Arlenda
Advance your business with statistics

Assessment of analytical biosimilarity: the objective, the challenge and the opportunities.

Bruno Boulanger
Arlenda

Basel, 13 September 2016

Agenda

- Working Group in Analytical Similarity
- Regulatory positions FDA and EMA
- Statistical status as of today
- General aim of analytical similarity
- Objective to achieve: what's the question?
- Statistical challenges and opportunities
- Integrating statistics early in biosimilar development
- QbD approach for developing biosimilars

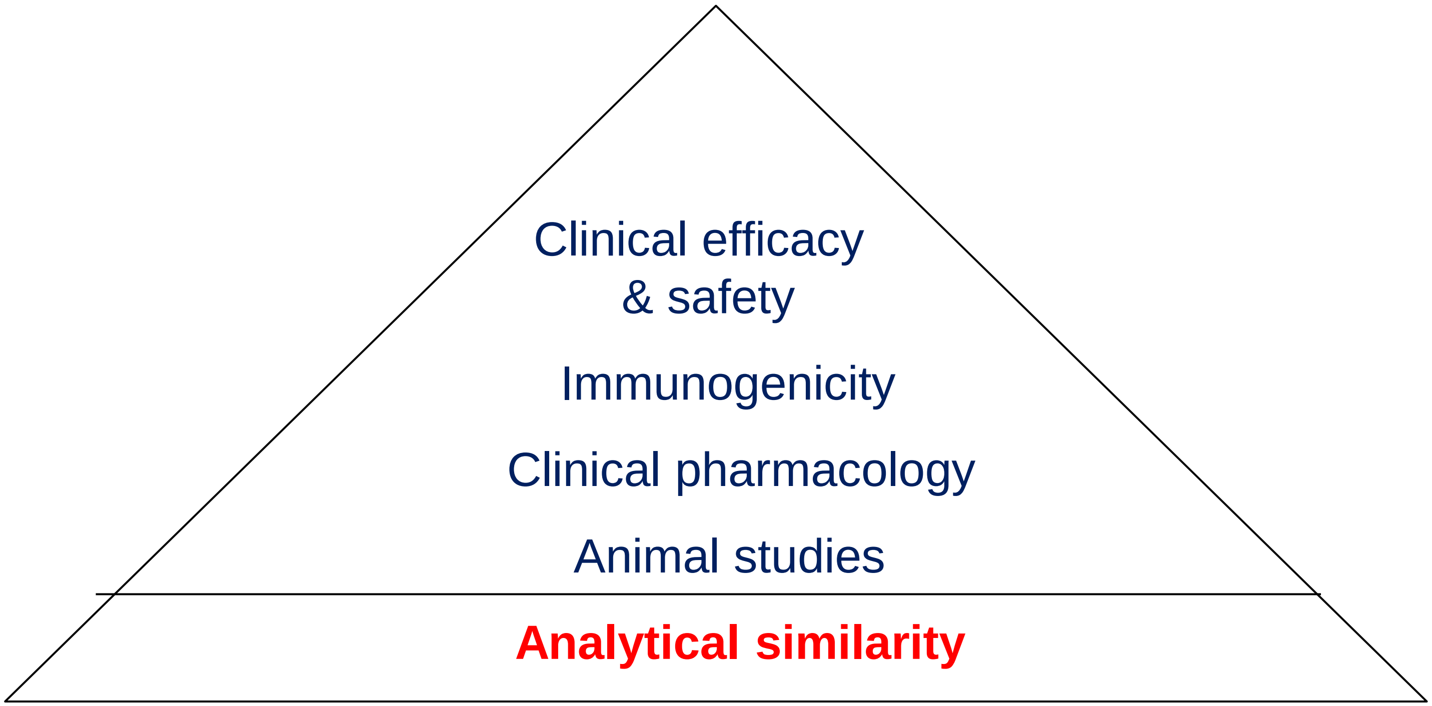
Membership

- Industry members
  - Martina Kron, Abbvie
  - Jens Lamerz, Roche
  - Mike Denham, GSK
  - Volker Schnaible, Roche
  - Christophe Agut, Sanofi
  - Timothy Mutsvari, Arlenda
  - Bruno Boulanger, Arlenda

WHAT?

Analytical Similarity

- Step-wise approach to data generation and the evaluation of residual uncertainty
- Totality-of-the-evidence to demonstrate biosimilarity



Clinical efficacy
& safety

Immunogenicity

Clinical pharmacology

Animal studies

Analytical similarity

Analytical similarity (FDA)

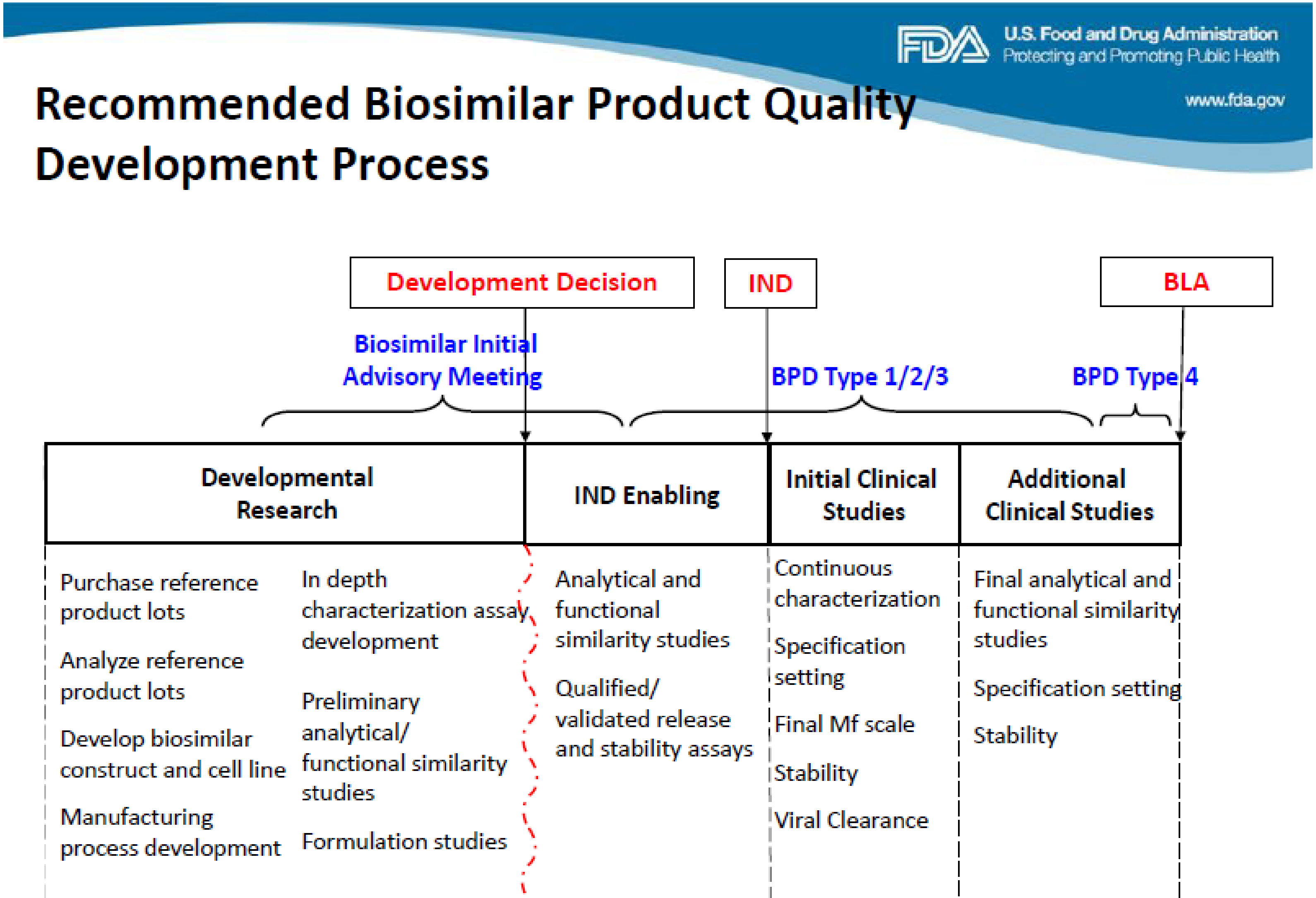
- Analytical similarity generally refers to an assessment of a proposed biosimilar product in comparison to a US-licensed reference product.
- Manufacturers should perform in-depth **chemical**, **physical**, and **bioactivity** comparisons with **side-by-side** analyses of an appropriate **number**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

→ A rather large number of Quality Attributes (> 50 CQAs)

→ Many lots of reference and test products (N lots << N CQAs)

WHEN ?

Development of a biosimilar product



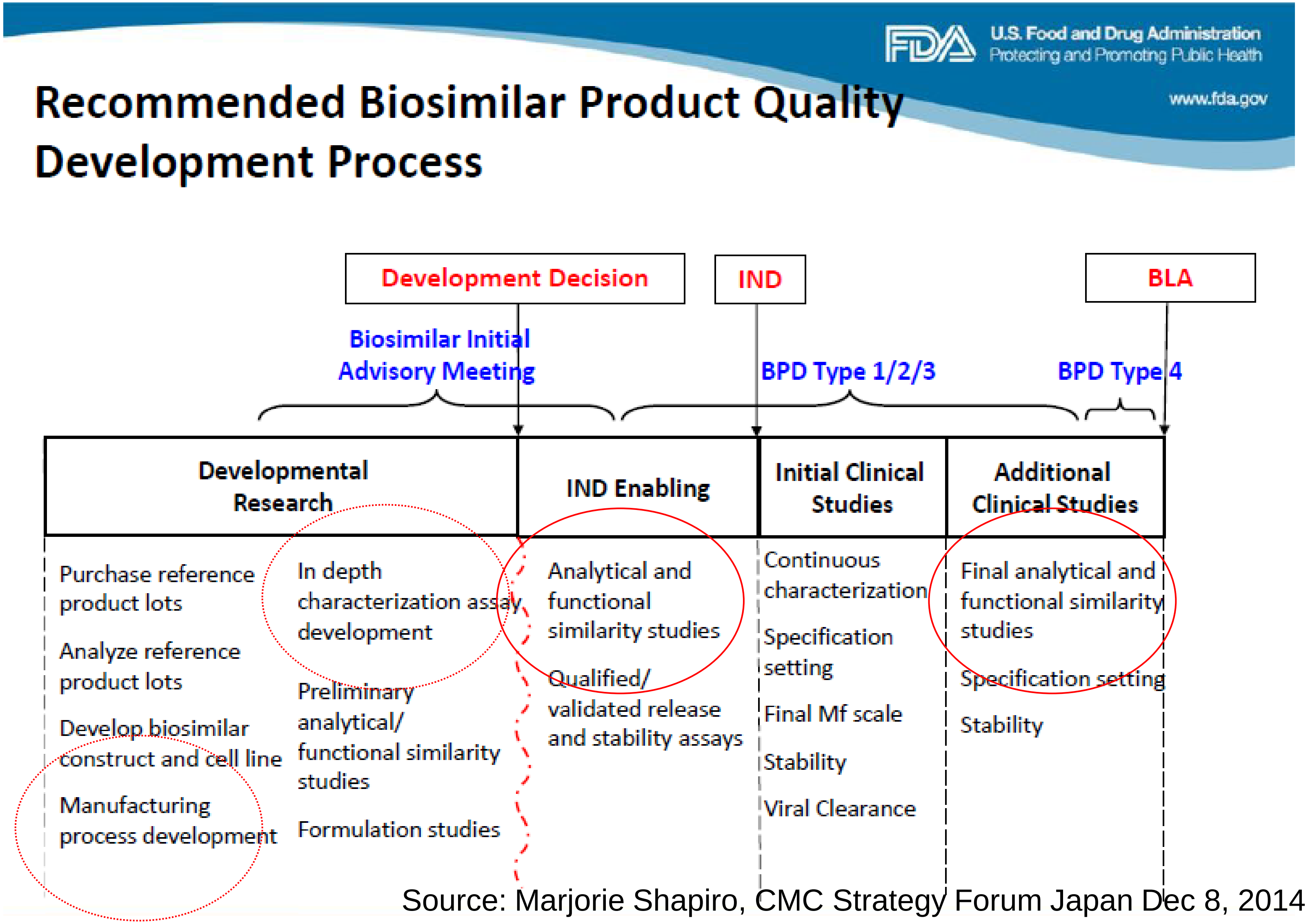
U.S. Food and Drug Administration
Protecting and Promoting Public Health
www.fda.gov

Recommended Biosimilar Product Quality Development Process

Developmental Research		IND Enabling	Initial Clinical Studies	Additional Clinical Studies
Purchase reference product lots Analyze reference product lots Develop biosimilar construct and cell line Manufacturing process development	In depth characterization assay development Preliminary analytical/ functional similarity studies Formulation studies	Analytical and functional similarity studies Qualified/ validated release and stability assays	Continuous characterization Specification setting Final Mf scale Stability Viral Clearance	Final analytical and functional similarity studies Specification setting Stability

Source: Marjorie Shapiro, CMC Strategy Forum Japan Dec 8, 2014

Development of a biosimilar product



U.S. Food and Drug Administration
Protecting and Promoting Public Health
www.fda.gov

