Recent Trends in the Evaluation of Analytical Biosimilarity

WCBP 2016, Washington D.C.

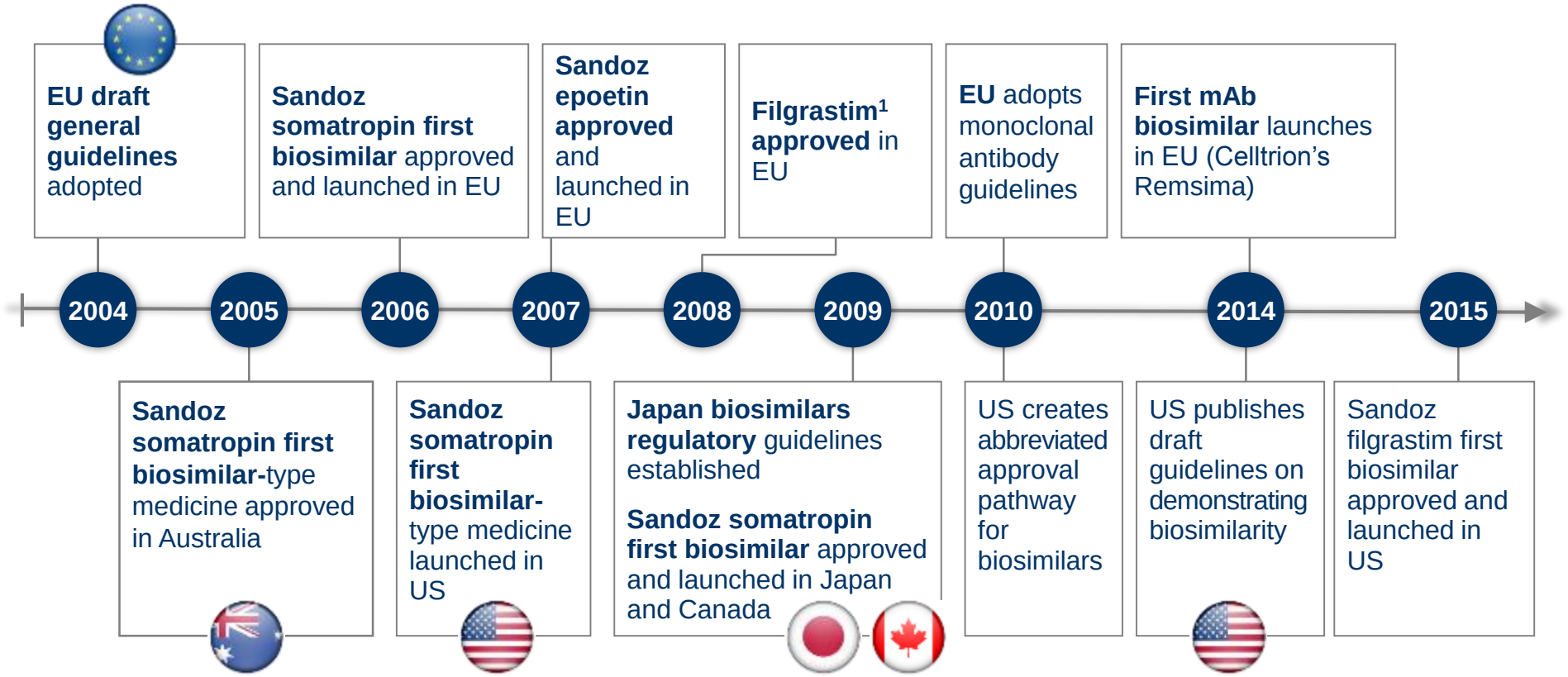
Thomas Stangler, Senior Scientist, Process Development Strategy

Sandoz Biopharmaceuticals

© 2016 Sandoz. All rights reserved. All trademarks are the property of their respective owners.

