

EMA Workshop

Case Study on Analytical Similarity

José G. Ramírez, Ph.D.
Amgen Inc.

EMA Workshop “Draft Reflection Paper on statistical methodology for the comparative assessment of quality attributes in drug development”

3-4 May 2018



**This is a joint industry presentation on
behalf of the trade associations shown**

Outline

- Analytical Similarity
- Herceptin ADCC Data Based on Kim et al (2017)
- Comparing RP vs BP Means
- Quality Range
- How do we define statistically similar?
- Summary

Analytical Similarity

- “Biosimilar candidate has ‘a highly similar’ quality profile as compared to the reference medicinal product.”¹
- “The goal to demonstrate equivalence is the focus in the biosimilar setting.”¹
- “Test of Mean Equivalence (FDA Tier 1) is typically recommended for quality attributes with the highest risk ranking.”²

¹Draft Reflection Paper on statistical methodology for the comparative assessment of quality attributes in drug development

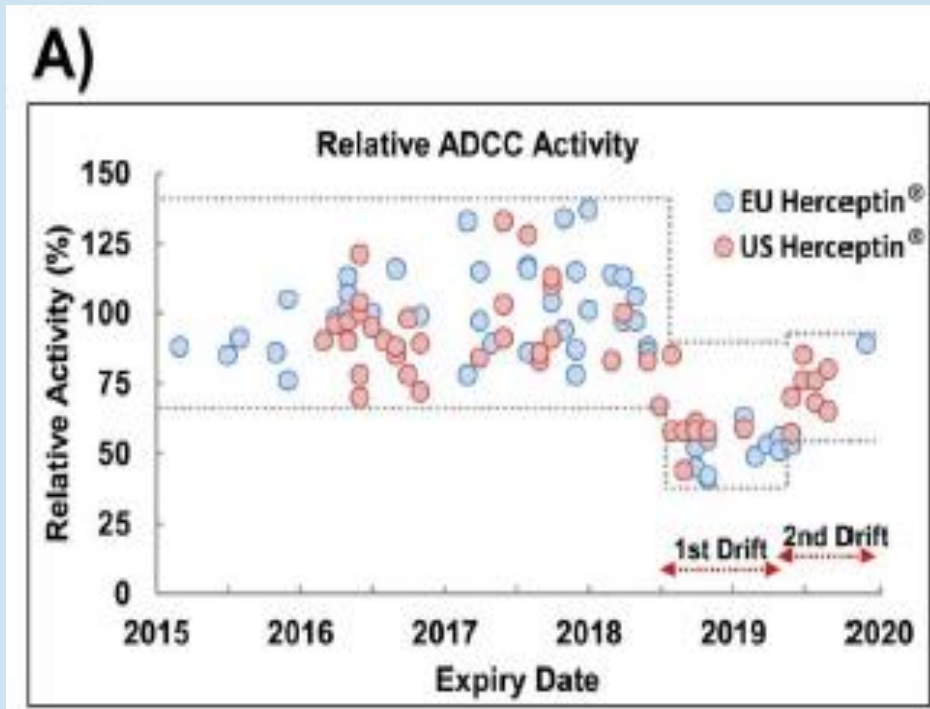
²FDA Statistical Approaches to Evaluate Analytical Similarity Guidance for Industry

Outline

- Analytical Similarity
- Herceptin ADCC Data Based on Kim et al (2017)

Analytical Similarity Example

Simulated ADCC Data Based on Kim et al (2017)