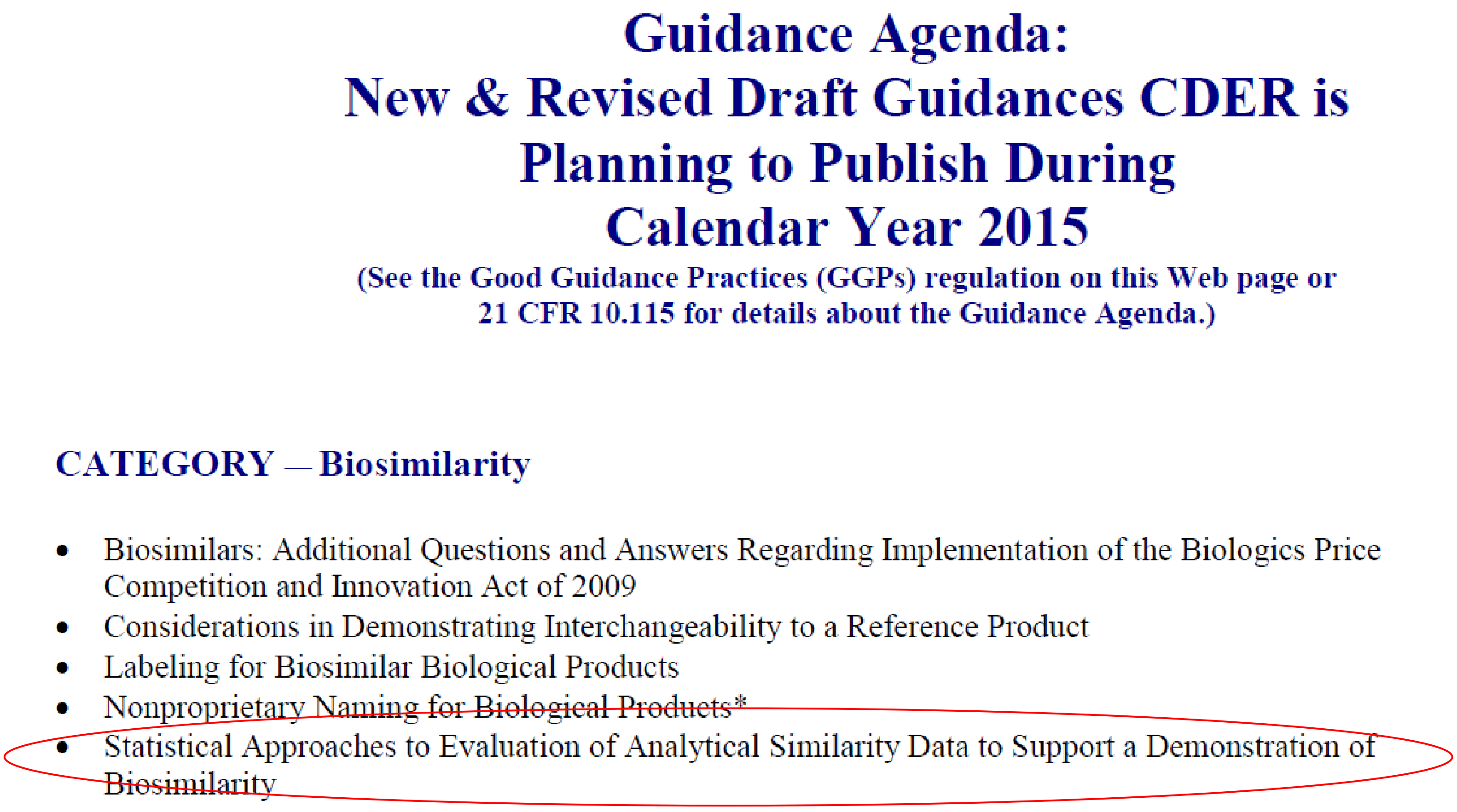
Recommended Biosimilar Product Quality Development Process

Developmental Research		IND Enabling	Initial Clinical Studies	Additional Clinical Studies
Purchase reference product lots Analyze reference product lots Develop biosimilar construct and cell line Manufacturing process development	In depth characterization assay development Preliminary analytical/ functional similarity studies Formulation studies	Analytical and functional similarity studies Qualified/ validated release and stability assays	Continuous characterization Specification setting Final Mf scale Stability Viral Clearance	Final analytical and functional similarity studies Specification setting Stability

Source: Marjorie Shapiro, CMC Strategy Forum Japan Dec 8, 2014

HOW ?

Regulatory positions today: FDA



Guidance Agenda:
New & Revised Draft Guidances CDER is Planning to Publish During Calendar Year 2015

(See the Good Guidance Practices (GGPs) regulation on this Web page or 21 CFR 10.115 for details about the Guidance Agenda.)

CATEGORY – Biosimilarity

- Biosimilars: Additional Questions and Answers Regarding Implementation of the Biologics Price Competition and Innovation Act of 2009
- Considerations in Demonstrating Interchangeability to a Reference Product
- Labeling for Biosimilar Biological Products
- Nonproprietary Naming for Biological Products*
- Statistical Approaches to Evaluation of Analytical Similarity Data to Support a Demonstration of Biosimilarity

Regulatory positions today: EMA

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

30 May 2013
EMA/CHMP/297149/2013 Rev. 1
Committee for Medicinal Products for Human Use (CHMP)

Concept paper on the need for a reflection paper on statistical methodology for the comparative assessment of quality attributes in drug development

- The EFSPI working group aims to make proposals for this reflection paper

EMA needs

"……inferential statistical approaches to compare quality attributes:

- ○ *of a (candidate) biosimilar product to that of a reference medicinal product;*
- ○ *of a particular biological drug compound in versions pre- and post-manufacturing changes."*

Analytical similarity ←→ Comparability

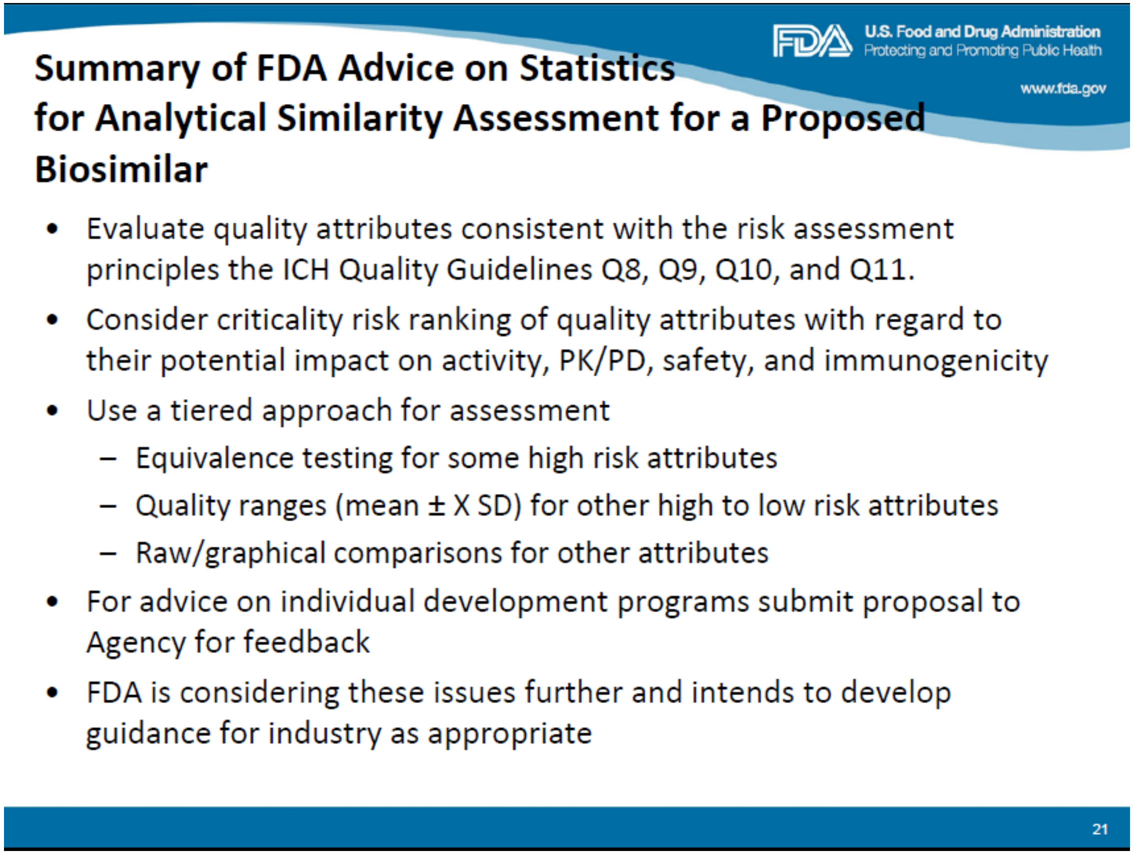
Same question:

ensure two different manufacturing processes will produce a « similar » drug products

- = similar activity, chemical and physical properties

FDA Current thinking for statistics

- Tiered approach by risk or criticality of Quality Attributes



U.S. Food and Drug Administration
Protecting and Promoting Public Health

www.fda.gov

Summary of FDA Advice on Statistics for Analytical Similarity Assessment for a Proposed Biosimilar

- Evaluate quality attributes consistent with the risk assessment principles the ICH Quality Guidelines Q8, Q9, Q10, and Q11.
- Consider criticality risk ranking of quality attributes with regard to their potential impact on activity, PK/PD, safety, and immunogenicity
- Use a tiered approach for assessment
  - Equivalence testing for some high risk attributes
  - Quality ranges (mean ± X SD) for other high to low risk attributes
  - Raw/graphical comparisons for other attributes
- For advice on individual development programs submit proposal to Agency for feedback
- FDA is considering these issues further and intends to develop guidance for industry as appropriate

21

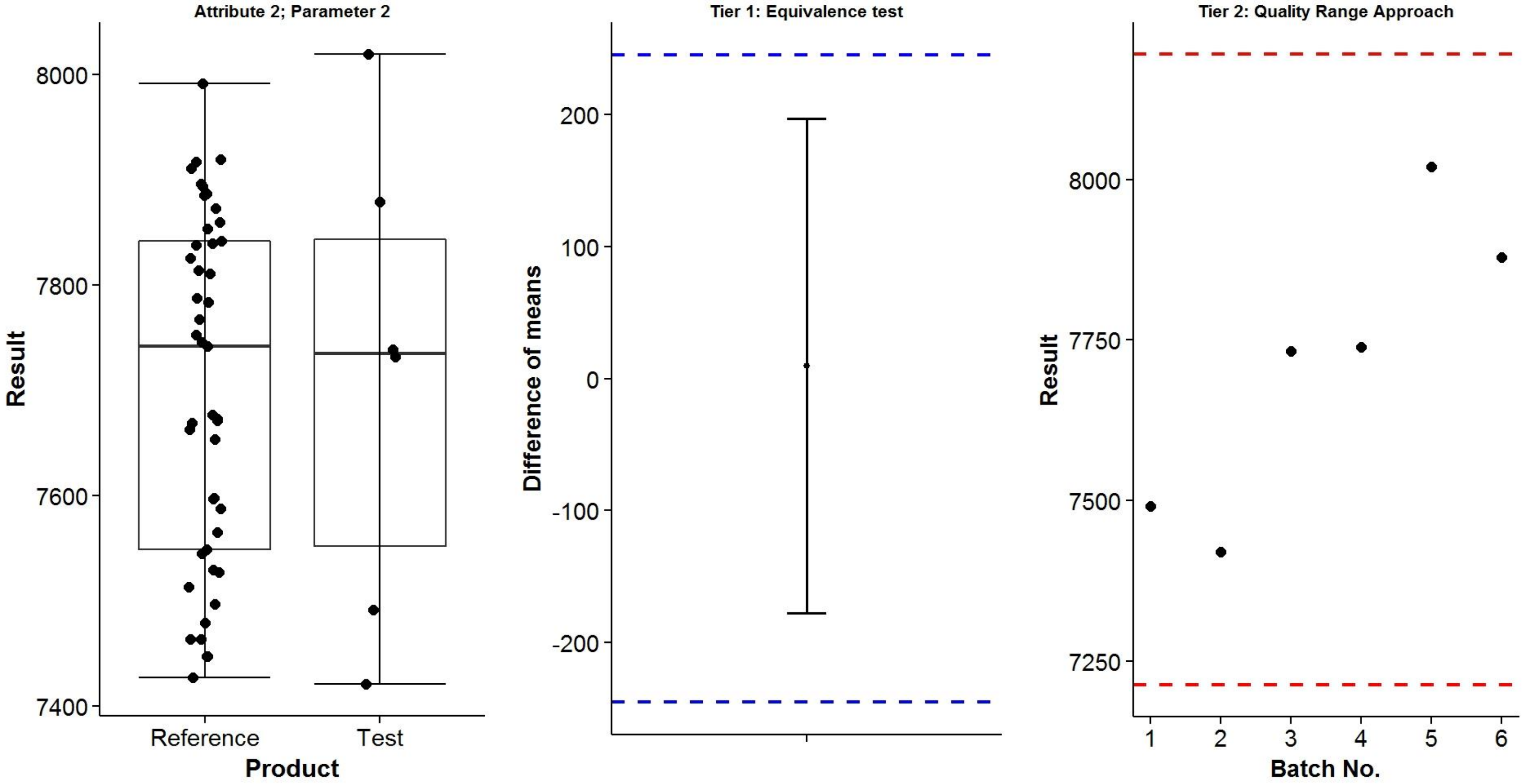
Source: Marjorie Shapiro, CMC Strategy Forum Japan Dec 8, 2014