SANDOZ
a Novartis company

Biosimilars are recognized around the world as safe and effective medicines



1 First competitor product (Sandoz product approved Feb 2009)

SANDOZ
a Novartis company

10 Years of Biosimilars...

... and evolving analytical technologies, development concepts and regulatory sciences

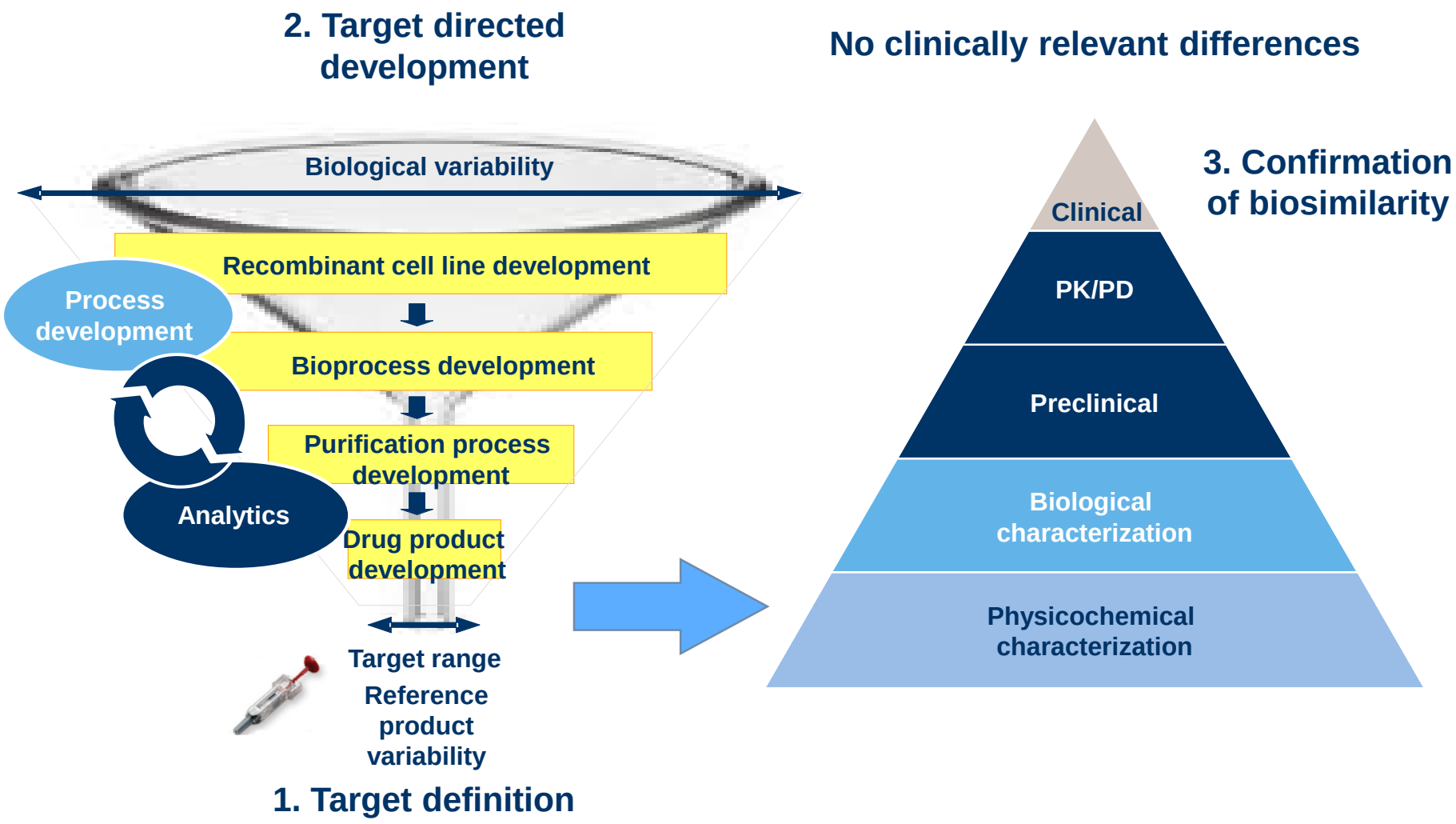
Increases from 2005 to 2015:

Number of reference product batches*:	×12
Number of analytical methods:	×4
Number of quantitative readouts:	×10
Pages of biosimilarity exercise:	×7
Molecule size (mol. weight):	×8

* as part of the comparability exercise

SANDOZ
a Novartis company

Biosimilars are systematically and iteratively developed to match the reference product



Adapted from McCamish M, et al. Worldwide experience with biosimilar development. Mabs. 2011;3(2):209–17; McCamish M, Woollett G. The state of the art in the development of biosimilars. Clin Pharmacol Ther. 2012;91(3):405-17

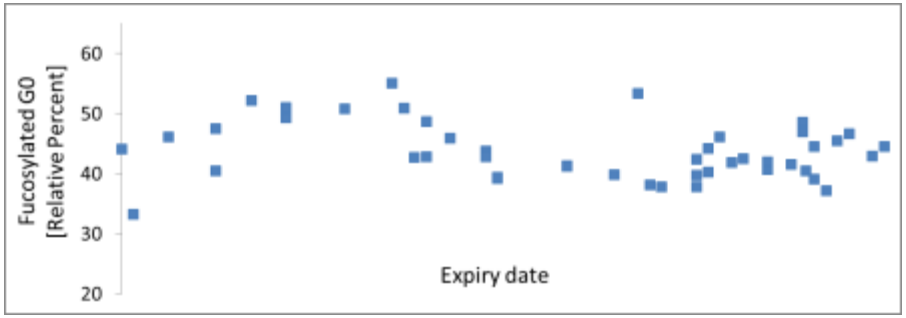
SANDOZ a Novartis company

Variability is inherent in biologics

Batch-to-batch

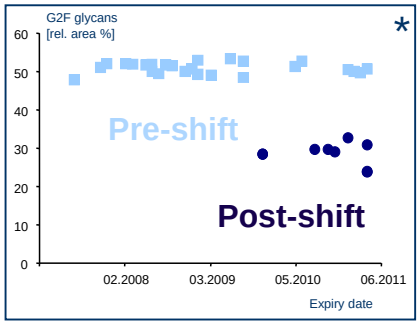
- Non-identicality is a normal principle in biologics
- No batch of any biologic is “identical” to the other batches
- Variability is natural even in the human body and usually not problematic

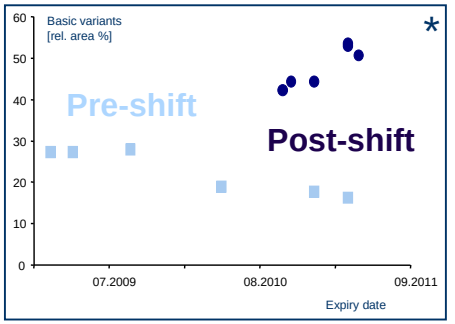
Variability of major glycan variant in commercial mAB



Manufacturing changes

- Manufacturing changes occur due to process improvements, scale up, etc
- Differences in attributes sometimes significantly larger than batch-to-batch variability





* M. Schiestl et al. Acceptable Changes in Quality Attributes of Glycosylated Biopharmaceutical; Nature Biotechnology (2011) 29:310

