Assessment of Analytical Biosimilarity: The Objective, the Challenge, and the Opportunities

Bruno Boulanger and EFSPI team
PharmaLex Statistical Solutions

Topics

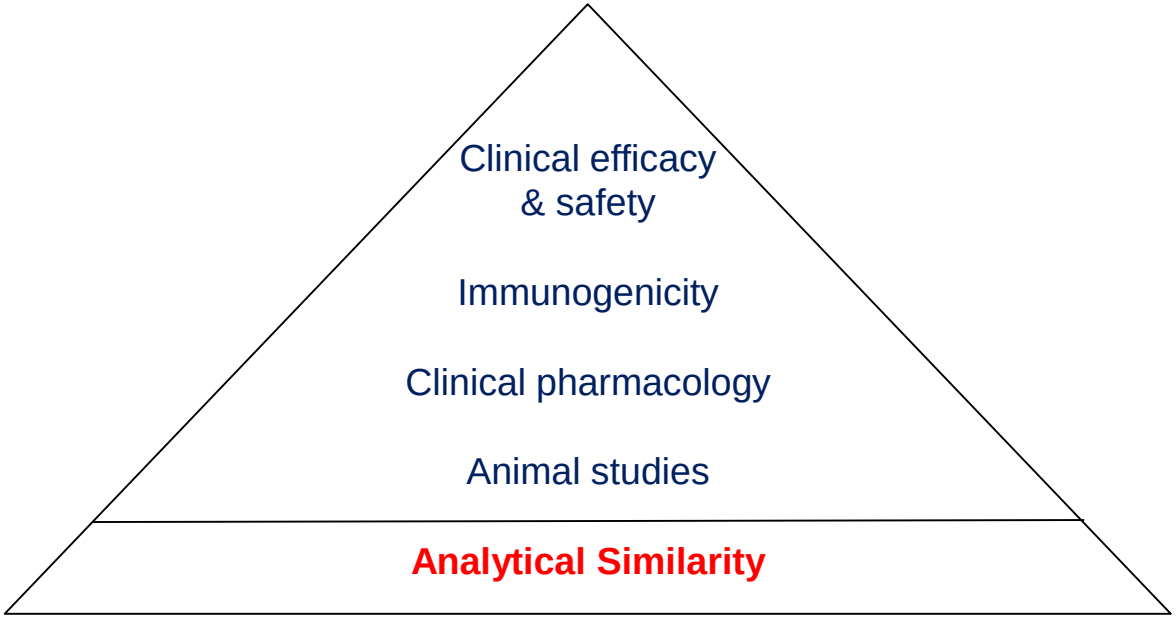
- EFSPI Working Group
- Introduction to analytical biosimilarity
  - What is it about?
  - What are the objectives?
- Regulatory perspectives
  - FDA – current status
  - EMA – Reflection paper released
- Statistical challenges
- Proposals
- Comparative operating characteristics
- Value of Bayesian statistics in analytical similarity

EFSPI Analytical Similarity Working Group

- Industry members:
  - Christophe Agut, Sanofi
  - Bruno Boulanger, Arlenda
  - Mike Denham, GSK
  - Piet Hoogkamer, Abbott
  - Martina Kron, Abbvie
  - Jens Lamerz, Roche
  - Timothy Mutsvari, Arlenda

What is Analytical Similarity?

- Step-wise approach to data generation and the evaluation of residual uncertainty
- Totality-of-the-evidence to demonstrate biosimilarity
- Analytical Similarity is the first step to demonstrate similarity of product



What is Analytical Similarity ?

- Analytical similarity generally refers to an assessment of a proposed biosimilar product in comparison to a licensed reference product.
- Manufacturers perform **chemical**, **physical**, and **bioactivity** comparisons
- with **side-by-side** analyses
- of an appropriate **number** of lots of the proposed product and the reference product

➔ A rather large number of Quality Attributes should be compared

➔ A large number of assays should be used

➔ Many lots of reference and test products

Statistical Approaches – FDA Tier Approach

- Different statistical/quantitative approaches are applied to each tier

Tier 1 – Most Critical
Statistical Equivalence Testing (TOST)

H0: $\mu_{Test} - \mu_{Ref} \leq -c$ or $\mu_{Test} - \mu_{Ref} \geq c$
H1: $-c \leq \sigma_{Ref}\ \mu_{Test} - \mu_{Ref} \leq c$

Tier 2 – Moderate Critical
Quality Range Method: mean +/- kσ

Tier 3 – Least Critical
Raw Data/Graphical Comparison

Equivalence small molecules ↔ Analytical similarity biologicals

- There are unambiguous ways to assess the API is the same
- Few assays (structure, dissolution…) ensure this is the same product
- The experimental units are the patients
- The patients are the “nuisance” parameter and the mean is the parameter of interest
- One lot is used and capability to produce the same (future) lots is assumed.

- It is the process that defines the product with its inherent variation
- A large range of diverse assays are required to evaluate the similarity
- The experimental units are the lots
- Both the mean and the variance of the lots are the parameters of interest
- Mean and variance are only estimates the capability of the new process.

Equivalence small molecules ⟷ Analytical similarity biologicals

- Control strategies ensure future lots and units will be within the equivalence limits
- One set of Equivalence Limits is predefined on the average of 2-3 PK parameters
  - Negligeable multiplicity and dependencies between between end-points
- Equivalence limits is on the mean PK parameters because accepted once as “clinically equivalent” (eg 80%-125%)
- **Equivalence Limits on mean parameter**

- The future capability of the new process to meet the limits is key in the evaluation
- There is a large number of Similarity Limits derived from numerous reference lots
  - Need to deal with multiplicity and important dependencies between assays
- How to define clinically justified limits based on a more or less limited number of Reference lots?
- **Similarity limits on individual lots**

Objective of analytical similarity

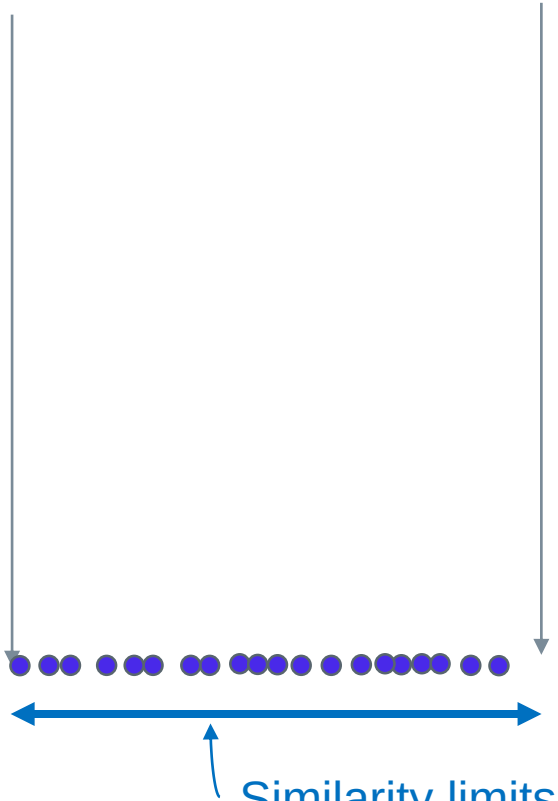
Demonstrate that **proposed new process will produce** lots of Test products that are analytically “comparable” in the future to several lots of Reference products.