Today’s FDA tiered approach

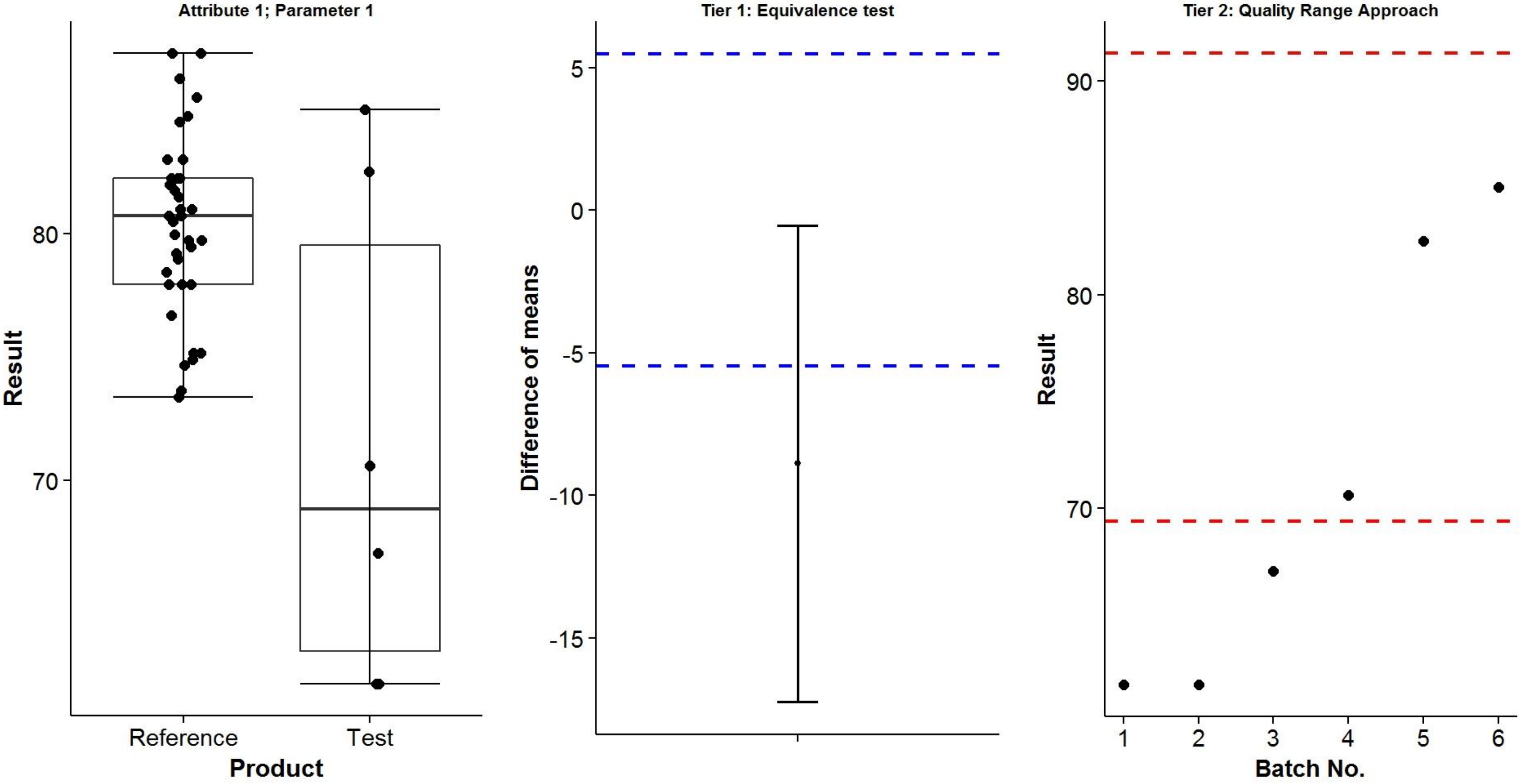
Use a tiered approach for assessment

- Tier 1 : Equivalence testing for some high risk attributes
- Tier 2: Quality ranges (mean ± c SD) for other high to low risk attributes
- Tier 3: Raw/graphical comparisons for other attributes

An example Tier 1& 2 (Pass)



An example Tier 1& 2 (Fail)



Sounds like pears and apples…..

- « Average» equivalence for high risk CQAs
  - $\mu_{Ref} \leftrightarrow \mu_{test}$
- « Individual» equivalence for medium risk CQAs
  - $Y_{i\ Ref} \leftrightarrow Y_{i\ Test}$

What is the question ?

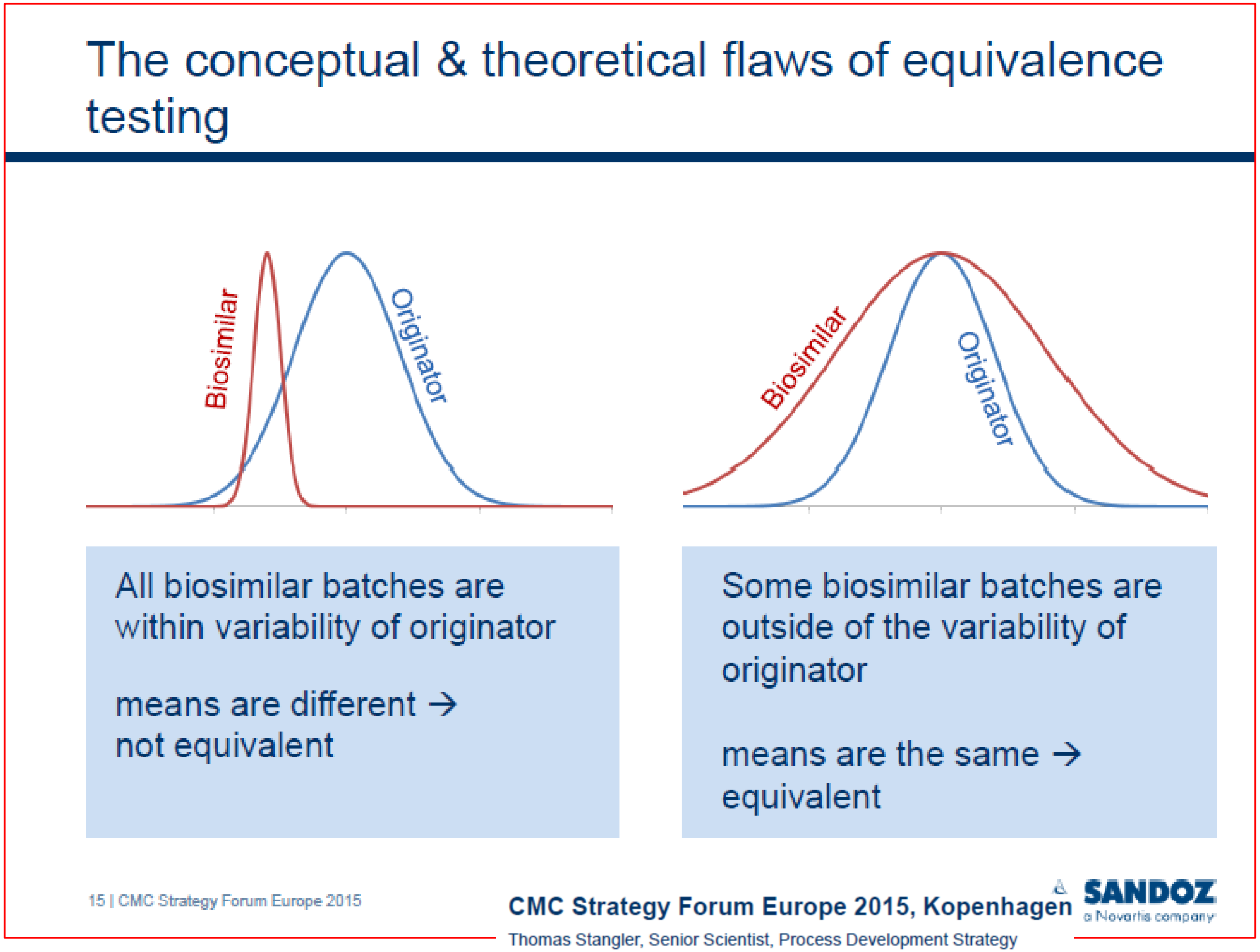
Objective of analytical similarity ? (1/3)

1 - Demonstrate that several lots of Test products are on **average** “**equivalent**” to several lots of Reference products.

- Justification of equivalence limits?
- Difficult to define limits on average based on clinical results
- Number of lots is not large….
- Multiplicity is challenging …. #CQAs >> # Lots
- Usually > 50 CQAs are considered
- Between lots and Within lot variances are important and not properly taken into account in this approach
- The “Comparability” is not achieved

Is it really the question ?

Average equivalence ?



The conceptual & theoretical flaws of equivalence testing

All biosimilar batches are within variability of originator

means are different → not equivalent

Some biosimilar batches are outside of the variability of originator

means are the same → equivalent

15 | CMC Strategy Forum Europe 2015

CMC Strategy Forum Europe 2015, Kopenhagen
Thomas Stangler, Senior Scientist, Process Development Strategy
Sandoz GmbH, Austria

Average equivalence testing

- How to define equivalence acceptable limits (EAC) on the difference of the **means** μR and μT?
- **Conceptual flaw when concluding about several batches**

- The conclusion is not about patients as in bioequivalence study
- It’s about lots of products in analytical similarity (not patients!)

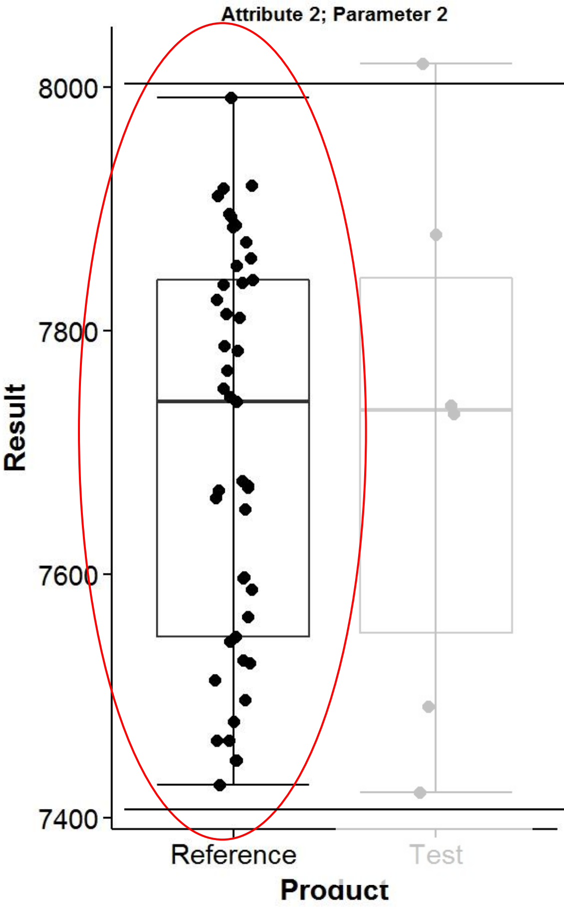
- Variance components are ignored
  - Assumes same Variability of processes R & T
  - Within lots (one unit per lot)
  - Assay Precision and format

Objective of analytical similarity ? (2/3)

2- Demonstrate that several lots of Test products are analytically "similar" or "comparable" to several lots of Reference products.

- Close to "quality range" for high to low risk QAs
- Close to a "individual equivalence" approach applied to lots, not to patients.
- Selection and number of lots is critical
- Justification of "equivalence" limits easier since linked to clinical effect
  - patients received individual lots, even units within lots
  - Several lots have been used in clinical studies
- Between lots and Within lots variance are important
- Conclusion only applies to past produced lots

Justification of acceptance limits



- All these Ref lots are released
- Clinical efficacy is recognized
- FDA propose 3 SDs
- It assumes σ ref is known
- Poor control of risk / confidence
- Accept is 90% observation in +-3SD…

Objective of analytical similarity ? (3/3)

3- Demonstrate that **proposed process will produce** lots of Test products that are analytically "comparable" to several lots of Reference products.