SANDOZ
a Novartis company

Variability is inherent in biologics

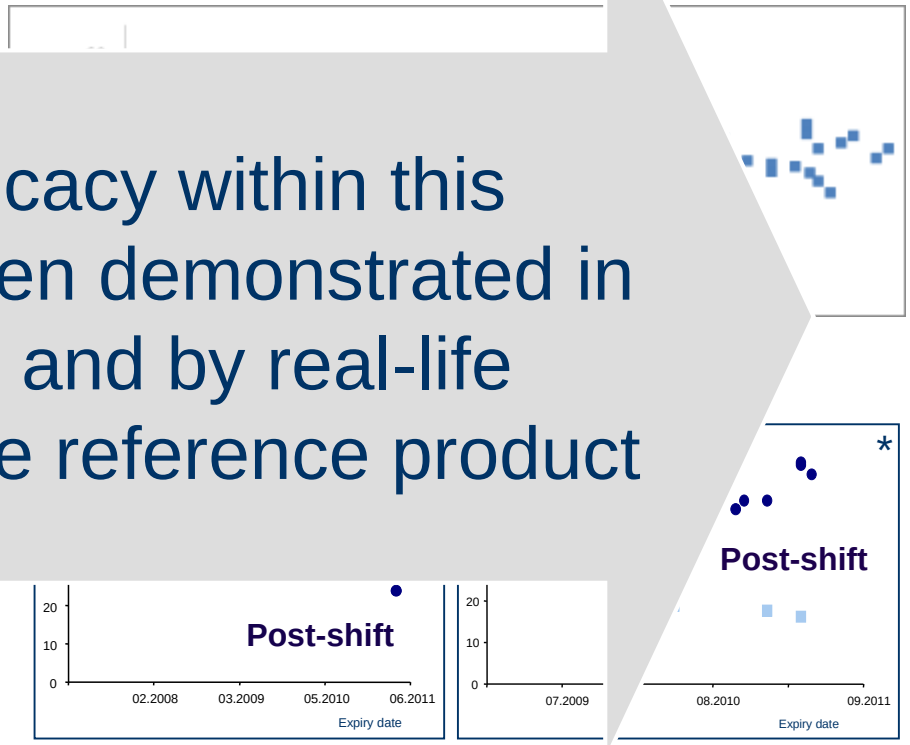
Batch-to-batch

- Non-iden ality is a normal principle in biologic
- No ba the c
- Va b

Safety and efficacy within this variability have been demonstrated in clinical studies and by real-life experience with the reference product

- Manufa process provements, scale up, etc
- Difference in attributes sometimes significantly larger than batch-to-batch variability

Variability of major glycan variant in c mercial mAB



* M. Schiestl et al. Acceptable Changes in Quality Attributes of Glycosylated Biopharmaceutical; Nature Biotechnology (2011) 29:310

SANDOZ a Novartis company

Considerations impacting biosimilarity evaluation

- The variability of the originator defines the goal posts for development
- Is there any difference between a
  - target range for development of a biosimilar
  - acceptance range for the biosimilarity exercise?
- Is every marketed batch from the originator defining acceptable quality with respect to its quality characteristics?
  - would a given quality characteristic of the originator lot be acceptable for a biosimilar lot?
- How to use the variability of the originator and the biosimilar to quantitatively assess for biosimilarity on a quality level

  → which statistical approach?

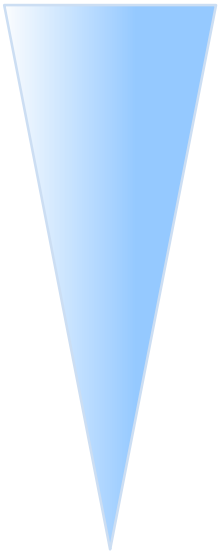
  → is statistics the deciding tool?

SANDOZ
a Novartis company

FDA‘s 3-Tier Approach

1. Evaluate the criticality of quality attributes
   - Impact on clinical performance
   - Degree of Uncertainty in Impact
2. Assign quality attributes to different tiers based on their criticality
3. Different statistical/quantitative approaches are applied to each tier

Statistical Rigor



Tier 1 – Critical QAs
Statistical Equivalence Testing

Tier 2 – Less Critical QAs
Quality Range Method: mean +/- X σ

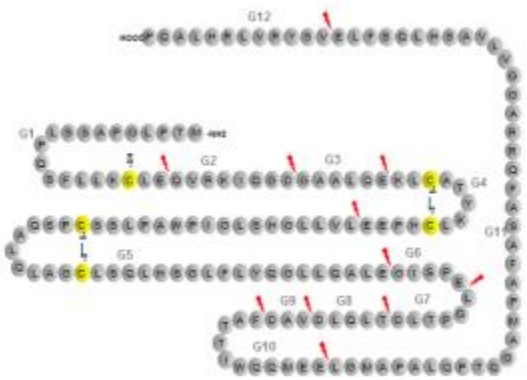
Tier 3 – Least Critical QAs
Raw Data / Graphical Comparison