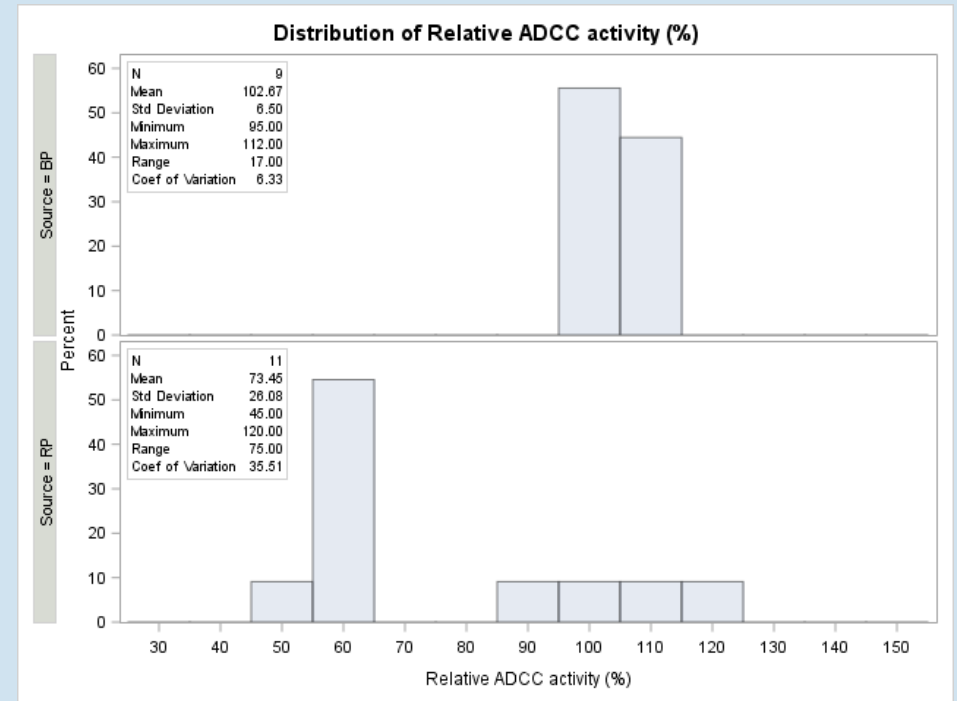
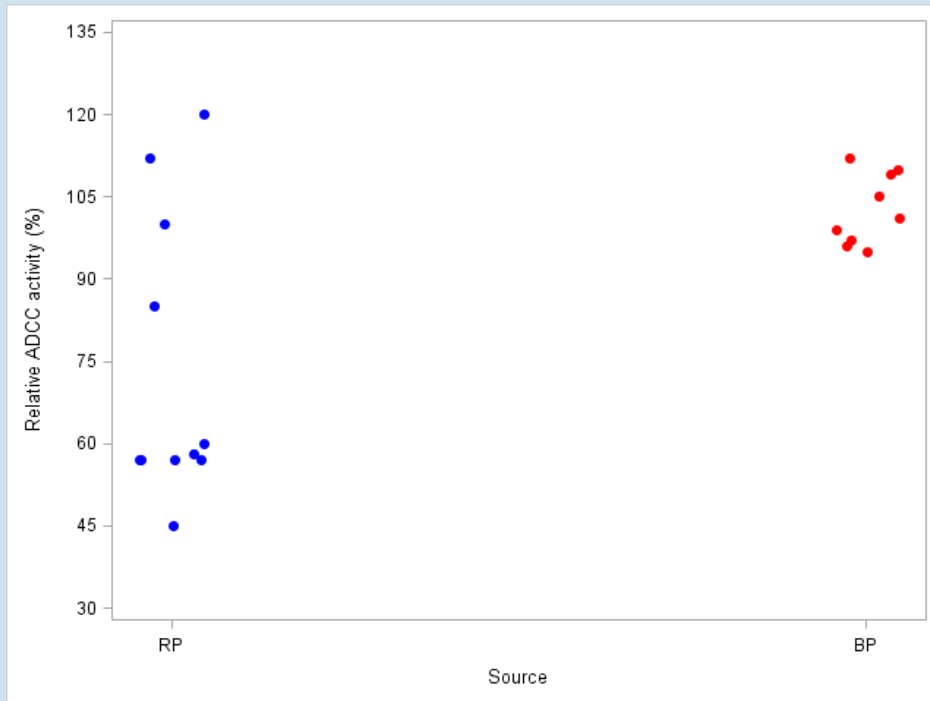


- Reference Product (RP)
 - Random
 - Mimics US Herceptin
- Biosimilar Product (BP)
 - Random
 - Centered on first RP group