- Close to a "individual equivalence" approach applied to lots.
- Justification of equivalence limits easier since linked to clinical effect
  - patients received individual lots, even units within lots
  - Several lots have been used in clinical studies
- Between lots and Within lots variance are important
- This is the very question
- This is consistent with ICH Q8-Q9 concepts of risks

The future biosimilar product is the current process and its capability

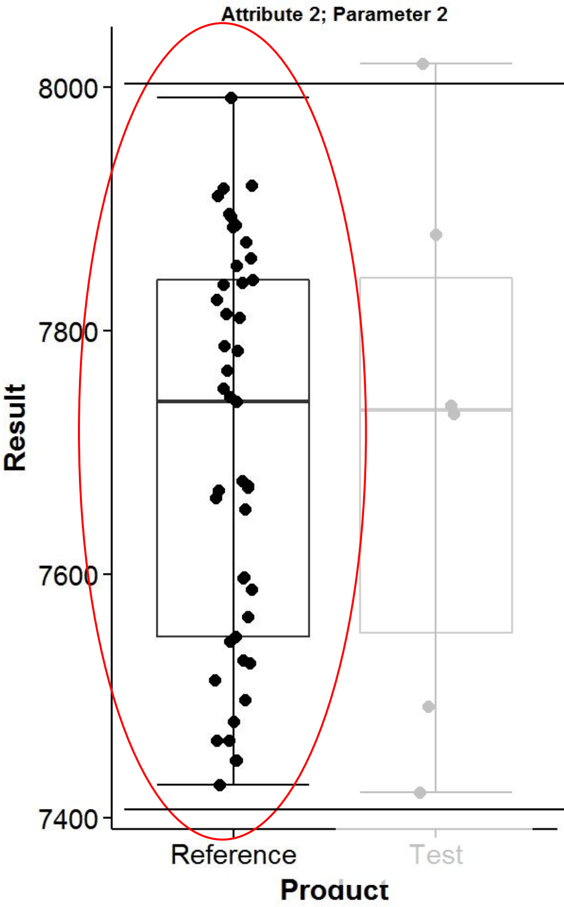
Note on Equivalence testing and Biosimilars

- Average bioequivalence (ABE)
  - Averaged over a number of patients
  - One T against one R, applicable to small molecules
  - Lot-to-lot variability was assumed –on purpose- to be under control
- Interchangeability (Population and Individual BE)
  - It’s about prescribability and Switchability
  - ➔ Produce the same clinical effect whatever the patient
- Analytical similarity for biosimilars
  - To ensure the product is the same whatever the lot of Test product

Statistical challenges/opportunities

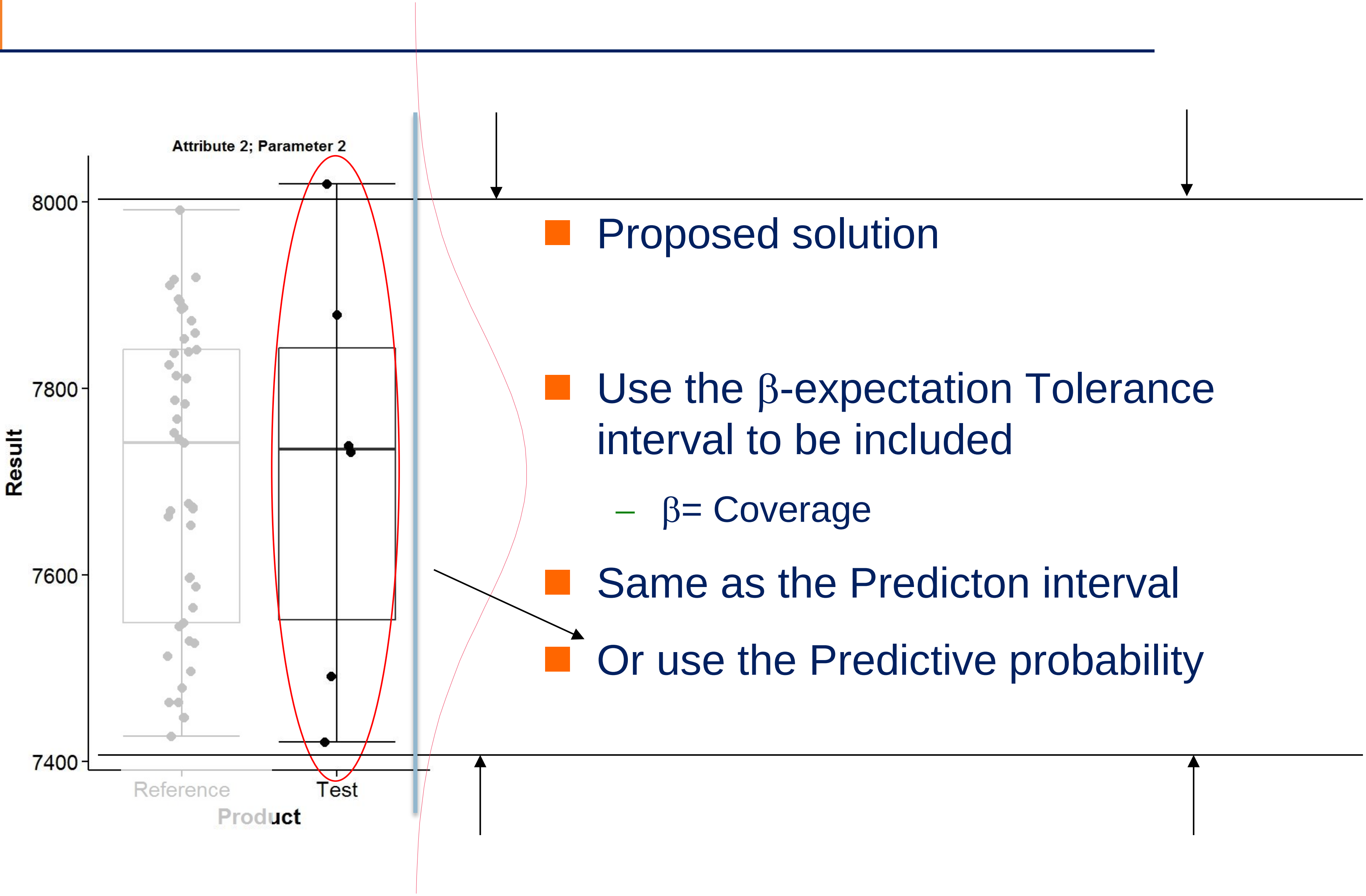
- Justification of “equivalence limits” connected to the clinical results
- Poor precision and large uncertainty of **bio**assays
- Variability of biological processes (between lots)
- Selection of lots and number of lots (R & T)
- Many correlated CQAs that should be **jointly** proven as “similar”
- Content uniformity (within lots)
- Advanced signal processing

Justification of acceptance limits



- Proposed solution
- Use the β-γ-Content Tolerance to define the acceptance limits
  - β= Coverage, say 90%
  - γ= Confidence, say 95%
- E.g. 90-95 Tolerance Interval

Decision



- Proposed solution
- Use the β-expectation Tolerance interval to be included
  - β= Coverage
- Same as the Predicton interval
- Or use the Predictive probability

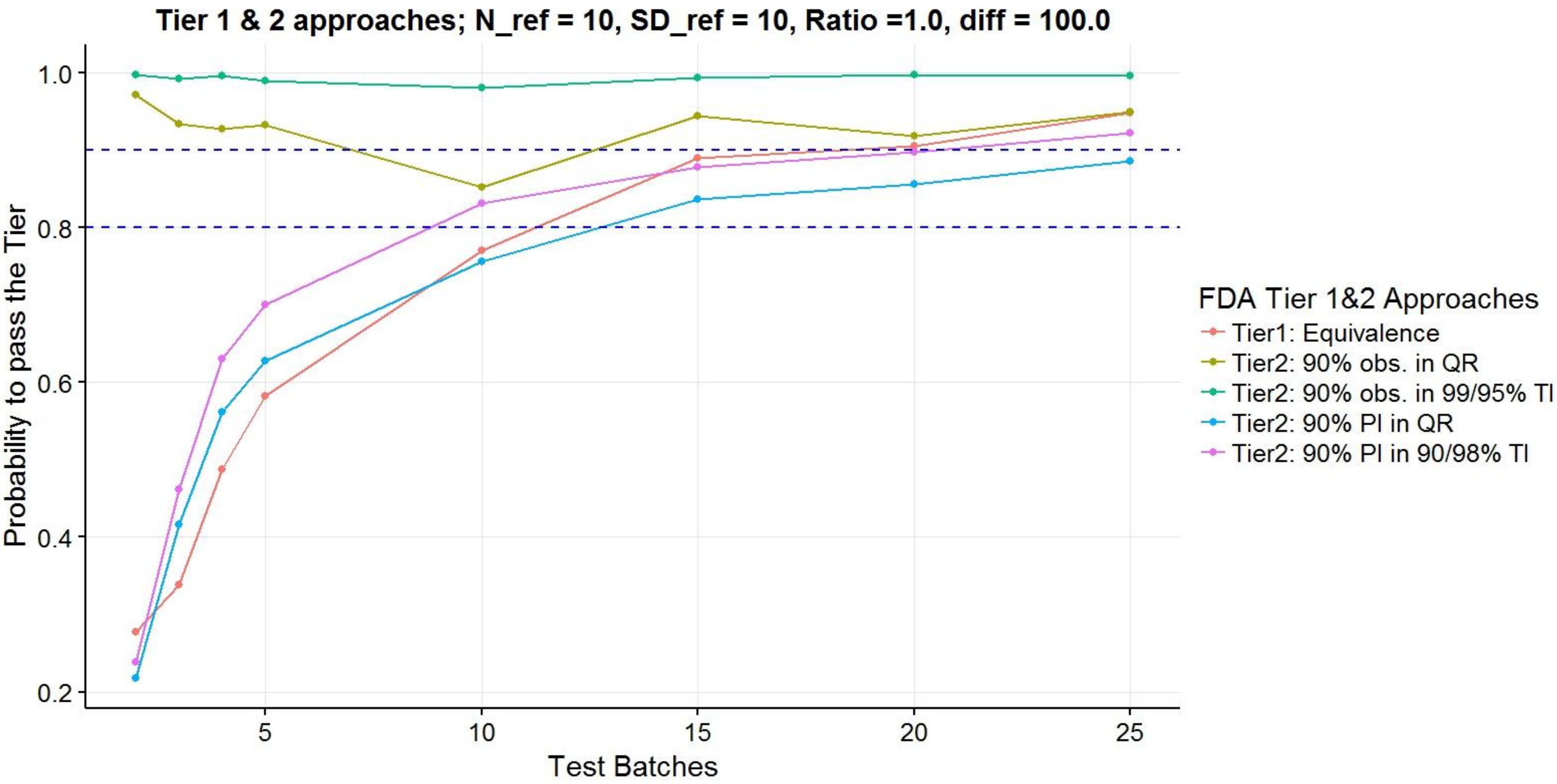
Comparison by simulations

- Assume
  - Test = Reference mean=100, SD=10
  - # Reference lots is 10

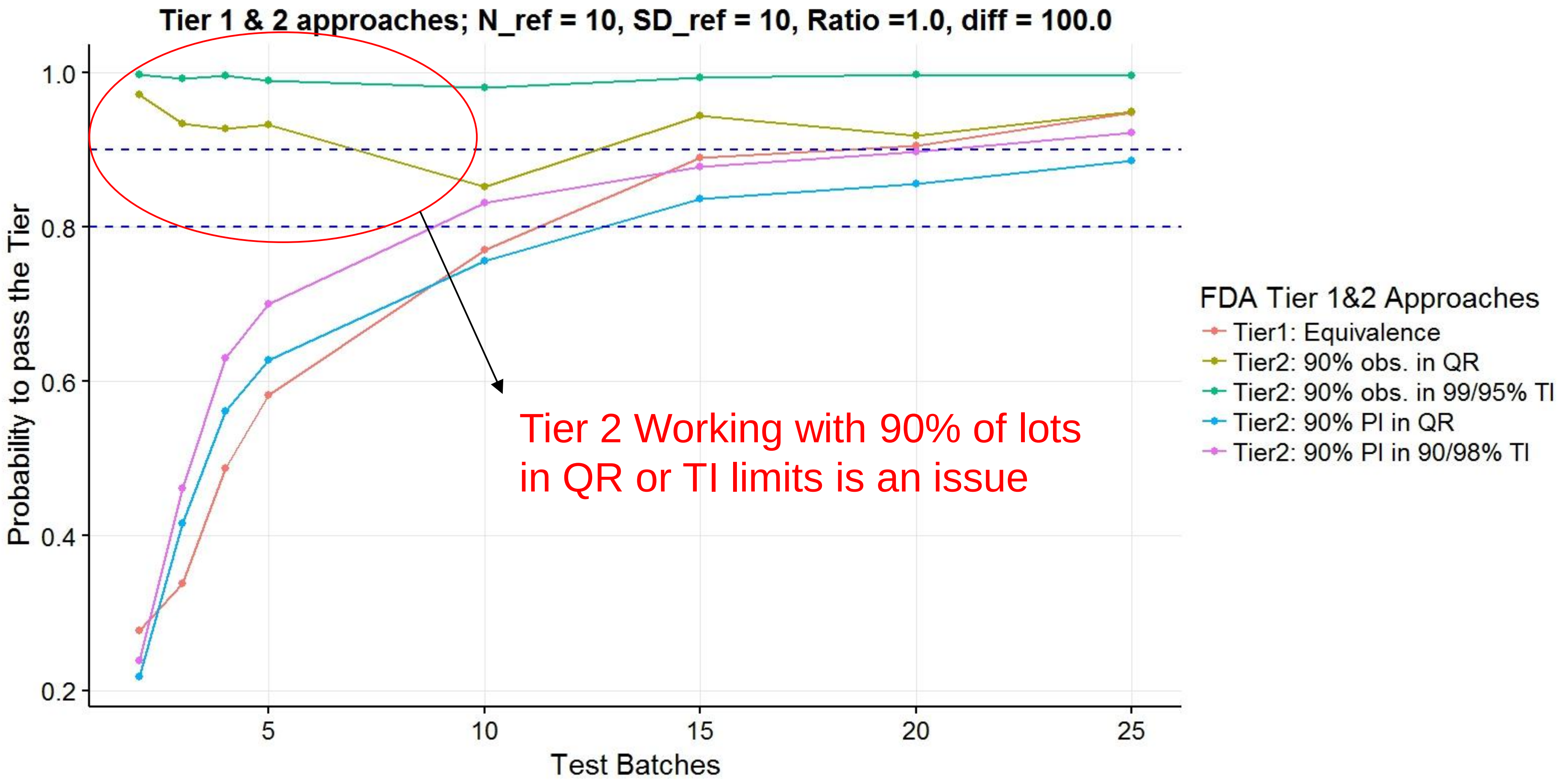
Decision methods

- Tier 1 FDA average Equivalence
- Tier 2 FDA 90% lots in +- 3 SD
- Tier 2 90% lots in 90/95 Tolerance Interval
- Tier 2 90 Prediction interval in +- 3 SD
- Tier 2 90 Prediction Interval in 90/98 Tolerance Interval