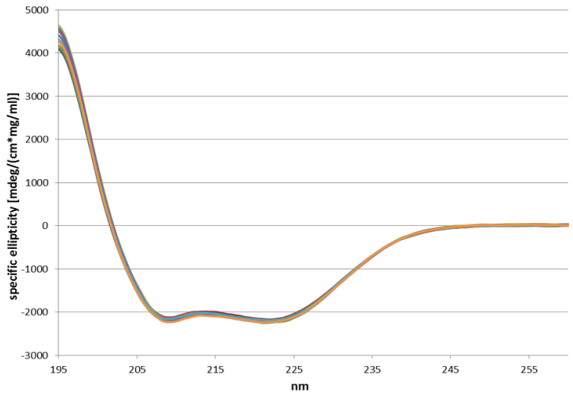
Source: X. Dong, IABS/FDA Statistical and Data Management Approaches for Biotechnology Drug Development, September 2015

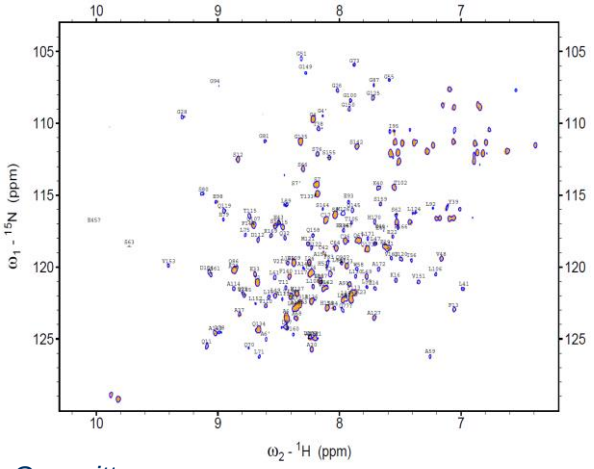
SANDOZ
a Novartis company

Not all quality attributes are evaluated best by statistical means

- Low criticality
- For some undesired quality attributes, „less than the maximum in reference product“ is better criterium than „equivalent“
  - Level of aggregates, deamidation, etc.
- For some quality attributes use of statistics less appropriate due to the nature of the data delivered by the particular analytical method