Kim et al (2017). Drifts in ADCC-related quality attributes of Herceptin® Impact on development of a trastuzumab biosimilar

Analytical Similarity Example

Simulated ADCS Data Based on Kim et al (2017)



Outline

- Analytical Similarity
- Herceptin ADCC Data Based on Kim et al (2017)
- Comparing RP vs BP Means
 - equivalence test

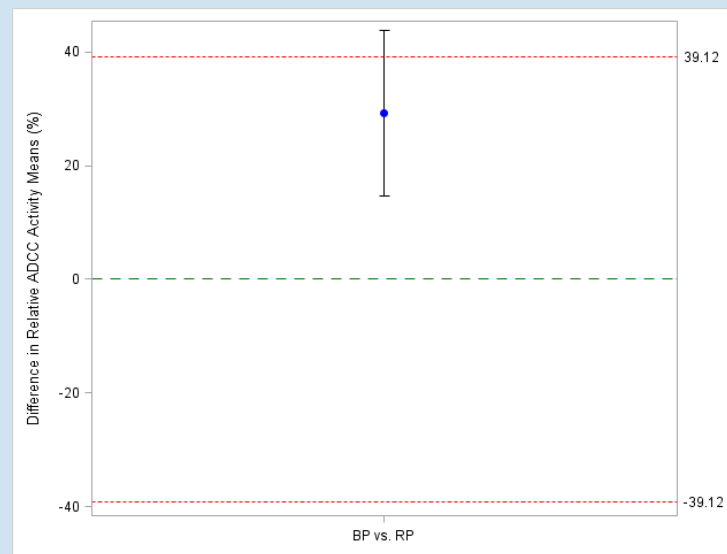
Analytical Similarity Example

Test of Mean Equivalence for Relative ADCC

- EAC = $\pm 1.5 \times \text{RP Std. Dev.}$
 - $\pm 1.5 \times 26.08 = \pm 39.12$

Label	BP Mean	RP Mean	Means Difference	90% LCL	90% UCL	EAC
BP vs. US	102.67	73.45	29.22	14.62	43.80	± 39.12