Comparison by simulations

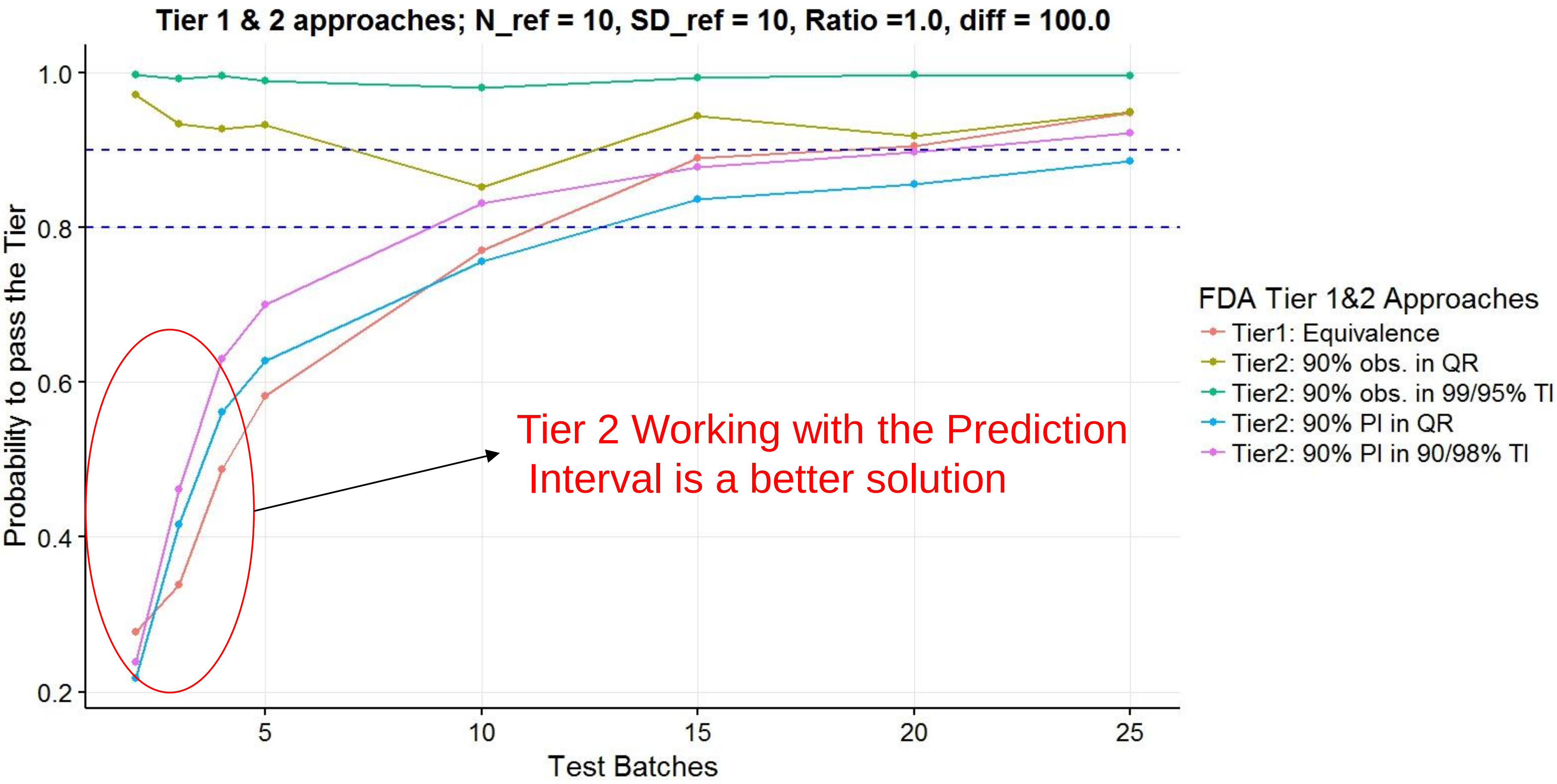


Comparison by simulations



Tier 2 Working with 90% of lots in QR or TI limits is an issue

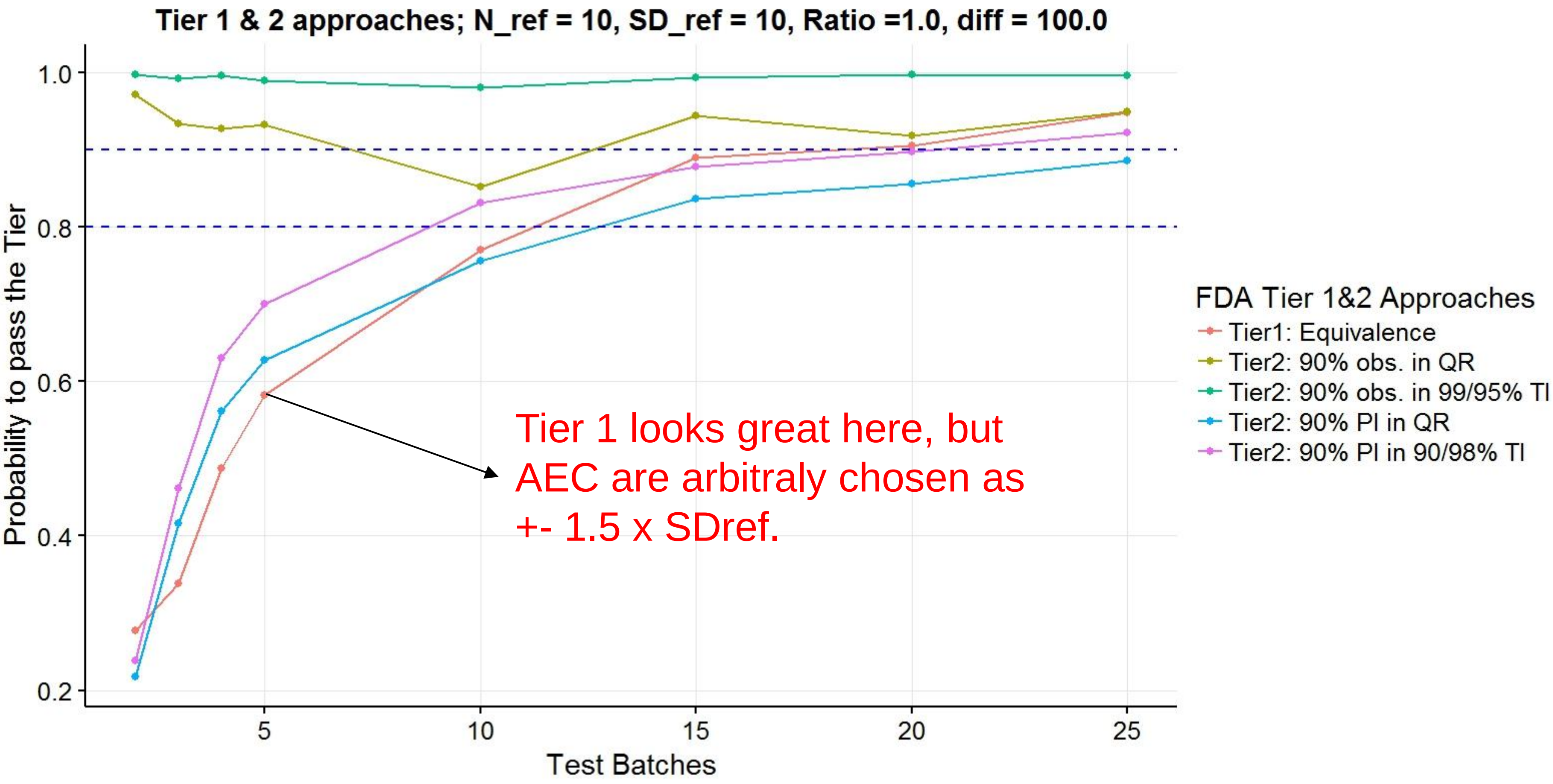
Comparison by simulations



Tier 1 & 2 approaches; N_ref = 10, SD_ref = 10, Ratio =1.0, diff = 100.0

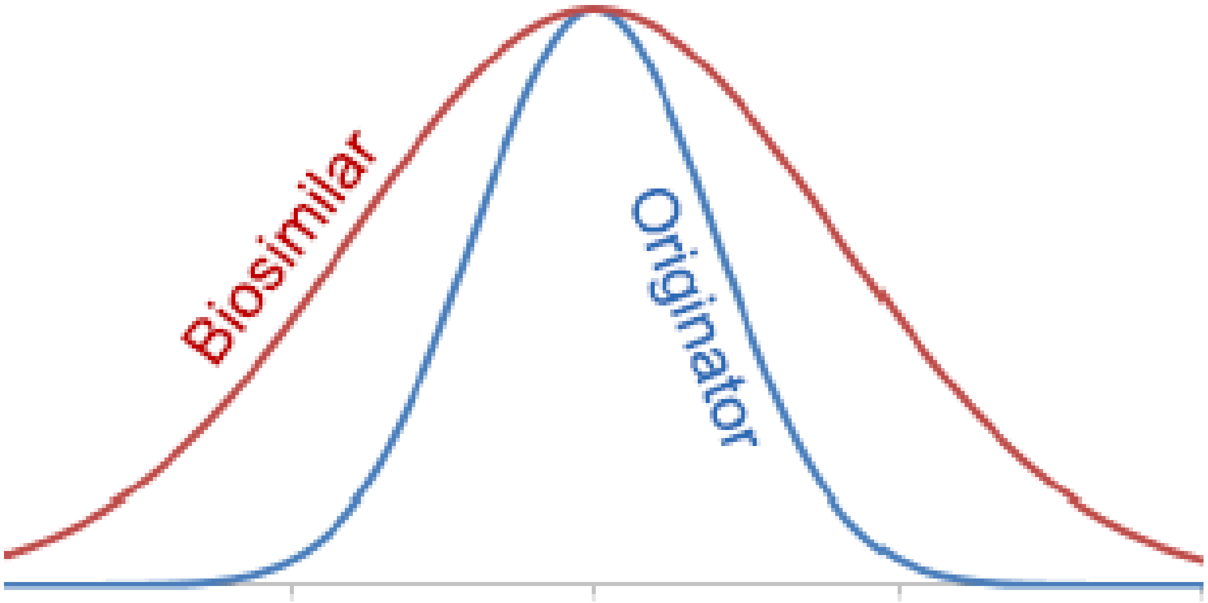
Tier 2 Working with the Prediction Interval is a better solution

Comparison by simulations



Tier 1 looks great here, but AEC are arbitraly chosen as +- 1.5 x SDref.

What if Var Test > Var Reference ?