- EMA paper puts “similarity” and “comparability” in the same document.
- Biologicals are different from small molecules
  - What defines the product is the process
- Justification of similarity limits easier since directly linked to clinical effect
- The future capability of the new process to meet the limits is the central decision
- Mean and variance are important to estimate to evaluate the capability of the new process

- This is consistent with ICH Q8-Q9 concepts of risks

The mean and variability of new process are estimated with **uncertainty**

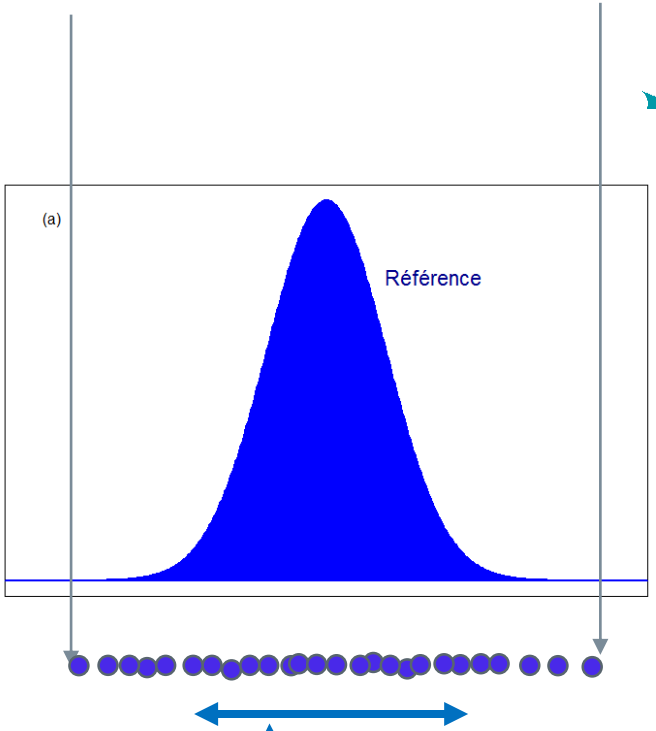
How to define the limits for Analytical Similarity (1)



- Use a number of independent Reference lots that are marketed (nR >= 10)
- These lots have been released and therefore clinically accepted
- The observed variability reflects the Reference process variability (& assay)
- The mean and variance of the Reference process is unknown
- The more Reference lots, the wider the min-max range
- Limits should be established wrt to all Reference products, not actual lots sampled

Similarity limits clinically justified (as in FDA Tier 2)

How to define the limits for Analytical Similarity (1)



- If Mean and Variance of Reference process is known, then using Mean +- 3 SD covers 99% of Clinically accepted lots.

  But mean and variance are estimated with uncertainty

  So using either min-max or Mean+-3SD is not a recommended risk-based approach

  Use of Tolerance interval is recommended to define clinically accepted limits (not specifications)

  It allows to take into account the uncertainty on the estimates

How to define similarity limits with Reference data?

- Mean and variance of Reference estimated with uncertainty
- The β-content γ-Confidence Tolerance Interval (TI) on reference is recommended

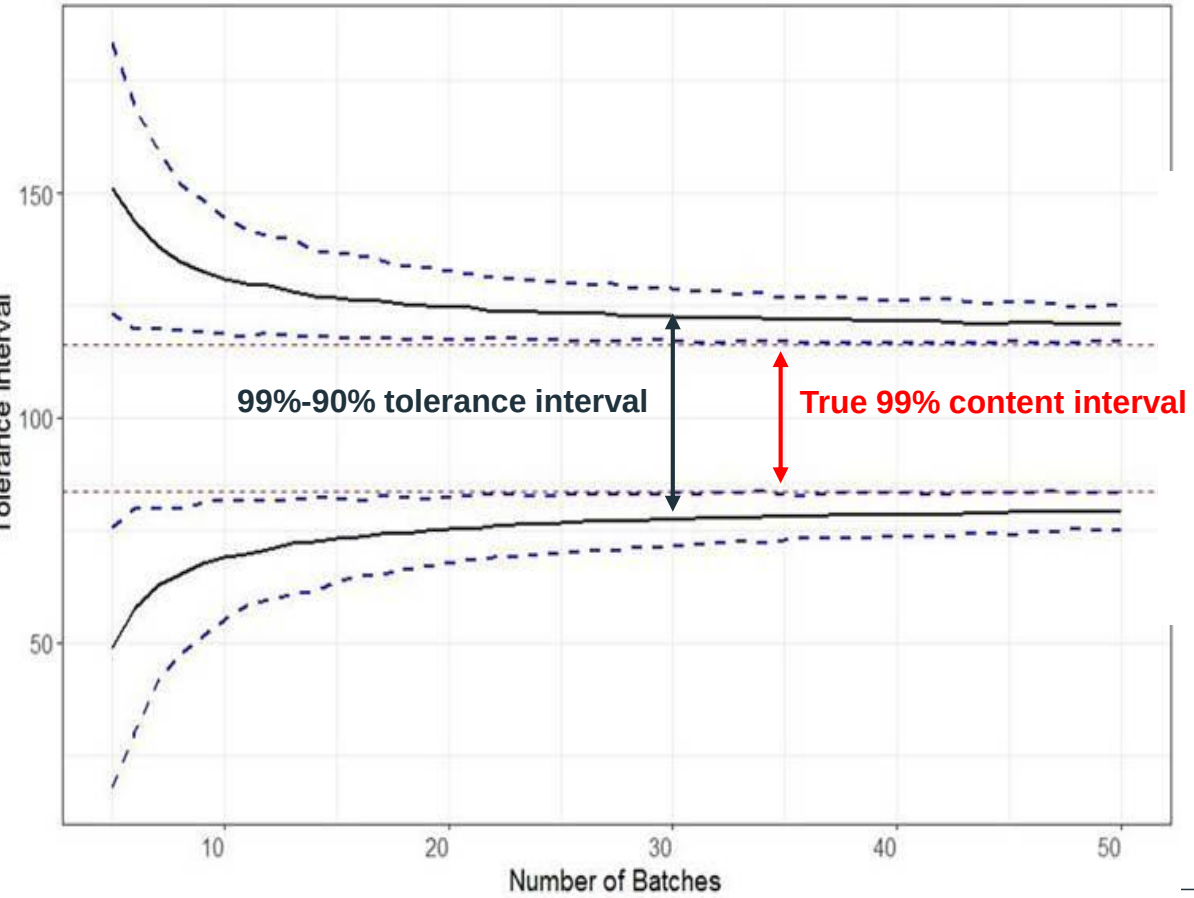
$$P_{\hat{\mu},\hat{\sigma}_X}\left\{P_X\left(\hat{\mu}-k_C\hat{\sigma}_X<X<\hat{\mu}+k_C\hat{\sigma}_X\,\middle|\,\hat{\mu},\hat{\sigma}_X\right)>\beta\right\}=\gamma$$

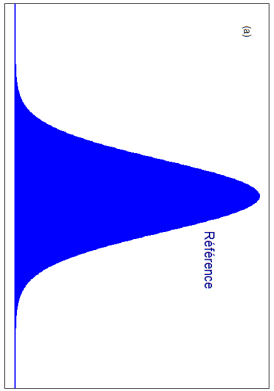
$$k_c=z_{\frac{1+\pi_{\min}}{2}}\sqrt{1+\frac{1}{\hat{\sigma}_X^2}}\sqrt{\sum_{j=1}^{q}c_j^2H_j^2MS_j^2}\sqrt{1+\frac{1}{N_E}}\quad\text{with}\quad \mathrm{H_j}=\frac{1}{F_{1-\gamma;c_j;\infty}}-1$$

- Take into account the uncertainty on the Mean and the Variance
- Better statistical properties than Min and Max
- Both Content and Confidence can be controlled

- A minimum sample size of Reference is recommended to make β-content γ-Confidence Tolerance Interval (TI) relevant for Similarity limits.