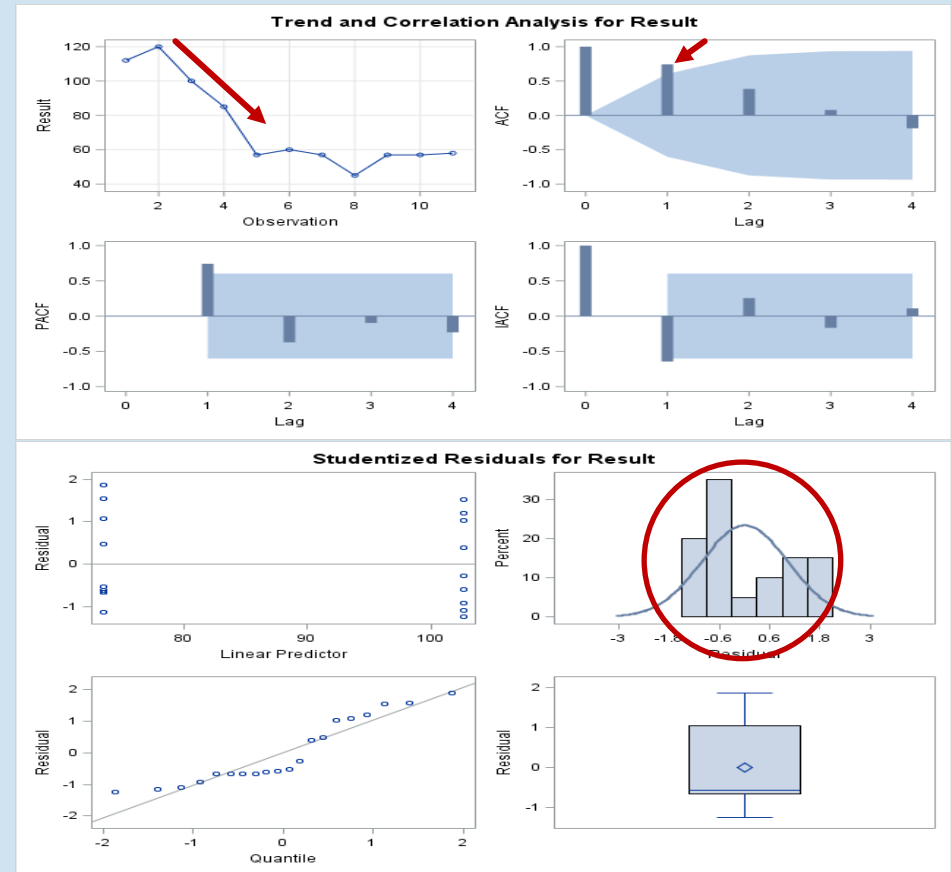
- BP not equivalent to RP



Means are not equivalent to ± 39.12

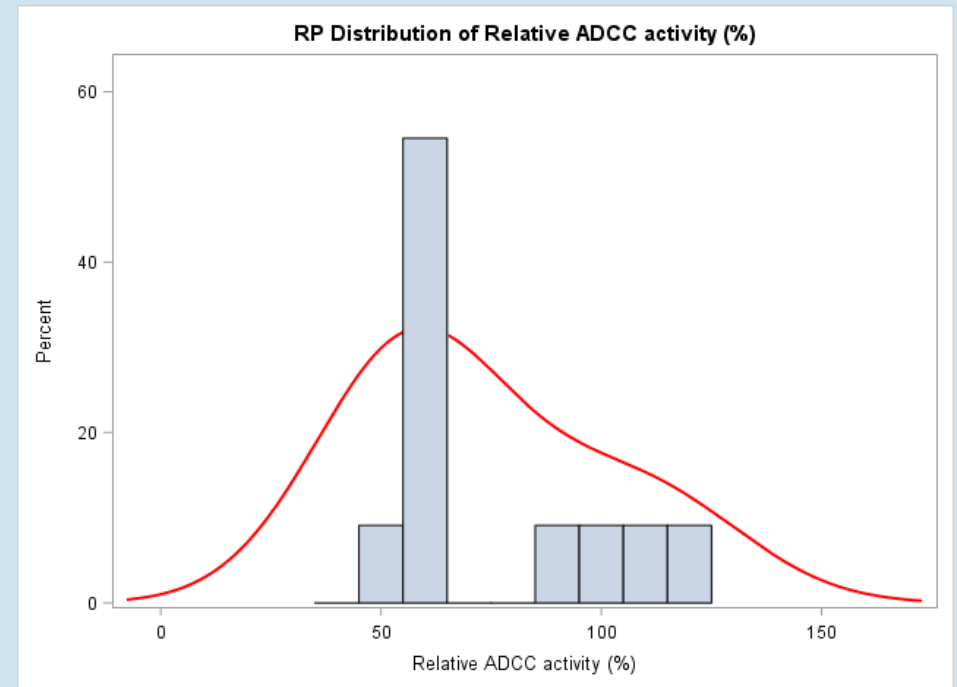
Equivalence Test Assumptions Check

- Downward shift in RP data
 - Manufacturing changes?
- Lag 1 Autocorrelation
 - Same DS batch?
- Residuals not normal
 - Bimodal distribution?