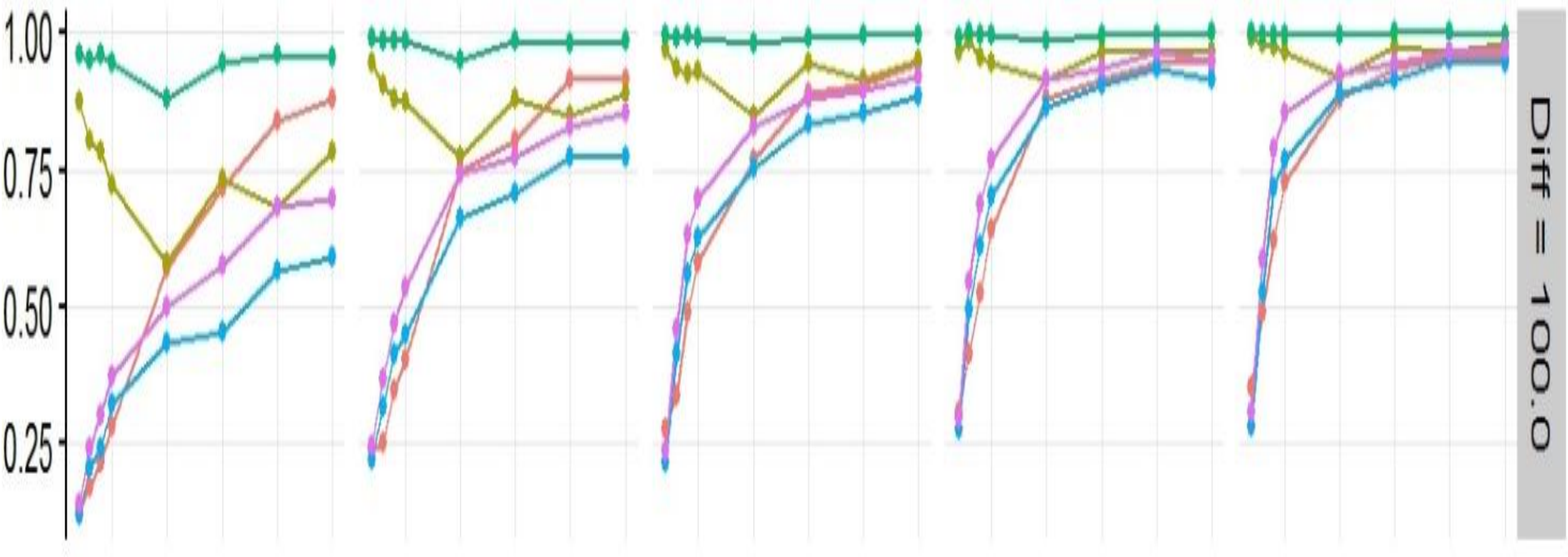


Some biosimilar batches are outside of the variability of originator

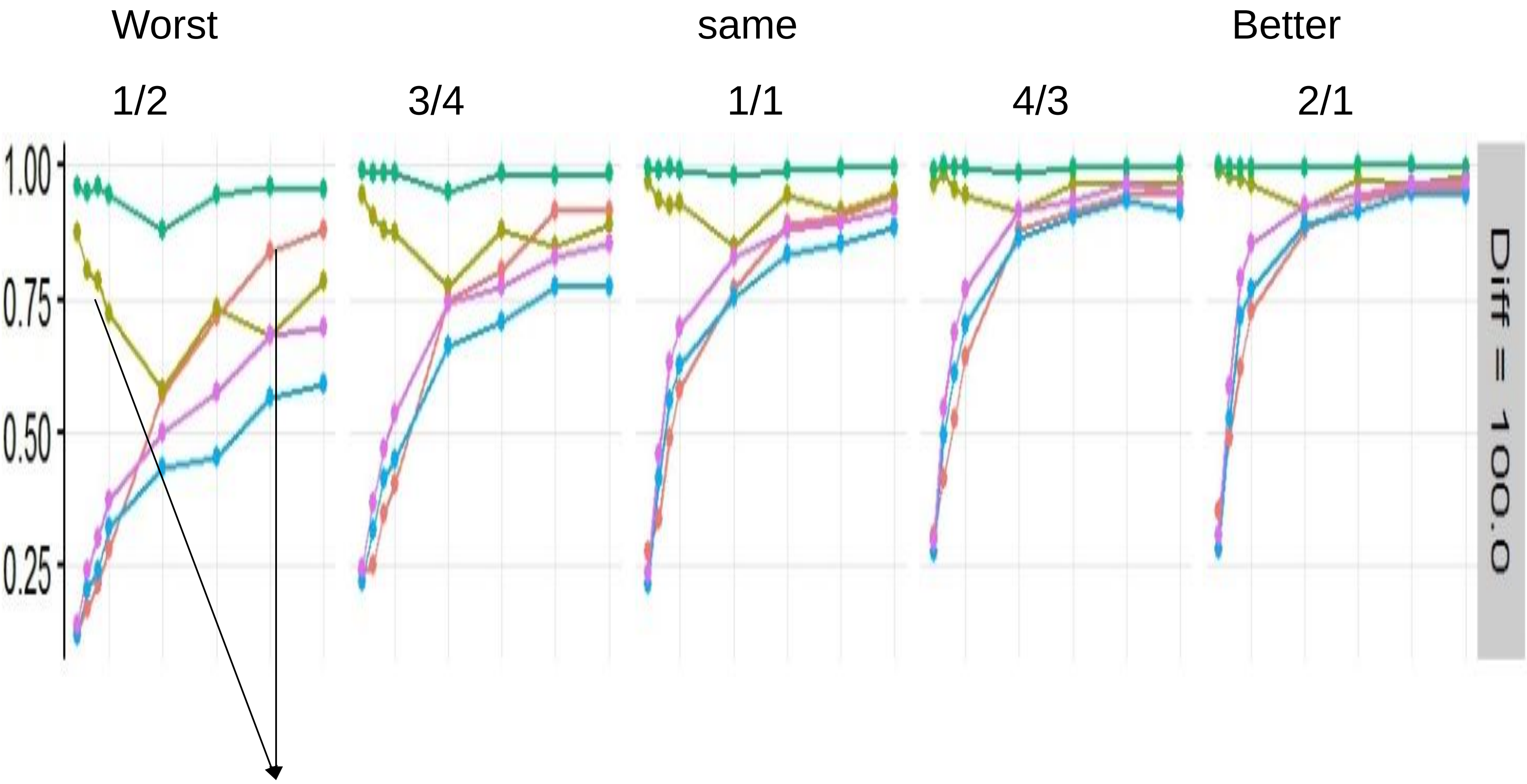
What if Var Test > Var Reference ?

Worst same Better

1/2 3/4 1/1 4/3 2/1



What if Var Test > Var Reference ?

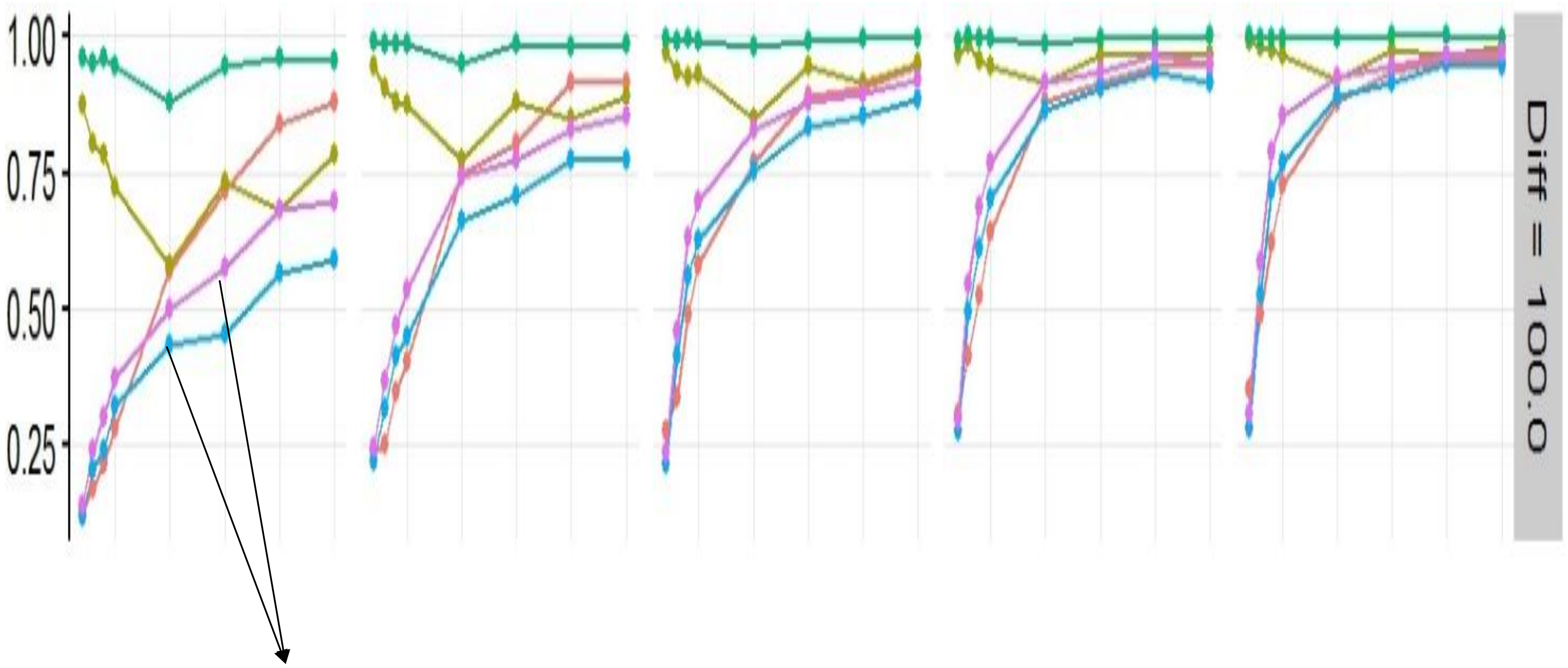


FDA Tier 1 and Tier 2 approaches are not sensitive to difference in process variability

What if Var(ref) / Var (std) is different

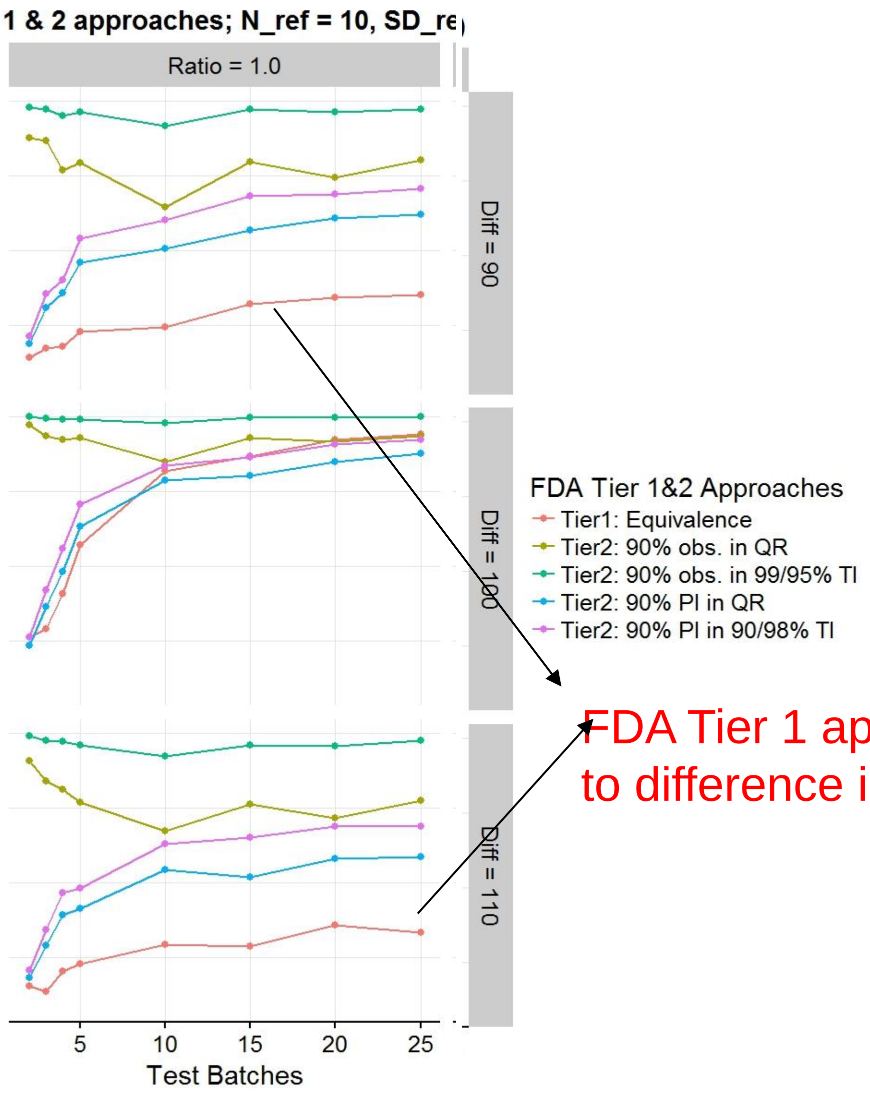
Worst same Better

1/2 3/4 1/1 4/3 2/1



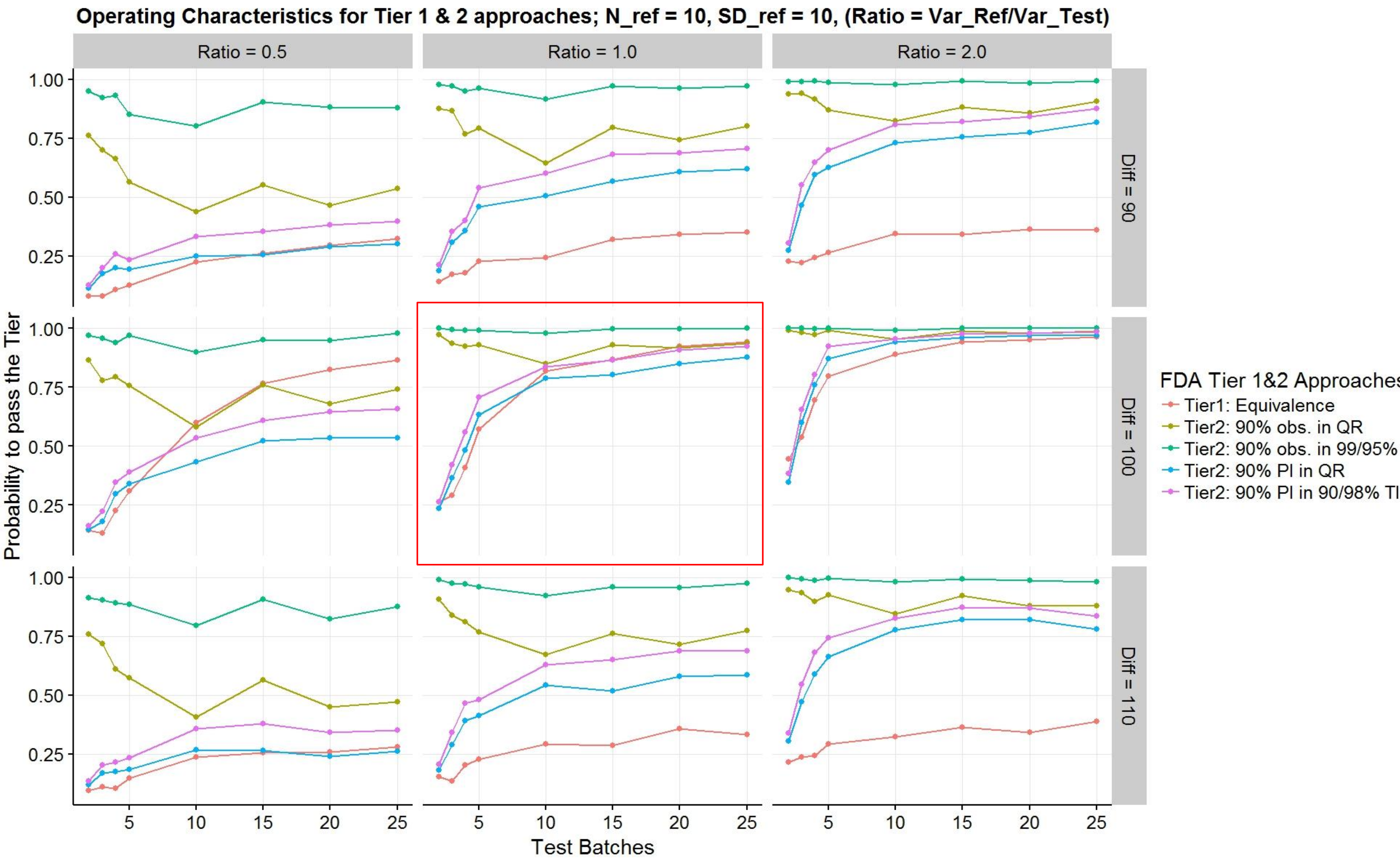
Prediction Interval based methods
are adequately sensitive

What if mean Test <> Var Reference ?