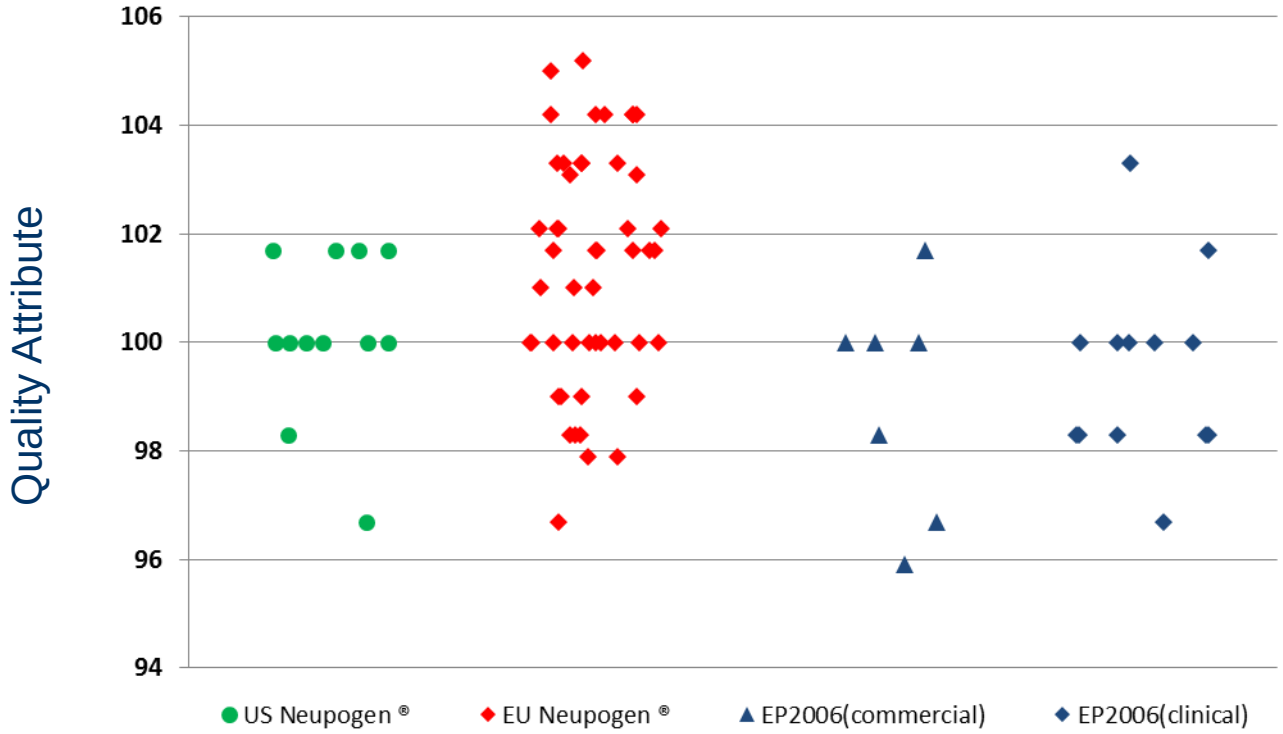


Source: Sandoz presentations for the January 7, 2015 Meeting of the Oncologic Advisory Committee

SANDOZ
a Novartis company

A graphical data comparison can already be very informative...

Comparing a biosimilar to its originator in different regions



Reference: Sandoz and FDA presentations for the January 7, 2015 Meeting of the Oncologic Advisory Committee

SANDOZ
a Novartis company

Tier 2 – Applying Quality Ranges
current practice based on reference product stdev

- Tier 2 testing is based on a quality range that depends on an estimate of the reference product standard deviation