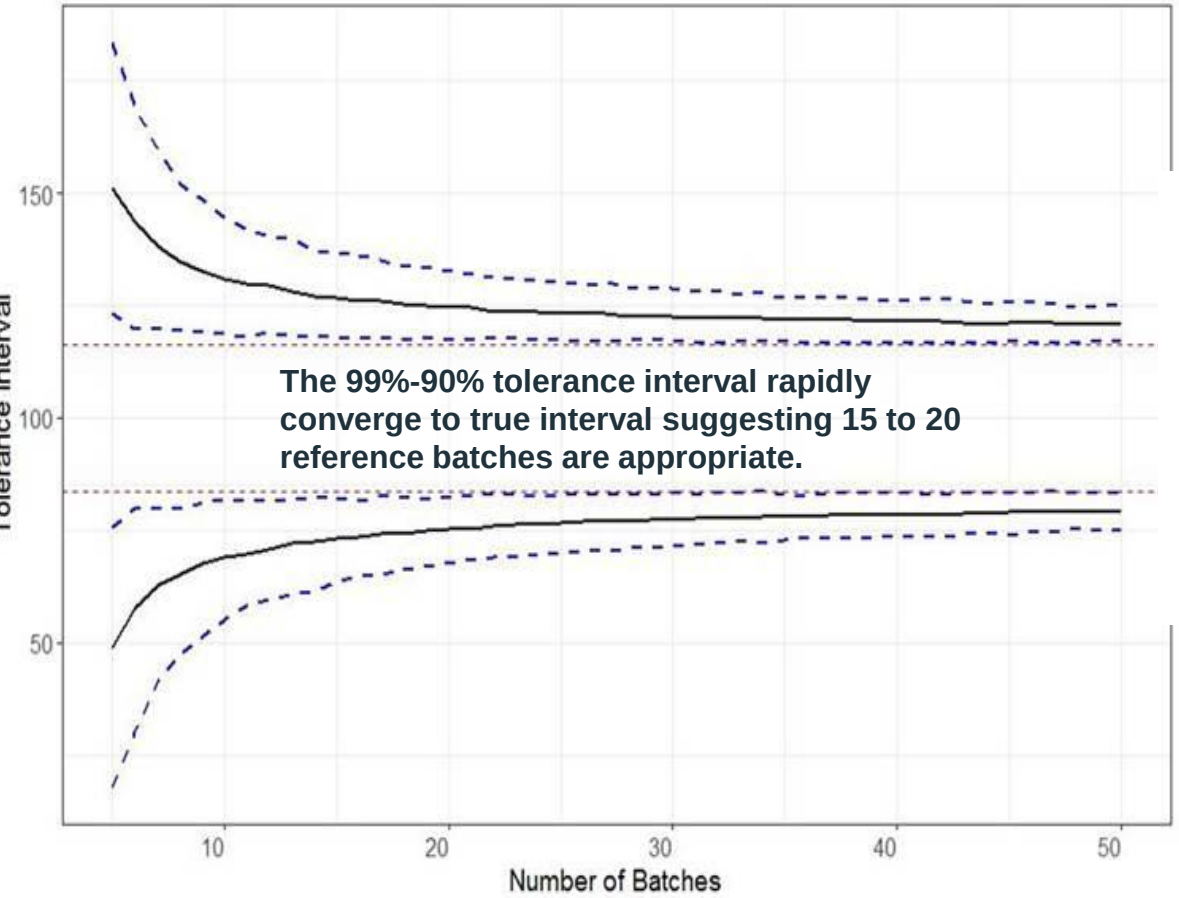
Convergence of Tolerance Interval β=0.99 γ=0.90

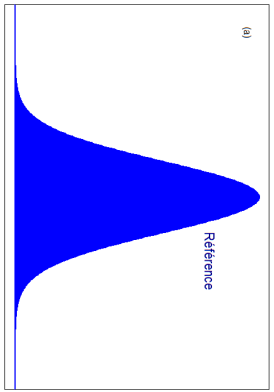




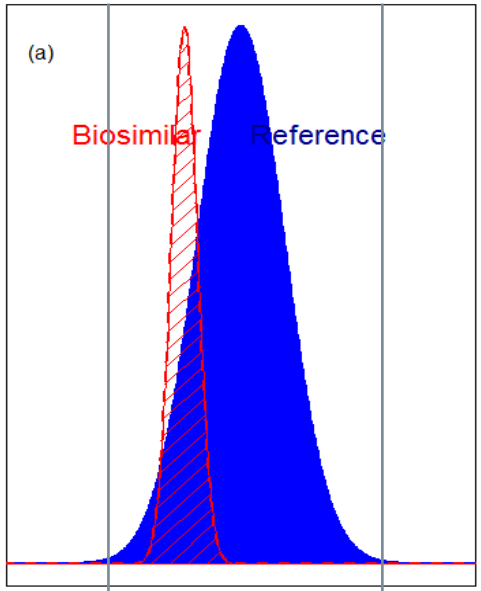
PHARMALEX

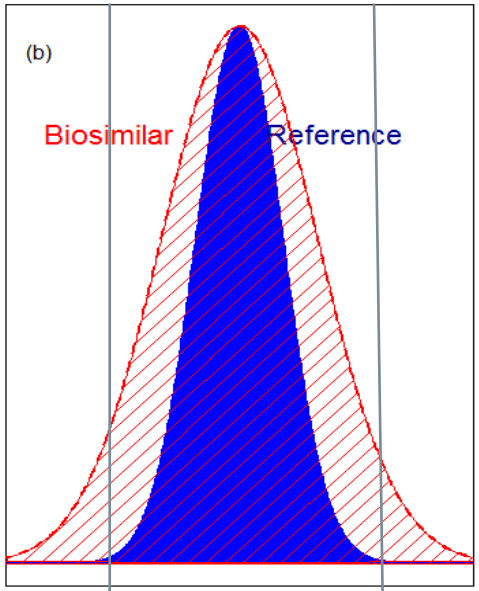
Convergence of Tolerance Interval β=0.99 γ=0.90





How to make a decision for the Test Product ?





Similarity limits clinically acceptable
(given uncertainty on reference parameter estimates)

- For the Test product, the Parameters are also estimated with uncertainty
- The central question is to know the future capability of Test process to make product within the similarity limits.
- Two options are available:
  - Compute the predictive probability to have lots within the similarity limits
  - Ensure the prediction interval is within the Similarity limits
- Using actual observed results doesn't control the risks

How to assess if Similarity is achieved ?

- Test if β-Prediction Interval (PI) of biosimilar is within β-γ-Tolerance Interval (TI) of reference

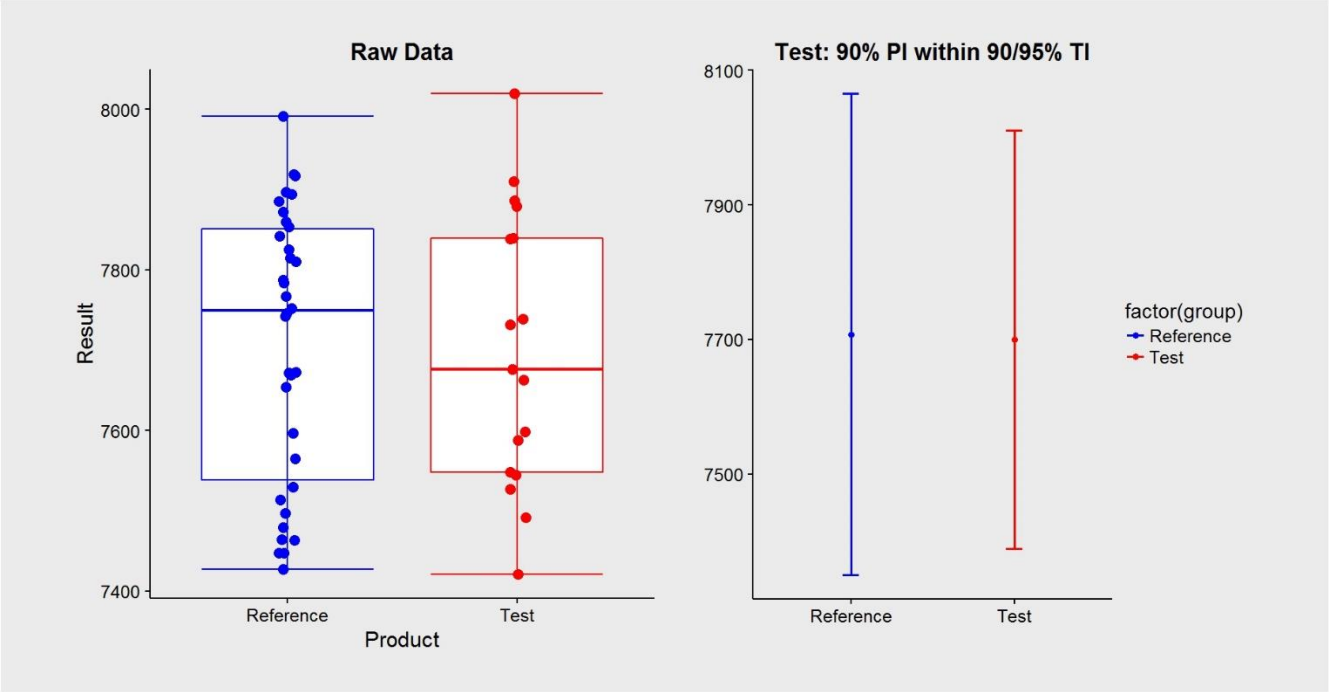
$$p(\tilde{x}|data) = \iint_{\sigma^2\,\mu} p(\tilde{x}|\mu,\sigma^2,data) \times p(\mu,\sigma^2|data)\ d\mu d\sigma^2$$