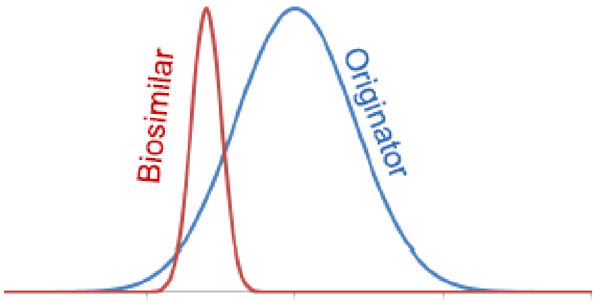


FDA Tier 1 approach is sensitiv
to difference in process mean

The big picture

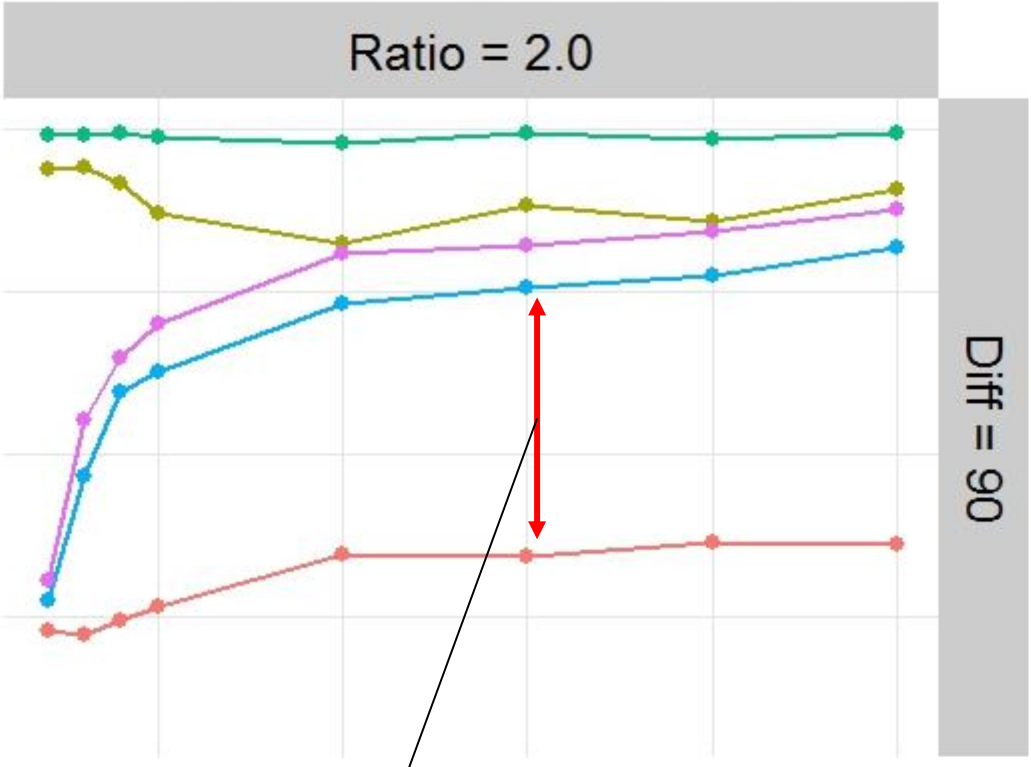


Zooming on a special case.



All biosimilar batches are within variability of originator

means are different → not equivalent



FDA Average equivalence would reject this case whilst Prediction based approach is appropriate.

FDA Tier 1

- Only sensitive to important difference in means
- Whilst equivalence test take into account different variances, its poorly sensitive to differences of capability of processes

- Equivalence limits remain arbitrary at this stage

FDA Tier 2

- Using the Prediction interval on individual lots is recommended
- No way to succeed with limited number of lots
- Closer to the question: where will future lots be

- Using the β-γ-Content Tolerance interval is the recognized way to define limits based on past observation and recognized clinical efficacy.
- γ can be tuned to optimize operating characteristics

Analytical similarity objective

- Ensure the population of patients are likely to receive a similar product, having the same clinical effect, whatever the lot
- Whatever the future lots made with a new process
- Given variability between lots and within lot
- Otherwise why requiring that analytical similarity studies should include several lots
  - It is the ability of the new proposed process to produce the same material that is targeted in a way
  - This new process could be of better quality

- It's closer to a "comparability" and "capability" assessment

Alternative proposal

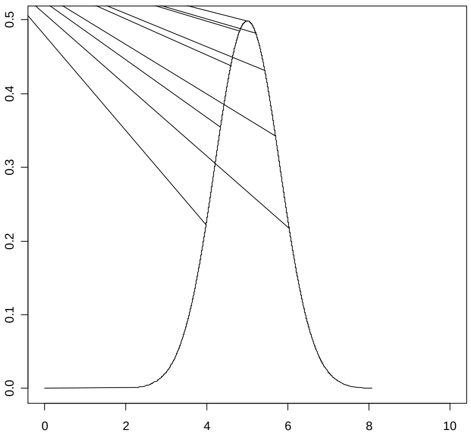
- Test if β-Prediction Interval is within β-γ-Tolerance Interval
- More relevant than using an arbitrary c factor (such as 3!)
- Take into account the variability of Test process (between-lots)
- Prove that all Test lots will be within the range of Reference lots with some level of confidence

The Bayesian route

- What is the predictive probability to be in specification ?
- How many test lots should be made to demonstrate it ?
- How to take into account the dependencies/correlation between the many CQA in the decision ?
- How to leverage in the information I have?
  - Eg about assay variability
- How to be confident about robustness of the process ?
  - I.e. producer’s risk !

Bayesian principle

PRIOR distribution distribution

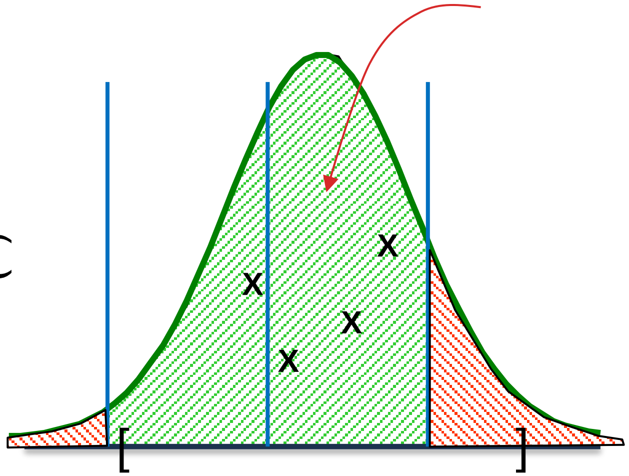


+

Analytical data

∝

POSTERIOR



P(potency in Specs)= P(quality)

- Frequentist ➔ P(data | assumed similarity)
- Bayesian ➔ P(similarity | data)
  - This is the question in fact!!

Proposal: Predictive analytical similarity

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

- Based on the Predictive Distribution of future Test lots
- The Bayesian theory provides a definition of the Predictive Distribution of a new lot given past data.

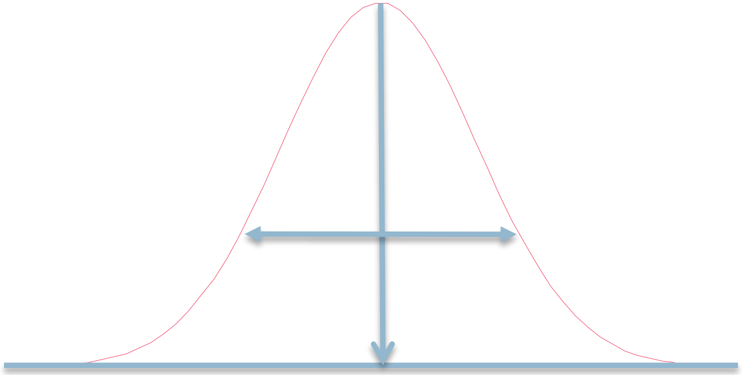
$$p(\tilde{x}|data) = \int_{\sigma^2}\int_{\mu} \underbrace{p(\tilde{x}|\mu,\sigma^2,data)}_{\text{Model}} \times \underbrace{p(\mu,\sigma^2|data)}_{\text{Joint posterior}} \; d\mu d\sigma^2$$

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

Note on Predictive distribution

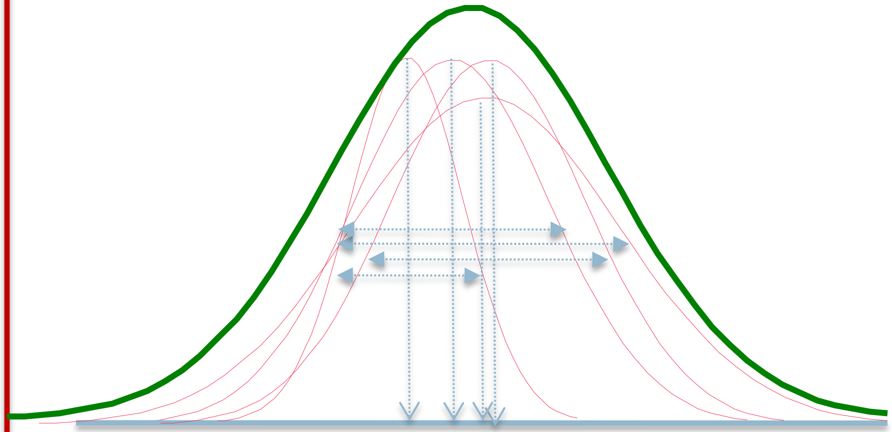
Simulations

the “new observations” are drawn from distribution “centered” on estimated location and dispersion parameters (treated as “true values”).

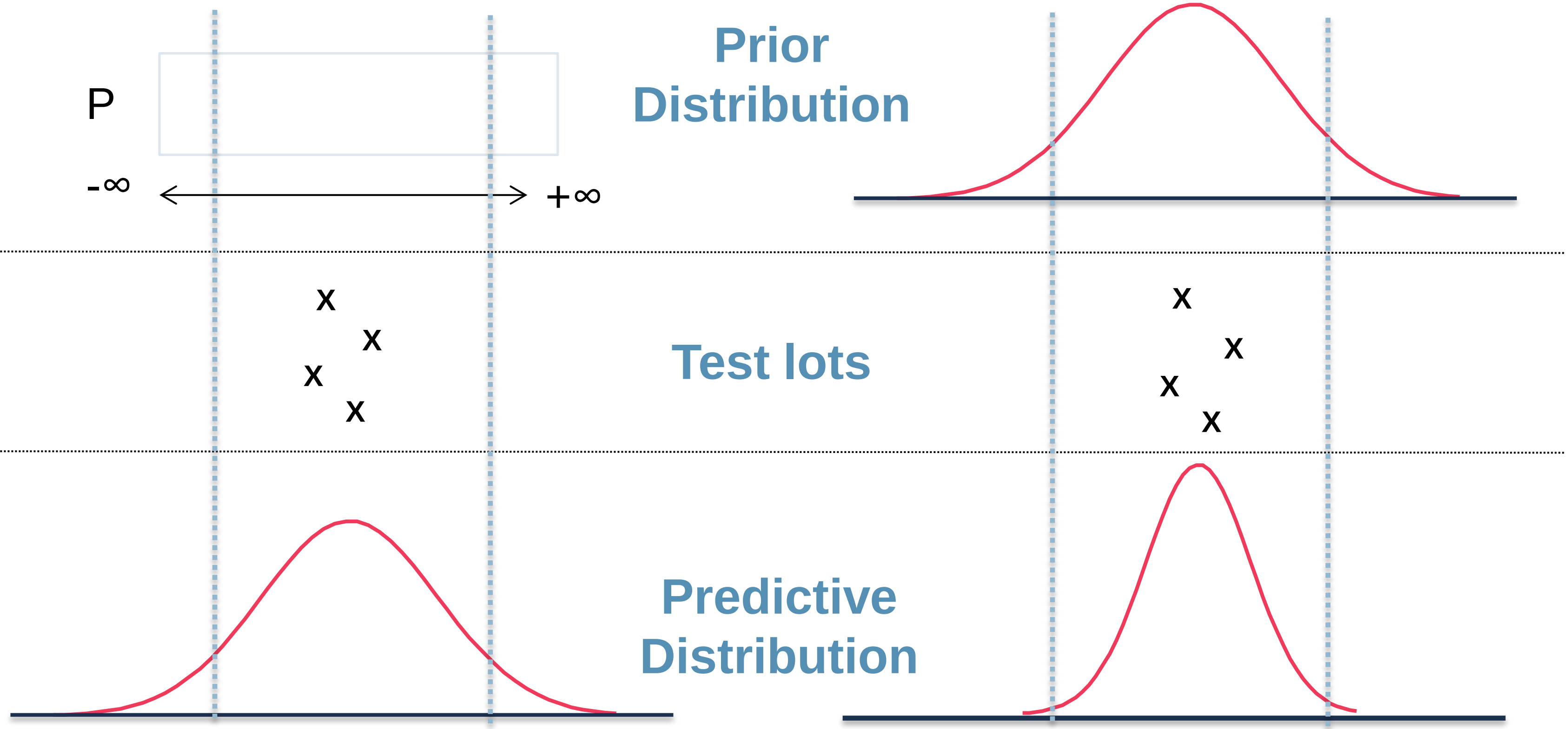


Predictions

the uncertainty of parameter estimates (location and dispersion) is taken into account before drawing “new observations” from relevant distribution