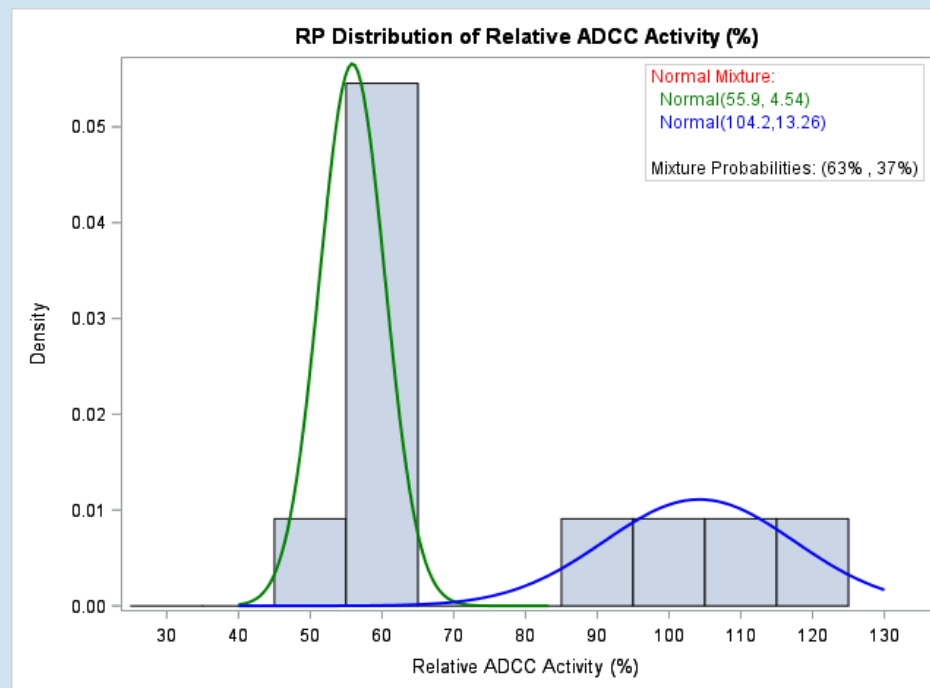
Equivalence Test Assumptions Check

- RP distribution skewed?
 - Mean (73.45) > median (58)
 - Std. Dev. = 26.08



Equivalence Test Assumptions Check

- RP distribution skewed?
 - Mean (73.45) > median (58)
 - Std. Dev. = 26.08
- RP distribution bimodal?
 - Mean_{RP1} = 55.9
 - Mean_{RP2} = 104.2
 - Mean_{BP} = 102.65



Data cannot be approximated by a unimodal distribution

Do We Meet Equivalence Test Assumptions?

- ✗ Samples are random and representative of each population
- ✗ Observations are independent and identically distributed (*i.i.d*)
- ✗ Data can be approximated by a normal (unimodal) distribution
 - Skewed or **bimodal distribution**
- ✗ Homogeneous variances
 - **Variances are different**

Assumptions Are Not Met

Why Do We Fail the Test of Mean Equivalence?

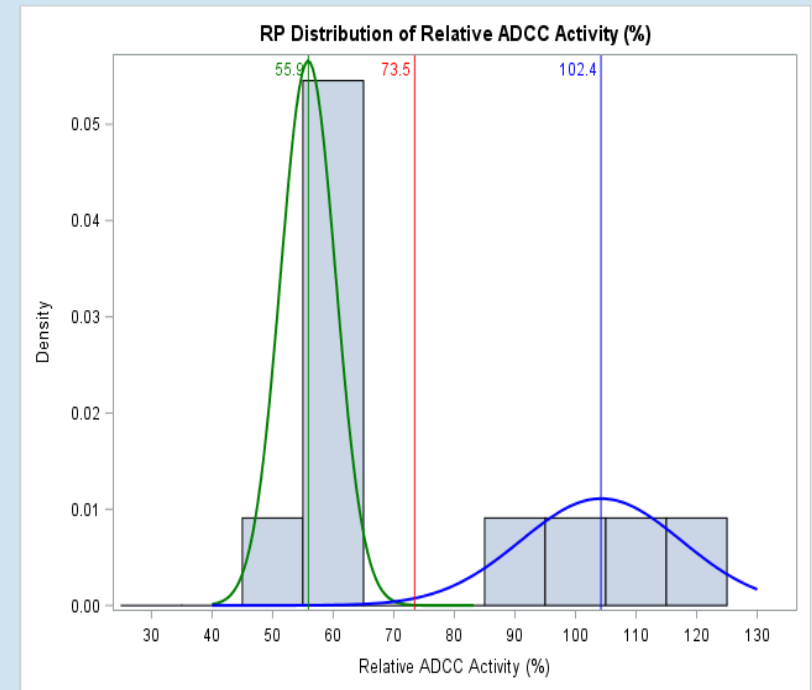
Bimodal Distribution

Equivalence of means

- Which RP mean?
- Overall average?
 - Can be too low or too high

Variances are different

- $\text{Std Dev}_{\text{RP}}=26.08 \gg \text{Std Dev}_{\text{BP}}=6.5$
 - Increases confidence interval width
- Large $\text{Std Dev}_{\text{RP}} \rightarrow$ Inflated EAC
- Does this matter?