$$\text{Mean of reference product lots} \pm k \times \sigma_{RP}$$

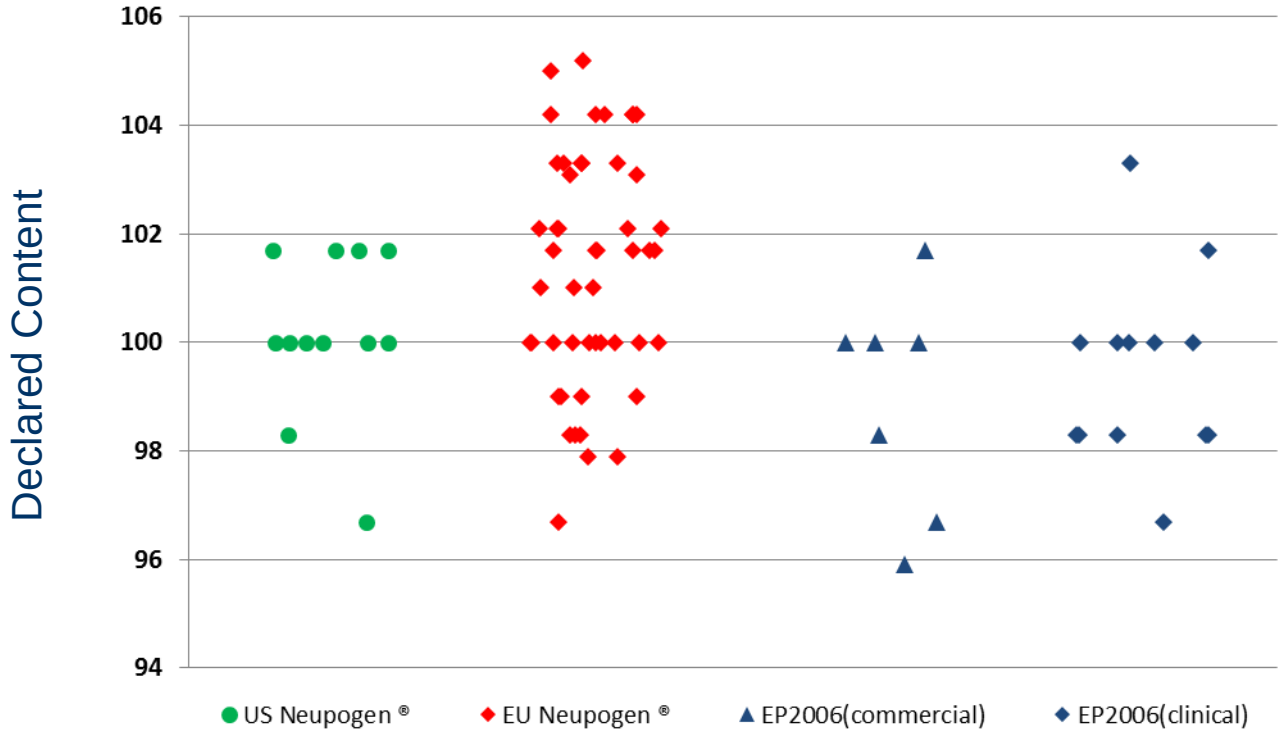
- A sufficient percentage of biosimilar batches (e.g. 90 %) required to fall into the quality range
- Multiplier k to be justified by sponsor (k may be 2,3,...)
  - difficult to find a scientific rationale for different multipliers for different quality attributes / analytical readouts other than criticality
  - k = 2 is too narrow to have a reasonable probability for two identical products (same μ and σ) to pass the criterion[1)]
  - k = 3 is widely accepted as reasonable estimator of the realistic variability
    - common standard in statistical process control
    - „three-sigma rule of thumb“: „nearly all“ values within 3 sigma [2)]

1) D. Weese & R. Burdick, IABS/FDA Statistical and Data Management Approaches for Biotechnology Drug Development, September 2015
2) Erik W. Grafarend, *Linear and Nonlinear Models: Fixed Effects, Random Effects, and Mixed Models*, Walter de Gruyter, 2006, p. 553

SANDOZ
a Novartis company

Protein content as tier 1 quality attribute

Comparing a biosimilar to its originator in different regions



Reference: Sandoz and FDA presentations for the January 7, 2015 Meeting of the Oncologic Advisory Committee

SANDOZ
a Novartis company

Statistical Equivalence Test for Protein Content

Protein content of the three product is statistically equivalent (mean values)

EP2006 vs US-Neupogen®
(-1.87, 0.15)

(-2.26 **2.26)**

EP2006 vs EU-Neupogen®
(2.89, 0.85)

(-3.23 **3.23)**

EU-Neupogen® vs US-Neupogen®
(0.27, 2.09)

(-2.26 **2.26)**

Results indicate that the products have the same strength and also support analytical similarity and analytical bridge

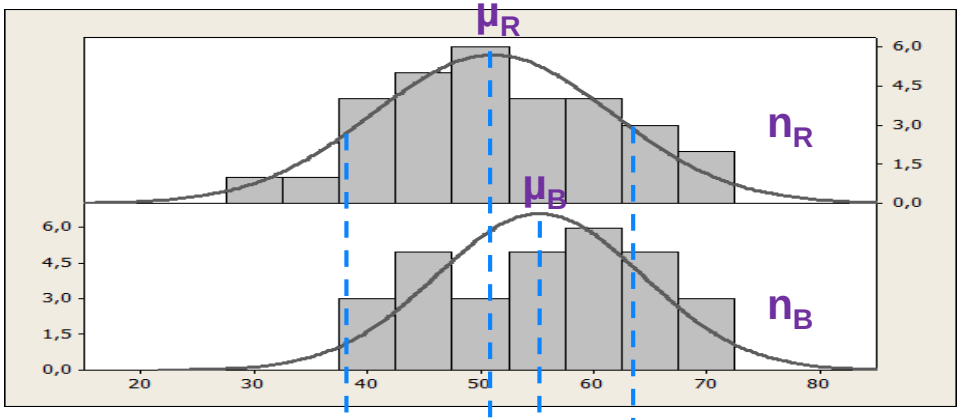
Source: FDA presentations for the January 7, 2015 Meeting of the Oncologic Advisory Committee