Bayesian Method – Prediction



Prior Distribution

Test lots

Predictive Distribution

Based on point estimates

Frequentist

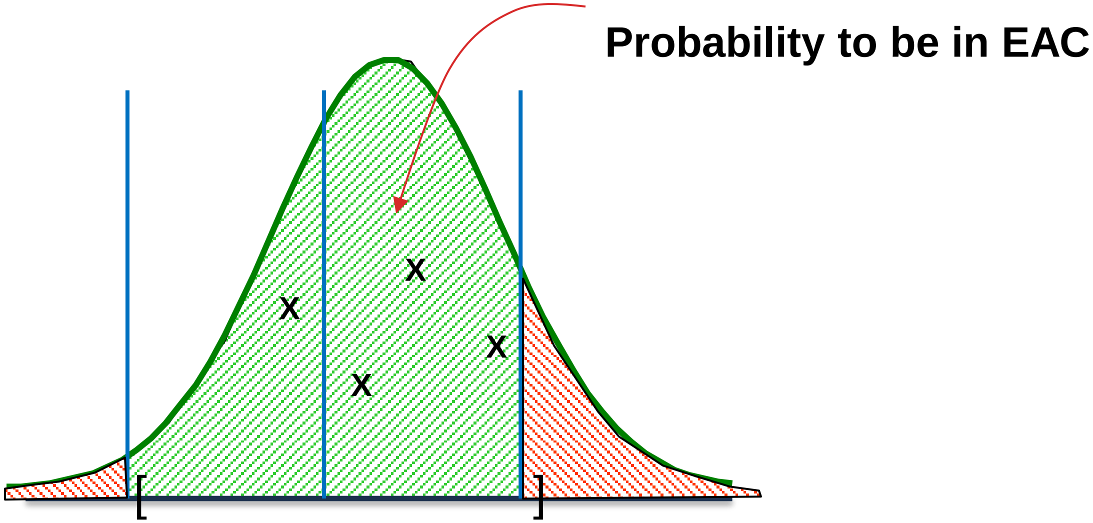
Based on distribution of parameters

Bayesian

Can compute directly Predictive Probability to be within acceptance limits

Probability being in specifications

- Use the Predictive distribution to compute the probability to in specifications.



- ➔ What’s the risk ?

Predictions and NOR: ICH Q8 & 9 Risk

- The known or assumed control/uncertainty on CPPs can be integrated into Predictions:

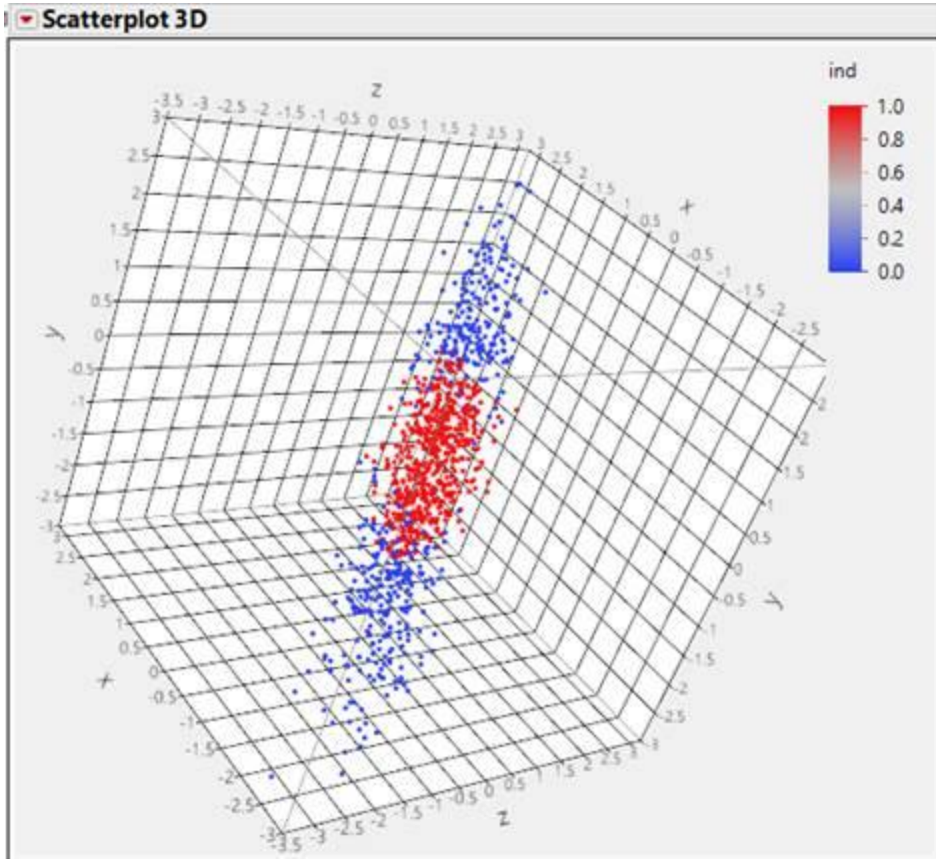
$$p(\tilde{y}|data) = \int_X \int_{\sigma^2} \int_\mu p(\tilde{y}|\mu,\sigma^2, X, data) \times p(X) \times p(\mu,\sigma^2|data)\ \mathrm{dX}\, d\mu\, d\sigma^2$$

Provide a distribution on CPP (NOR)

- This predictive distribution allows to compute the P(Lot in EAC) or Capability under realistic/industrial conditions to produce biosimilars.
- The use of a distribution on CPP depends on designs used during the Stage 1 according to QbD principles

Multi-Criteria decision method

- When there is several CQAs to analyze jointly
  - Use the joint probability of acceptance



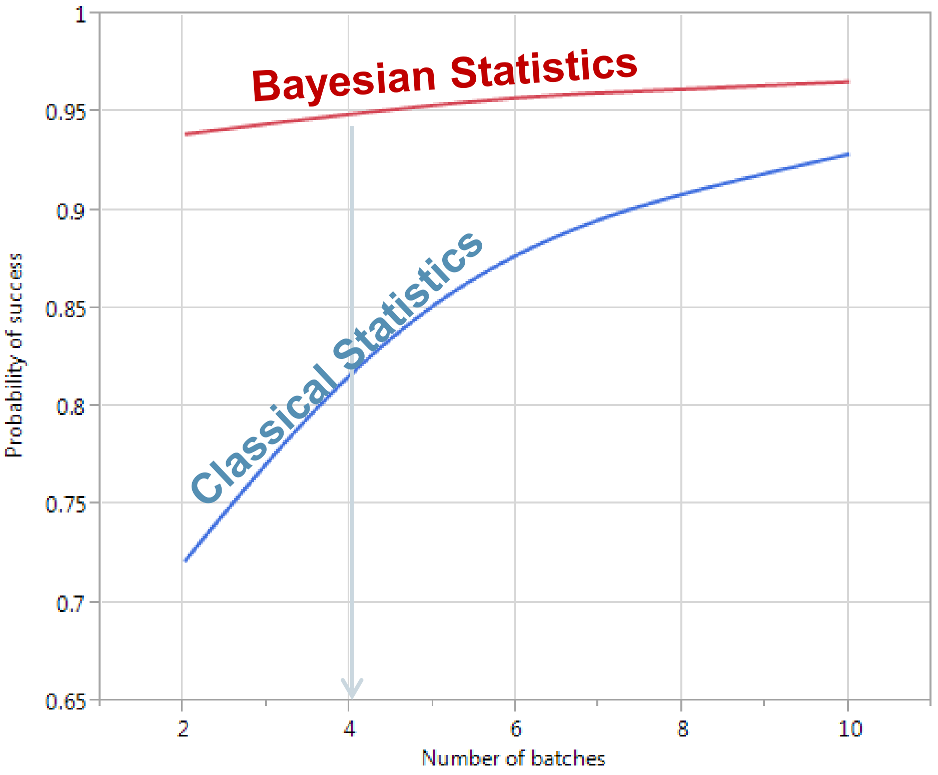
in red : in AEC
in beu : out of AEC

About Priors

- Priors could be used and justified.
- They should be established and fixed beforehand
- Predictive distribution could be envisaged with non-informative priors
- Recommended to be weakly non-informative on parameters of interest
  - Mean Test process
  - Variance test process
- Could be informative on Precision of assays
- Could be informative on dependencies between assays
- Could be weakly informative on CU

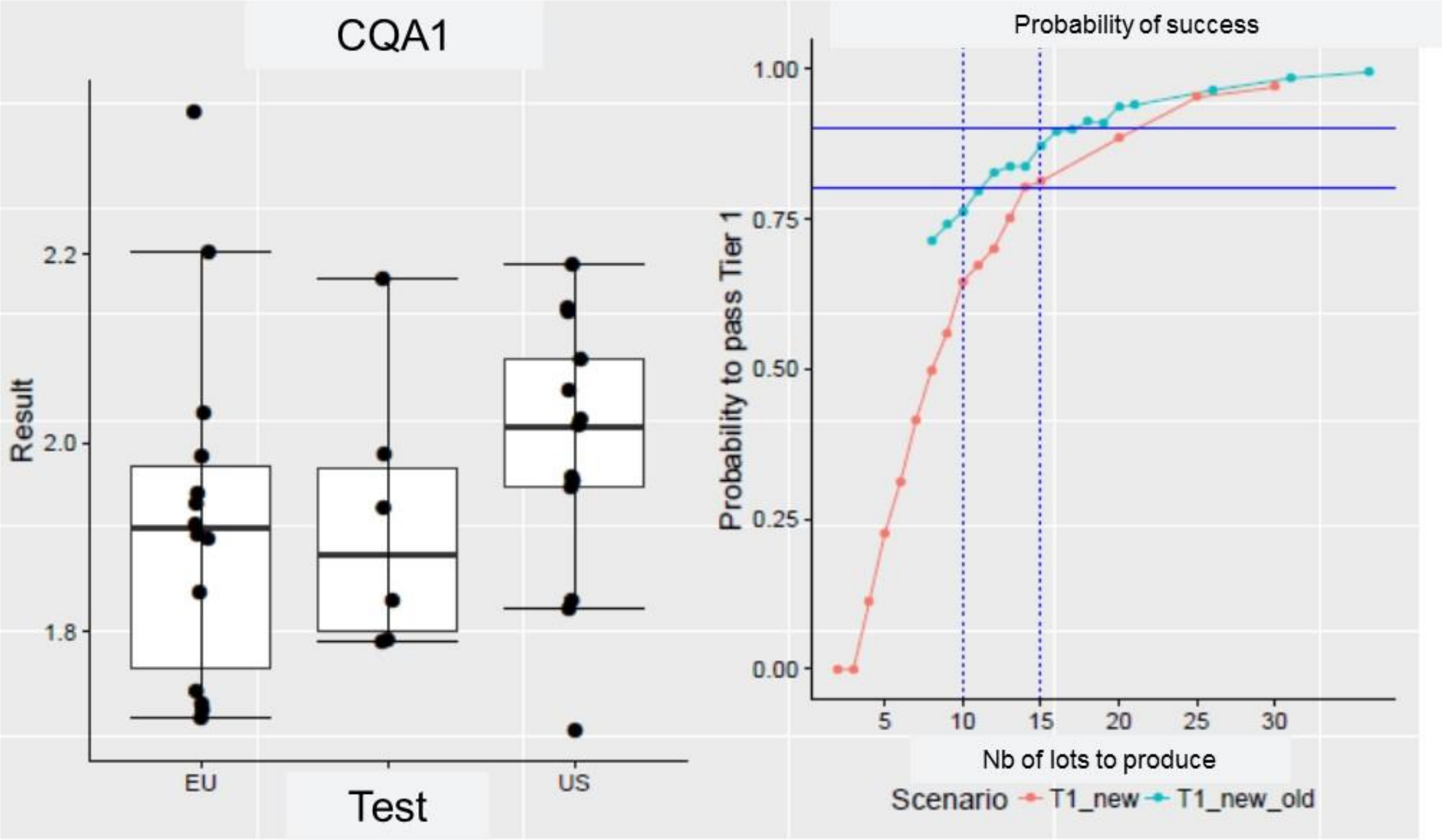
Number of Batches to be used

- Number of batches required to guarantee 95% of success of future results will be within EAC.
- Classical Stats requires more batches
- Bayesian statistics using prior (defendable/obvious) information requires less batches.
- Why? Because the Posterior of performance parameters is more precise.
- Use weak priors on parameters of interest



An example

- How many new lots given past lot results ?



Conclusions

- What's the real objective here ?
- Similarity should be proven whatever the future lots and units
- Bayesian methods using the predictive distribution answers the very objective
- Bayesian models can be used during development to justify the number of lots to perform
- Bayesian methods easily handle multiplicity: the predictive joint probability can be computed
- Informative priors on some parameters can be justified and recommended
- ➔ Ensuring future Test products will be biosimilar

Bruno.Boulanger@arlenda.com

THANK YOU