Manufacturing changes can induce multimodal distribution

Outline

- Analytical Similarity
- Herceptin ADCC Data Based on Kim et al (2017)
- Comparing RP vs BP Means
- **Quality Range**

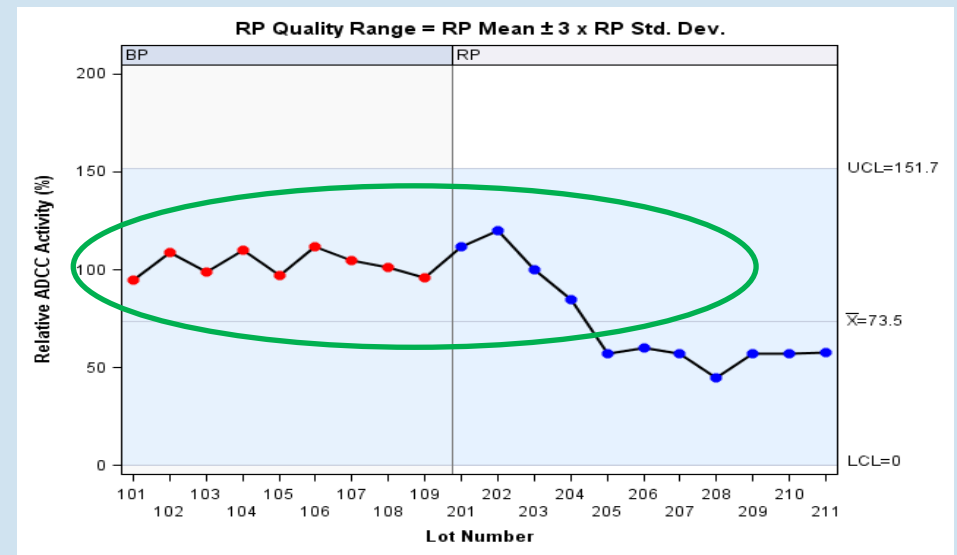
But the BP data is within the RP Quality Range...

