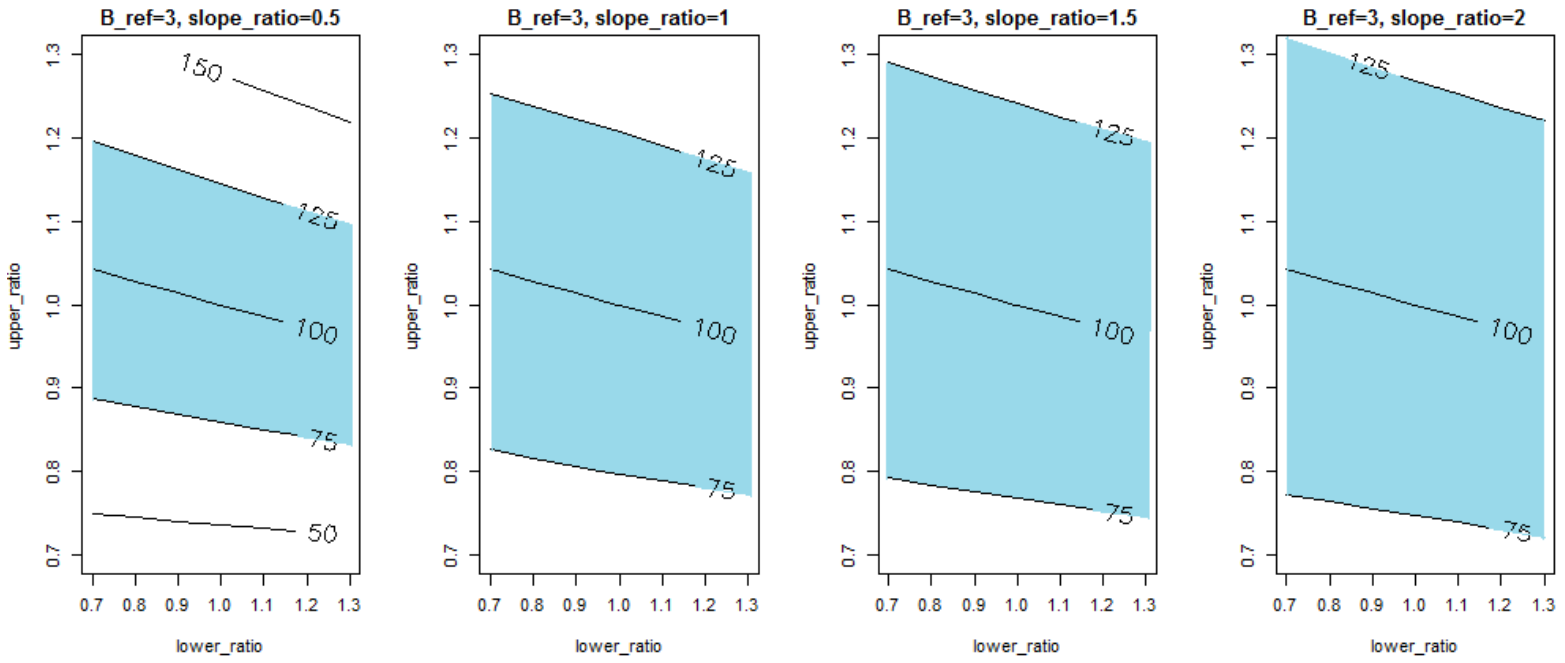
Simulation Study: Impact of Non-Parallelism on Potency Estimation – Results (B=2)

Slope ratio is 0.5, 1.0, 1.5, and 2.0



Simulation Study: Impact of Non-Parallelism on Potency Estimation– Results (B=3)

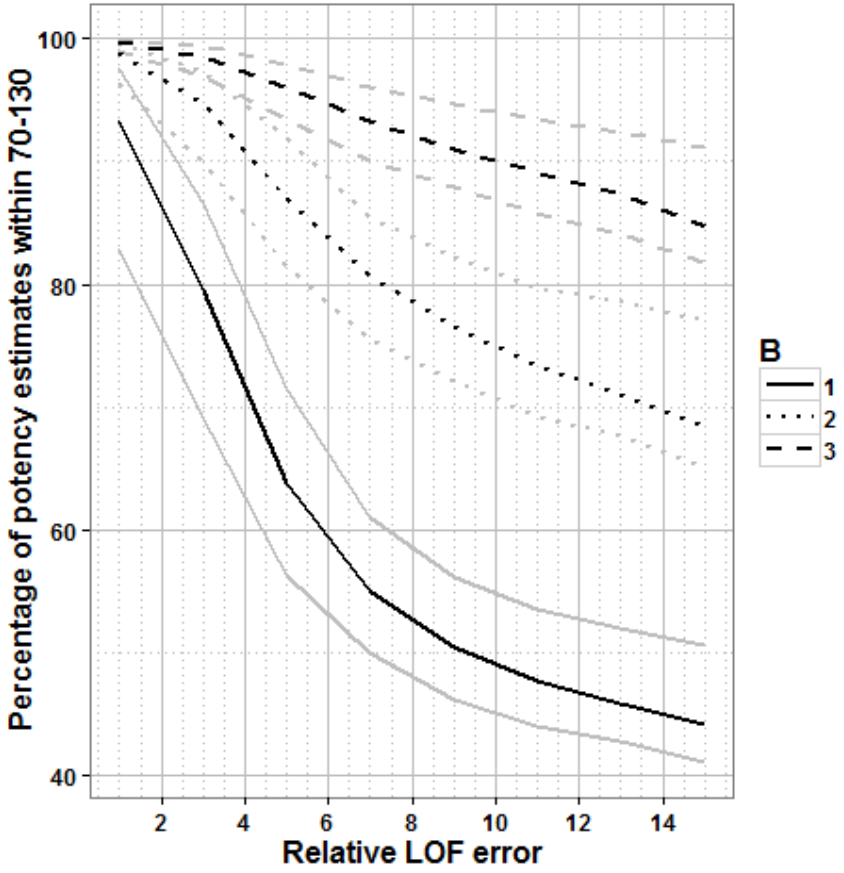
Slope ratio is 0.5, 1.0, 1.5, and 2.0



Simulation Study: Impact of Non-Parallelism on Potency Estimation – Summary

- The range of 75%-125% contours of potency
  - wider as the reference curve gets steeper.
- Upper ratio has more significant effects on potency estimation than lower ratio and slope ratio, thus needs to be more tightly controlled.
- Similar results were obtained for nominal potency 70% and 130%.

Simulation Study: Impact of Relative LOF Error on Potency Estimation



Simulation study results:

The median (black lines) and 5%, 95% percentiles (surrounding grey lines) of the probability of having relative potency within 70-130% are plotted separately for dose-response curve slope factor B of 1, 2 and 3.