Integrate over parameter distribution
Meaning that the uncertainty of those performance parameters are integrated into the computation of the risks

- More relevant than using an arbitrary c factor (such as 3!)
- Take into account the variability of Test process (between-lots)
- Take into account uncertainty on means and variability of new process
- Prove that all Test lots will be within the range of Reference lots with some level of confidence even in the future

Proposal - Justification of acceptance limits

- Test if β-Prediction Iinterval is within β-γ-Tolerance Interval



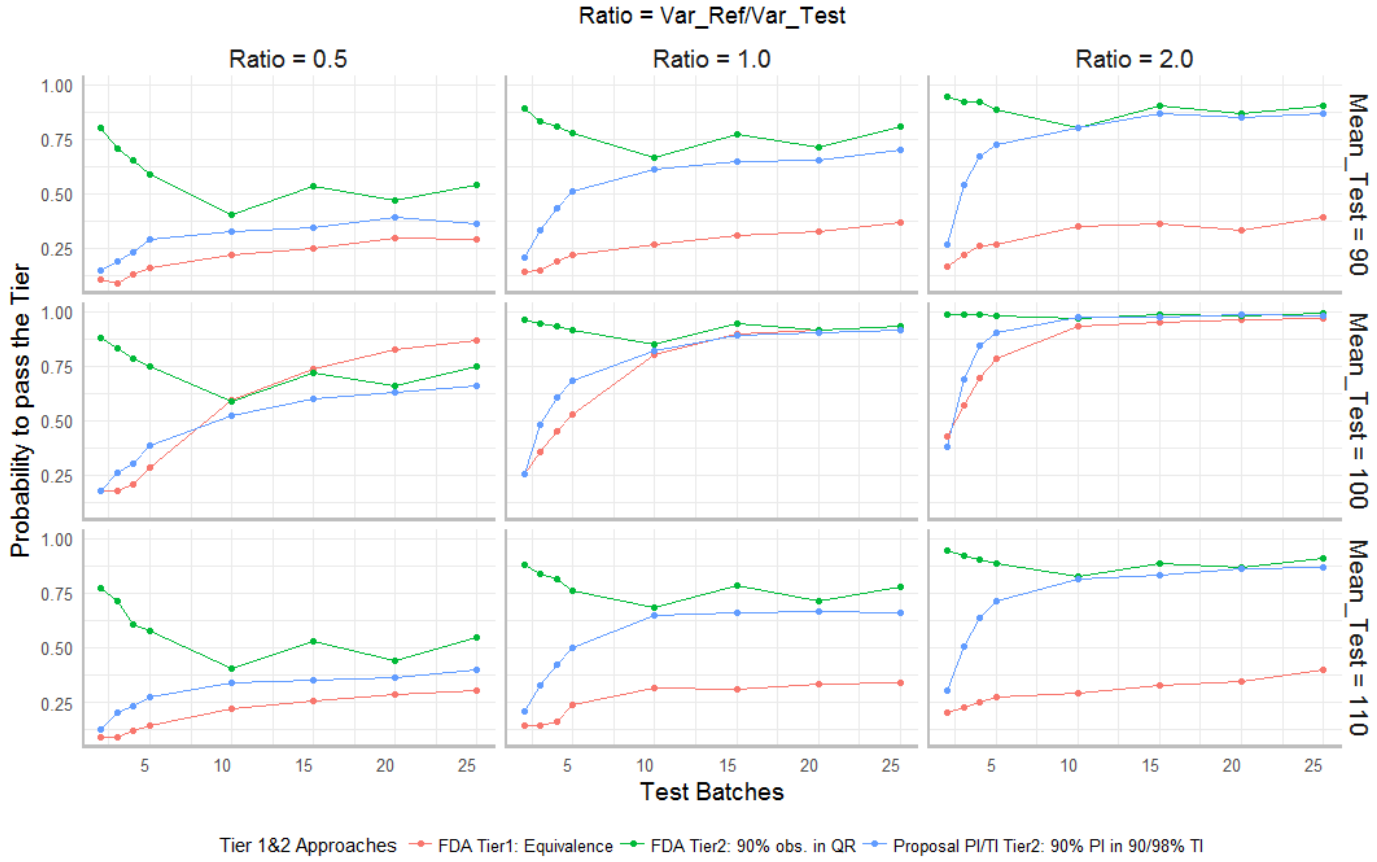
Comparison of Operating Characteristics by Simulations

- Assume
  - Reference Mean = 100, Reference SD = 10
  - # Reference lots is 10

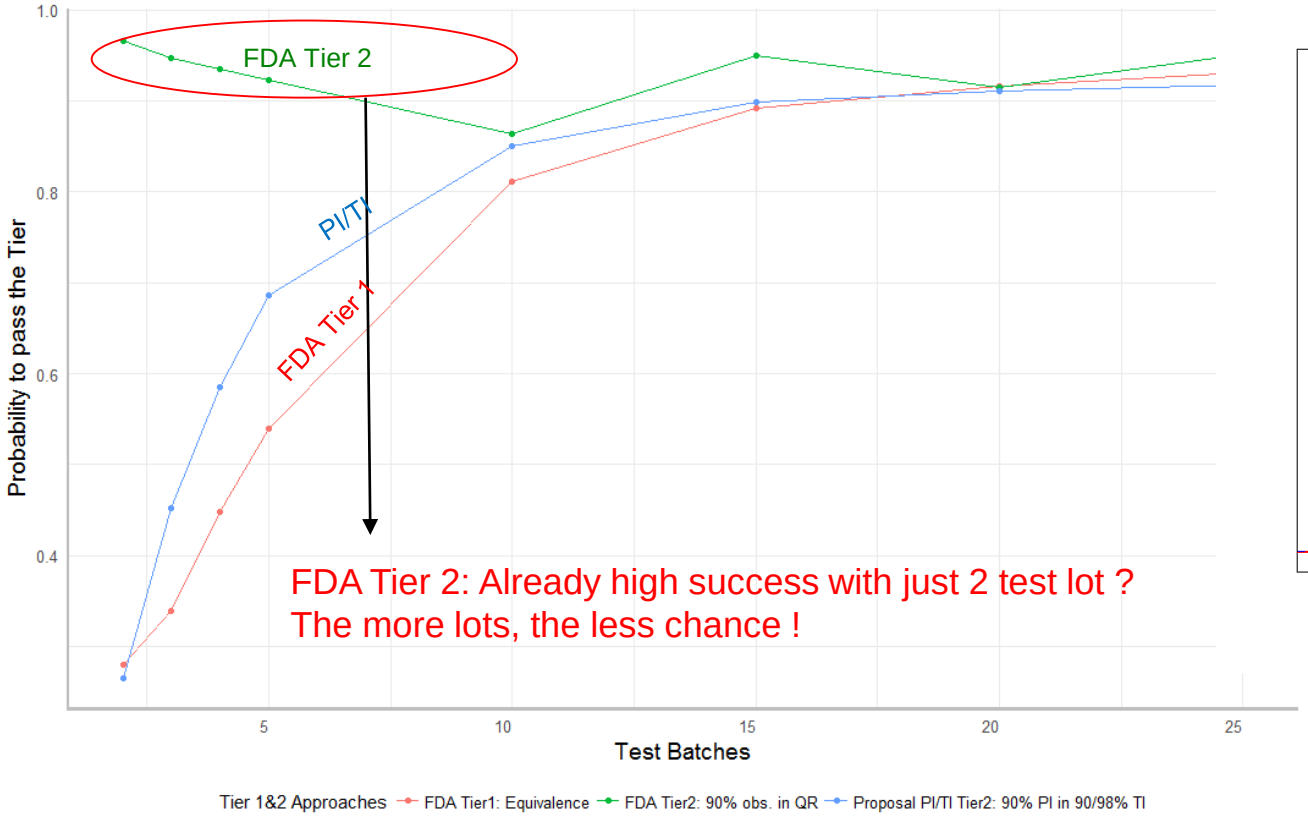
Decision methods:

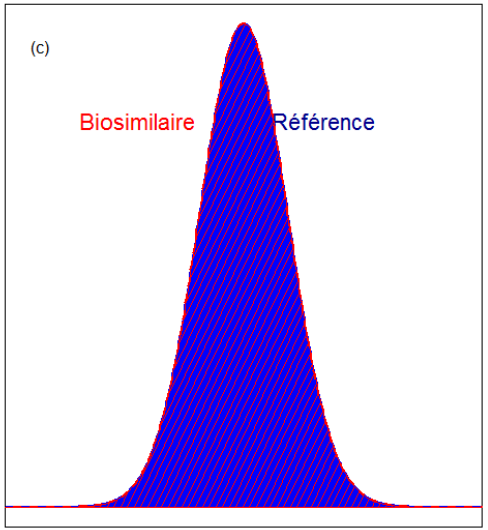
- FDA Equivalence Testing
- FDA 90% lots in +- 3 SD
- Proposal 90% PI within 90/98% TI

Simulation - Results

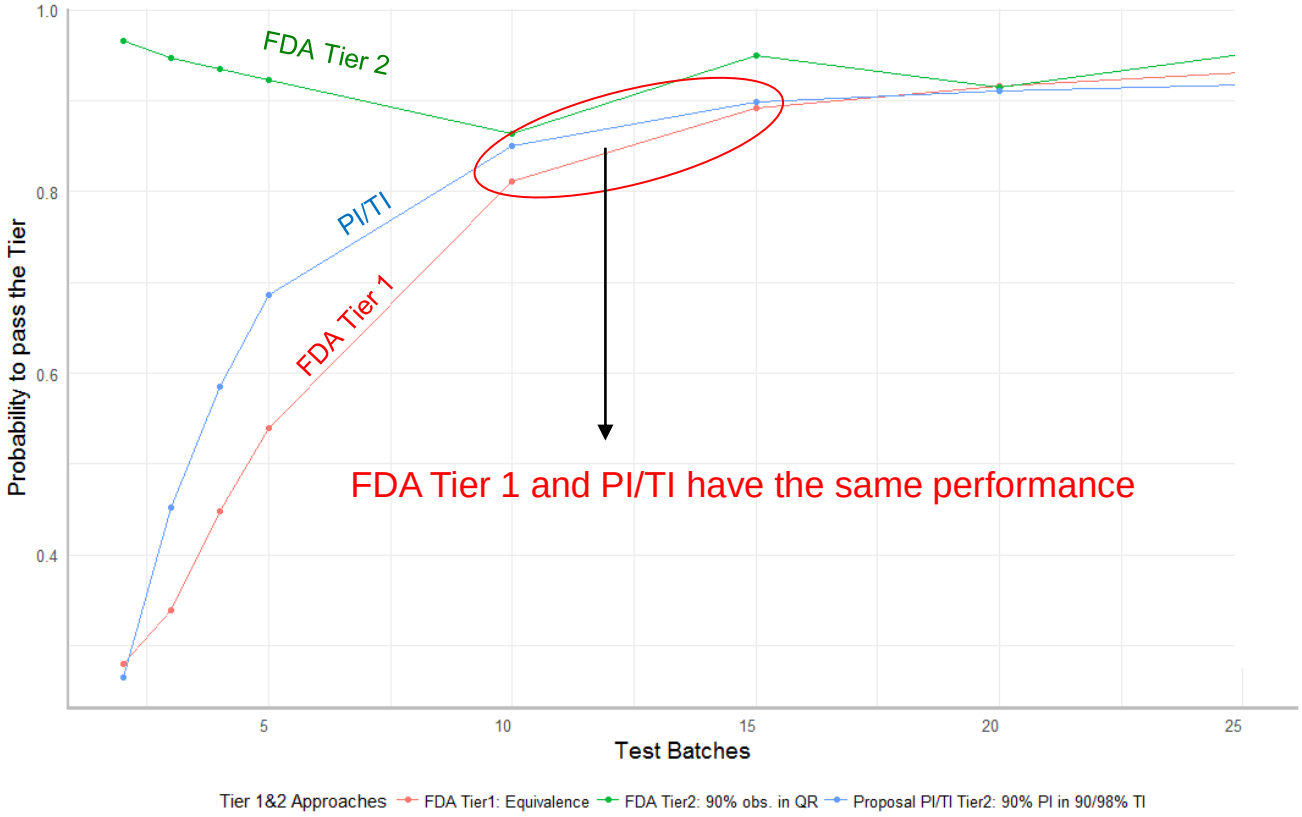


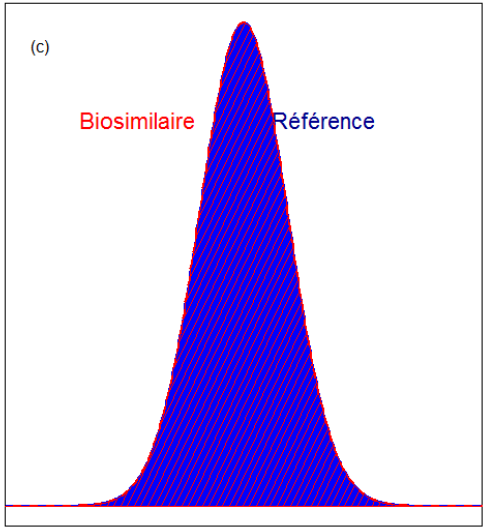
Best Case Scenario: Test = Reference



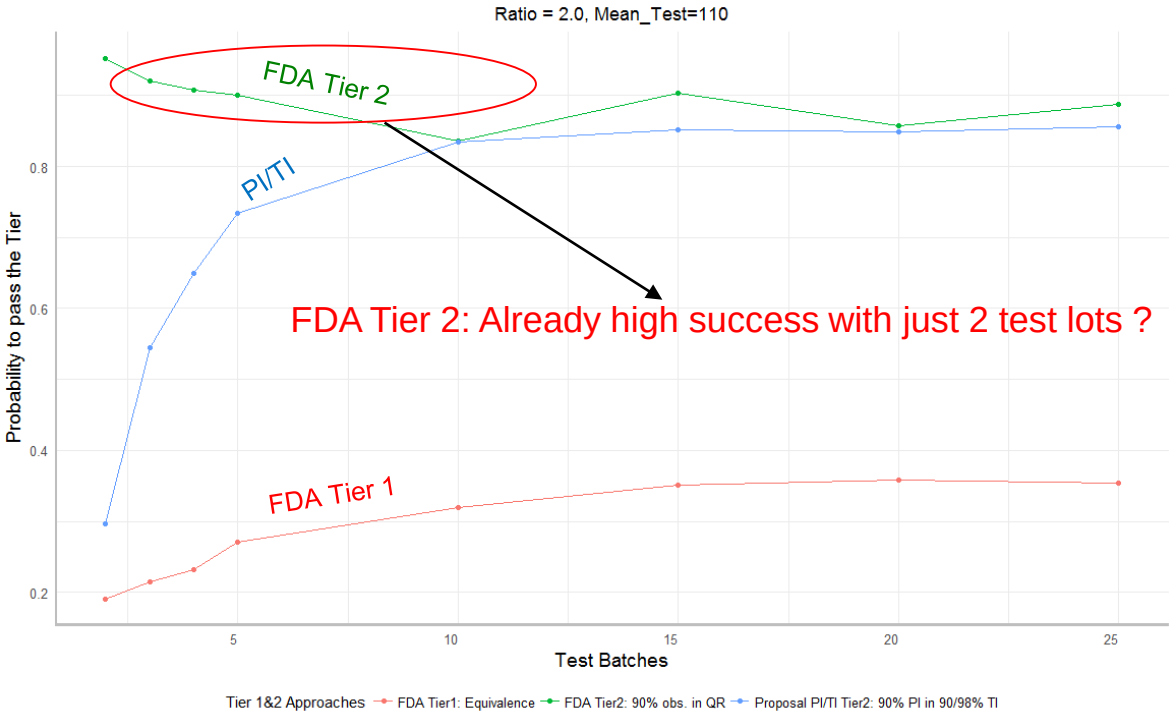


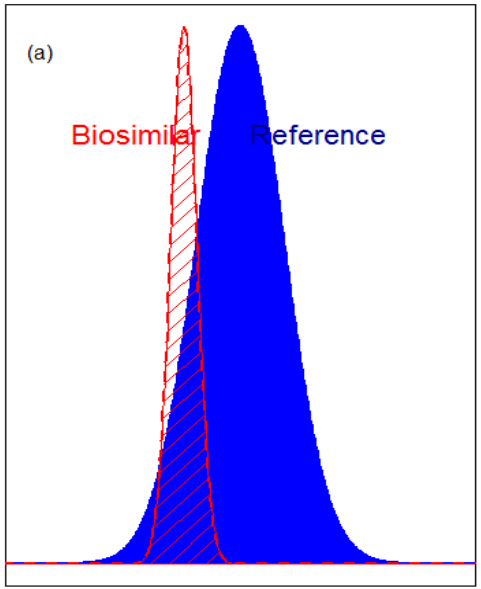
Best Case Scenario: Test = Reference



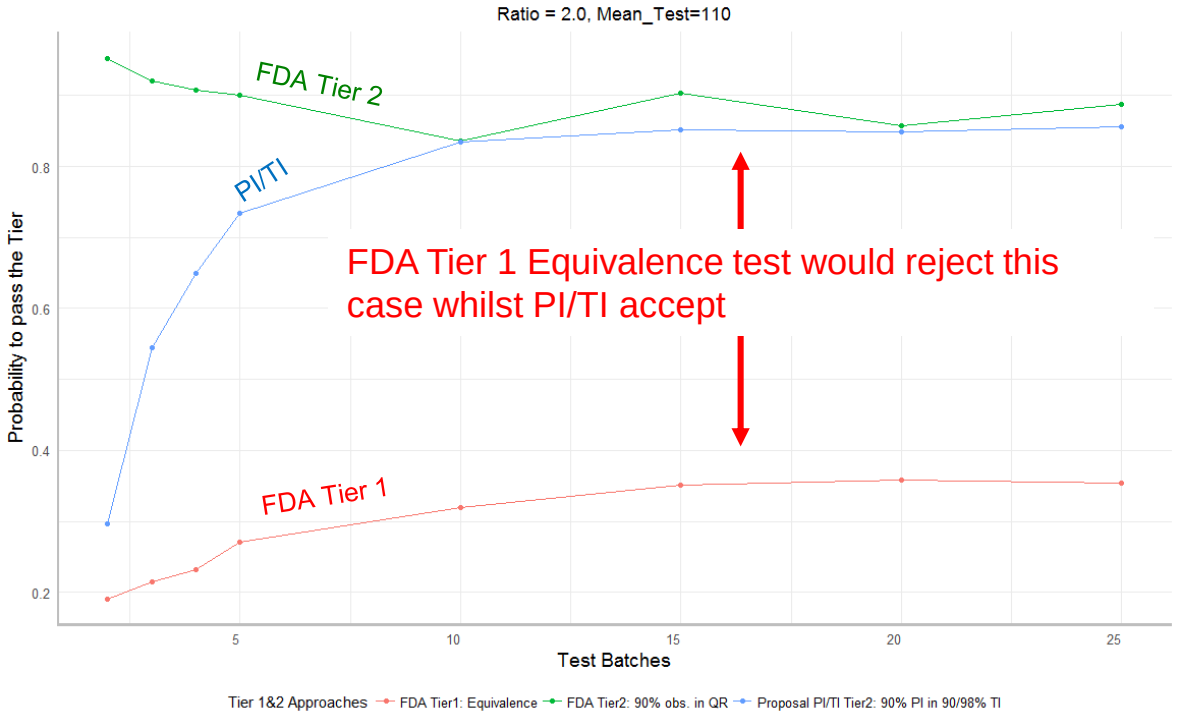


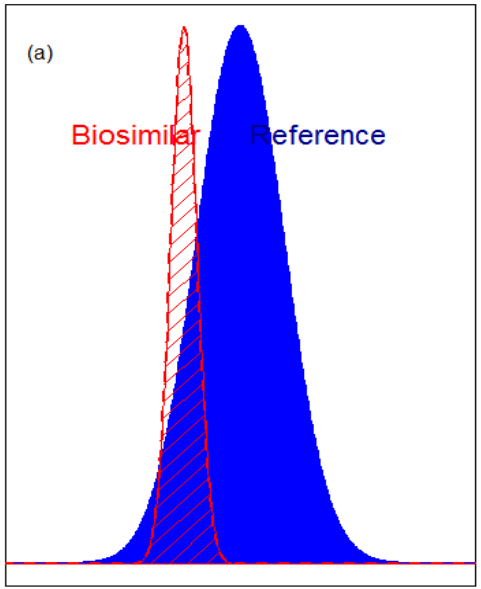
The next scenario



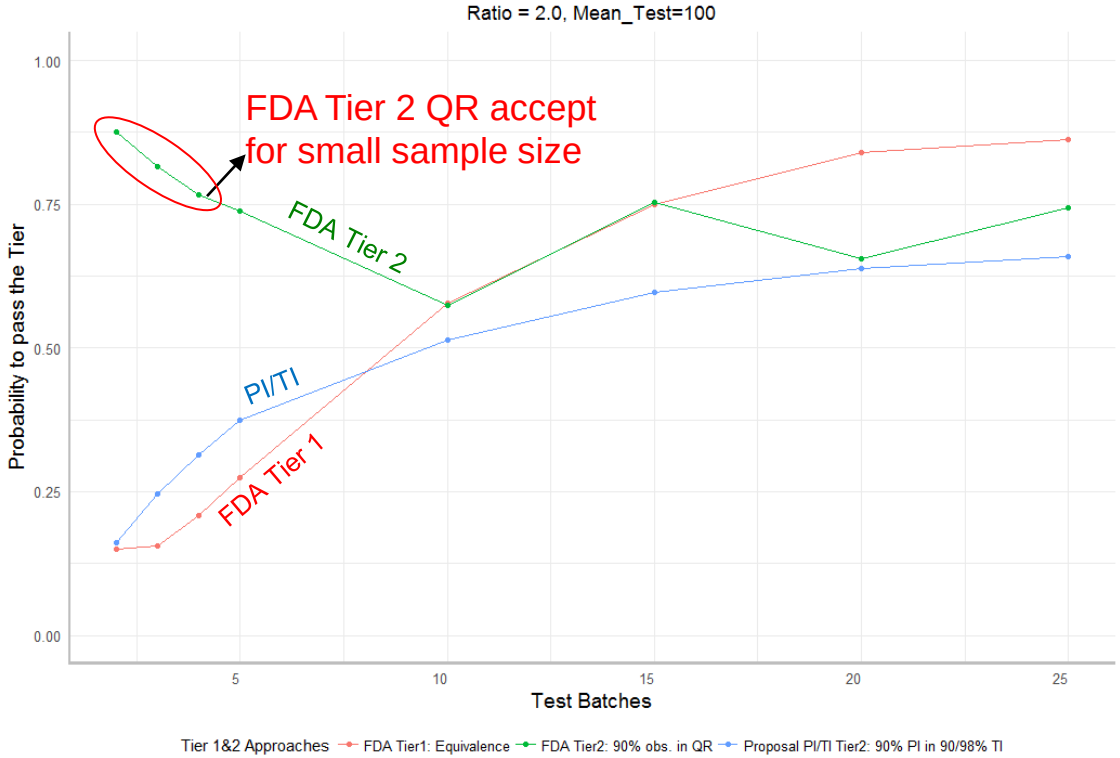


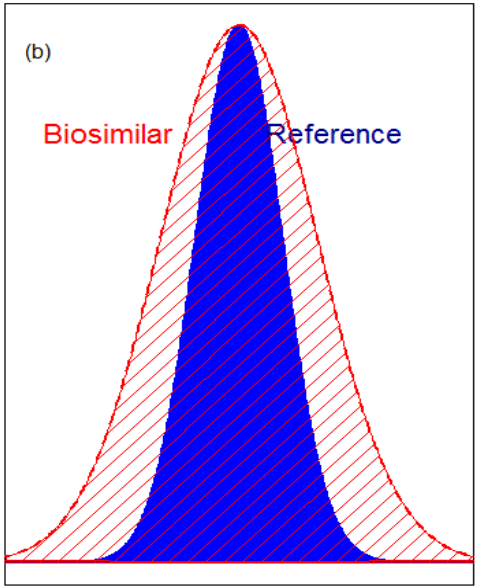