RP quality range

- RP mean $\pm 3 \times$ RP std. dev.
 - 0 (−4.8) to 151.7

BP data 100% inside quality range

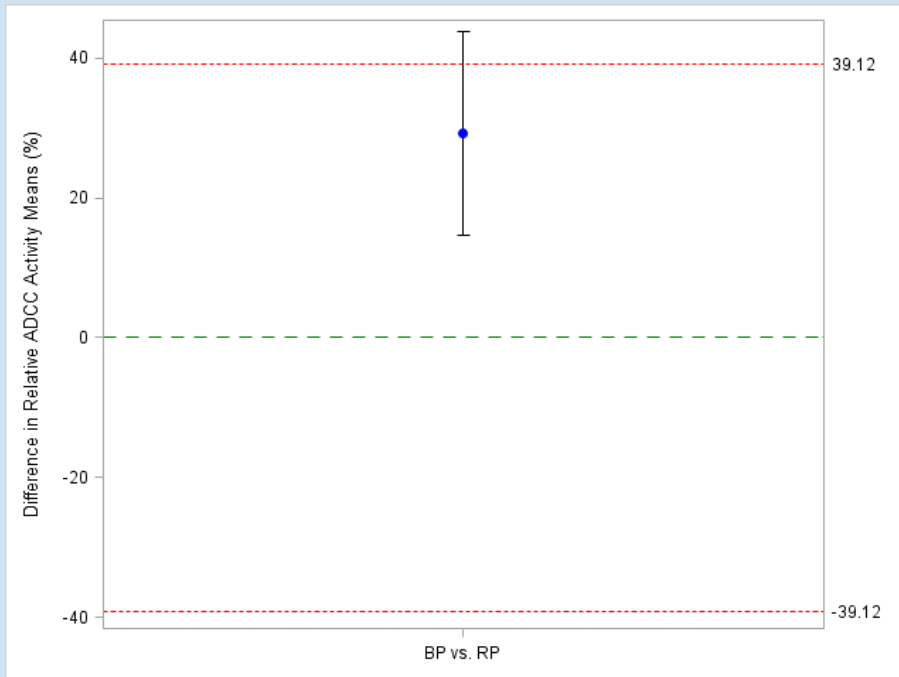
- Same performance as RP



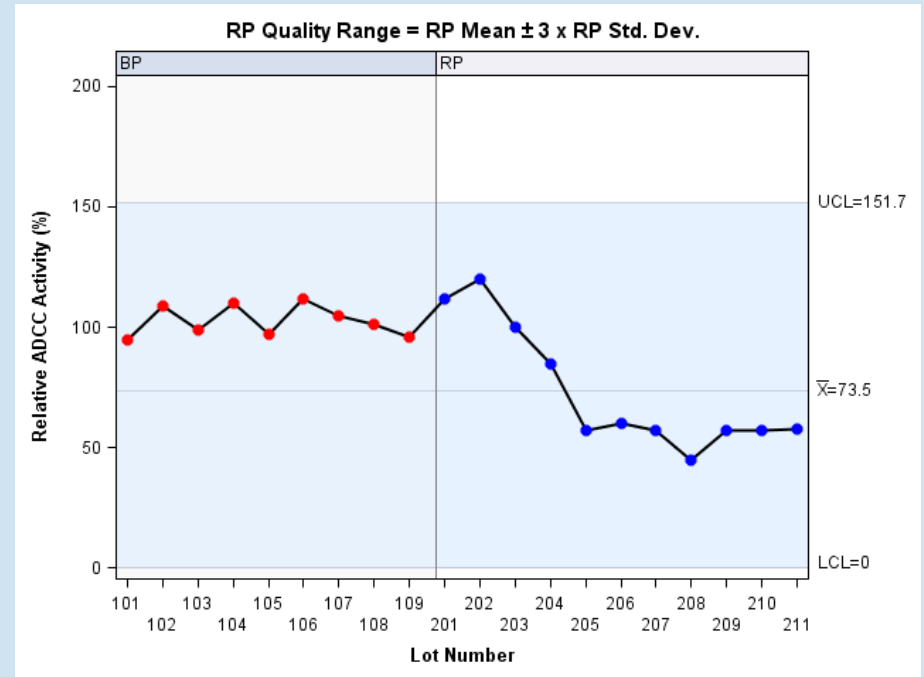
Quality range “represents” specification limits

Equivalence Test or Quality Range?

Compares average performance



Compares individual performance



Means vs individual values: not the same comparison

Outline

- Analytical Similarity
- Herceptin ADCC Data Based on Kim et al (2017)
- Comparing RP vs BP Means
- Quality Range
- How do we define statistically similar?

How do we define statistically similar?

- Manufacturing changes, process improvements, over time can induce a multimodal distribution.
 - Tests assumptions may not hold
 - Several average values are possible
- Biosimilar is “similar” to reference product
 - To which mean?
 - Within a quality range?
 - Do differences in variance matter?

We want the RP and BP distributions to be “similar”

Summary

- Statistical tests depend on assumptions
 - Need to verify assumptions to insure test results are reliable
- How do we define statistically similar?
 - Distance between means?
 - Individual values within a “quality range”?
 - We want to show that the RP and BP distributions are “similar”
- Manufacturing changes can induce multimodal distribution
 - Several means to choose from
 - Overall mean too low or too high
 - Increases overall variance

Totality of evidence: One statistical test is not enough

Acknowledgements

- Franz Innerbichler, Novartis