The next scenario



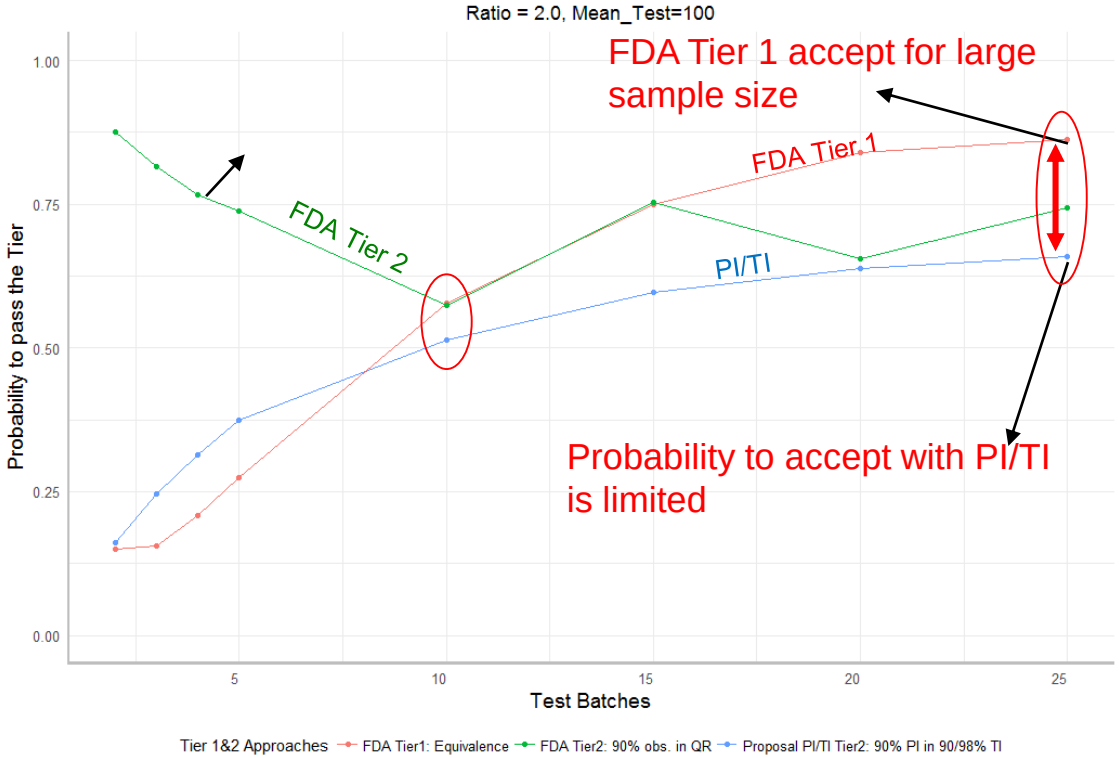


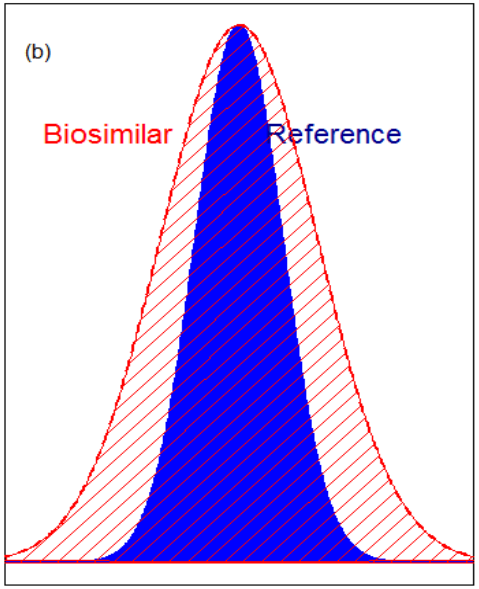
The next scenario





The next scenario





Preliminary conclusions

- What is the question ?

- Prediction and tolerance Interval are showing appropriate operating caracteristics
- FDA Tier 2 approach raises some questions
- FDA Tier 1 Equivalence approach is only about the mean, not the variance
- Multiplicity is not addressed yet !

- ➔ Make it Bayesian

Why make it Bayesian ?

- That's the question:
  - **P(Biosimilar | Data) vs P(Data | assuming not biosimilar)**
  - **what is the predictive probability having future lots within the limits given available data.**
- The Bayesian approach provides naturally the predictive distribution of future observation: it's a function of parameters and their uncertainty.
- The parameter estimation is an intermediate steps to figure out future lots
- The Bayesian approach can easily handle multivariate problems
- Simplification occur drastically when moving from comparison of interval to predictive probability: the number of dimension doesn't matter anymore.

Model - Univariate Case

$$CQA_{Test} \sim N\left(\mu_{Test},\ \sigma^2_{Test} + \sigma^2_{assay}\right)$$ → Model for Biosimilar

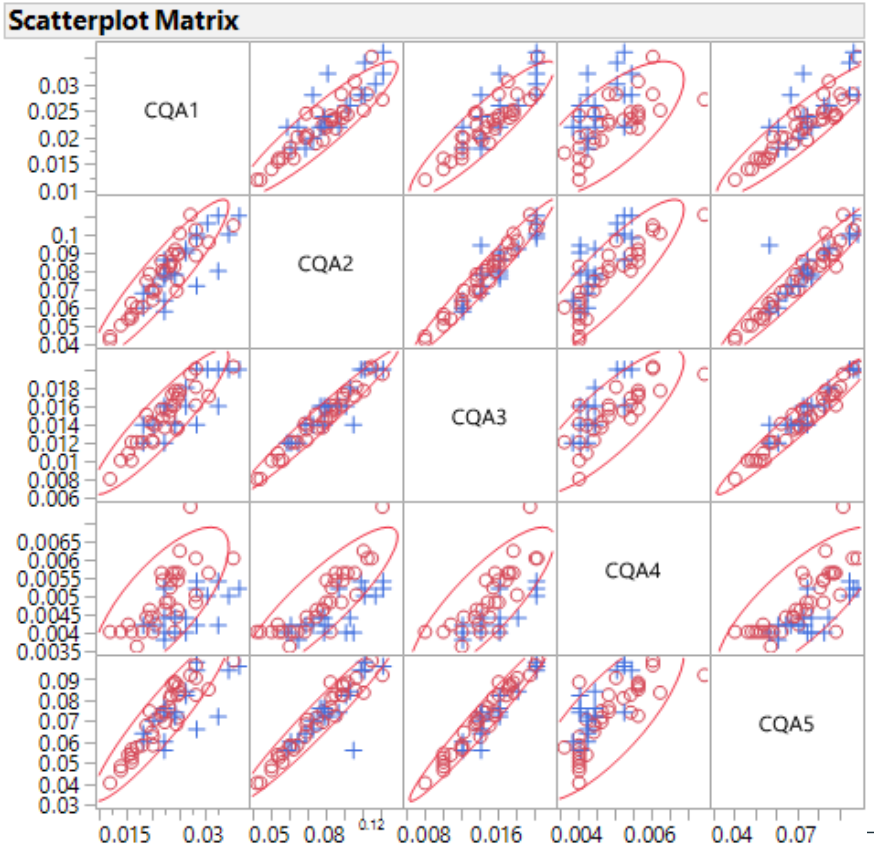
$$CQA_{Ref} \sim N\left(\mu_{Ref},\ \sigma^2_{ref} + \sigma^2_{assay}\right)$$ → Model for Ref

$$\sigma^2_{Test} = \alpha_0 * \sigma^2_{ref}$$ → Test will not extremely different from Ref

$$\alpha_0 \sim \text{Uniform}\,(0.01, 20)$$

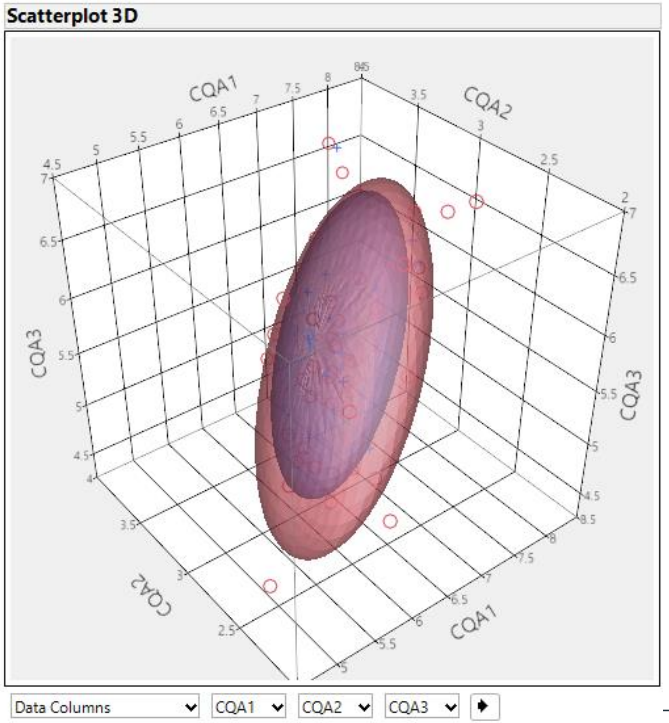
- Possibility to leverage prior for assay variability using historical data
- Directly derive the PI/TI from predictive dist. – future batches will still prove biosimilarity

Multivariate - Exploratory



Multivariate CQA Model – PI/TI Proposal

- Tier 2: Joint Prediction Interval region within joint Tolerance Interval region

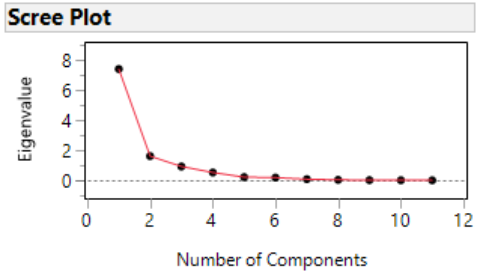


Multivariate – PCA in Analytical Similarity

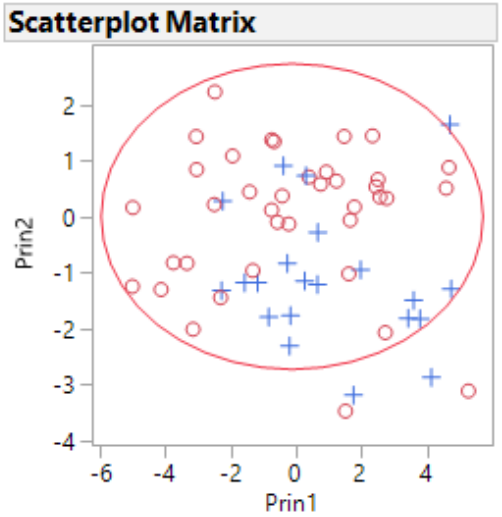
- Build the Principal Component Analysis (PCA) with ref data alone

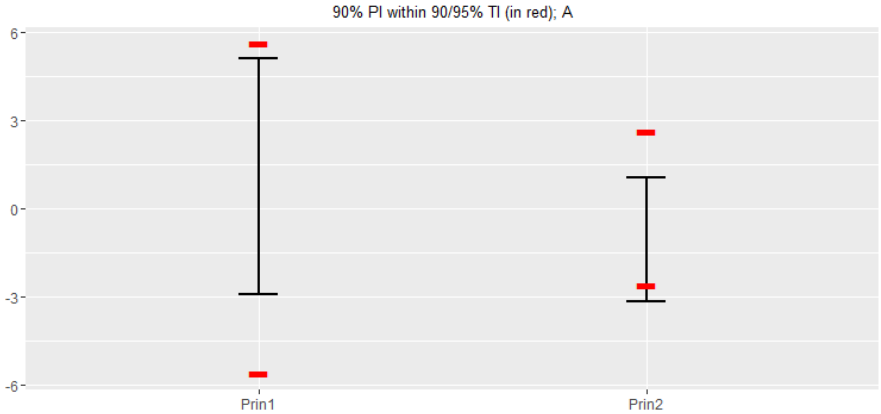
Principal Components: on Correlations

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	7.3964	67.240		67.240
2	1.6122	14.656		81.897
3	0.9191	8.356		90.252
4	0.5217	4.743		94.995
5	0.2175	1.978		96.973
6	0.1825	1.659		98.632
7	0.0810	0.736		99.369
8	0.0356	0.323		99.692
9	0.0174	0.158		99.850
10	0.0108	0.098		99.948
11	0.0058	0.052		100.000



Multivariate – PI/TI proposal on PCA





Conclusions

- What is the objective ?
  - Is it about the “averages” ?
  - Is it about the individual (future) lots ?
- Bayesian approach answers the very question
- Similarity should be proven whatever the future lots and units
- Bayesian methods using the predictive distribution answers the very objective
  - In most cases CQAs are correlated
  - Easily compute the predictive joint probability
  - Informative priors on some parameters can be justified and recommended
- Ensuring future Test products will be biosimilar given current available data!

When SIMILAR is not the SAME!