SANDOZ
a Novartis company

Equivalence testing for a practical difference in the means

Concluding equivalence by rejecting the null hypothesis $H:|\mu_R-\mu_B|\geq\delta$

means & sample sizes



- δ acceptable difference of the means μ_R and μ_B
- Significance level α
- Power 1-β

0

Equivalence Margin

Confidence interval
for the difference of the means

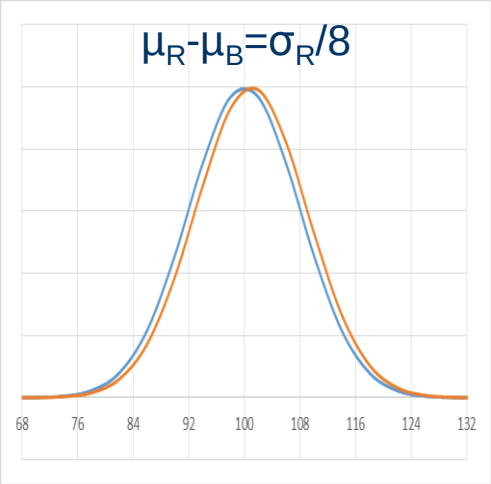
The equivalence margin interrelates strongly to sample sizes, allowable difference, significance level and power

SANDOZ
a Novartis company

A reasonable choice of the equivalence margin is key for meaningful equivalence testing

- Determining the margin is challenging:
  - scientific justification usually not feasible
  - no standard statistical approach for determining the margin
- FDA‘s proposed equivalence margin is 1.5 σ_R independent of the sample size
  - easy to implement, no power calculation

Background on the equiv. margin determination: [1)]
- Sample sizes $n_R=n_B=10$
- Difference of means between reference product and biosimilar $\mu_R-\mu_B=\sigma_R/8$
- Probability to conclude „equivalent“ (power): 87 %

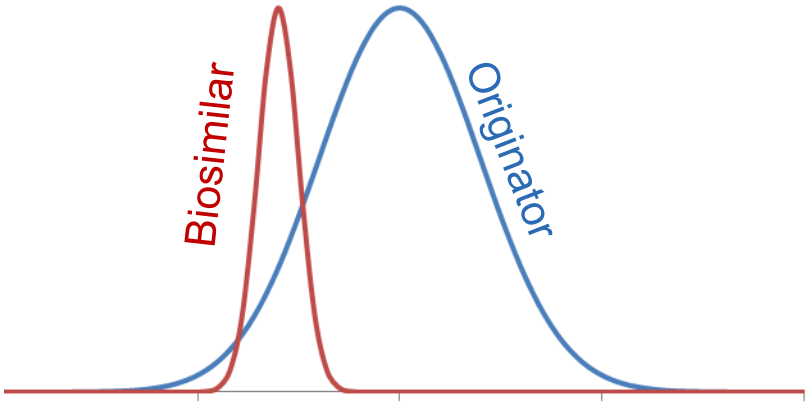


→ strong driver for large sample sizes for statistical (not necessarily scientific) reasons[2)]

1) Y. Tsong, IABS 2016; Tsong, Dong & Shen J Biopharm Stat. 2015
2) M. Horvat, AAPS 27.10.2015, Statistical Analysis of Comparabiltiy Data

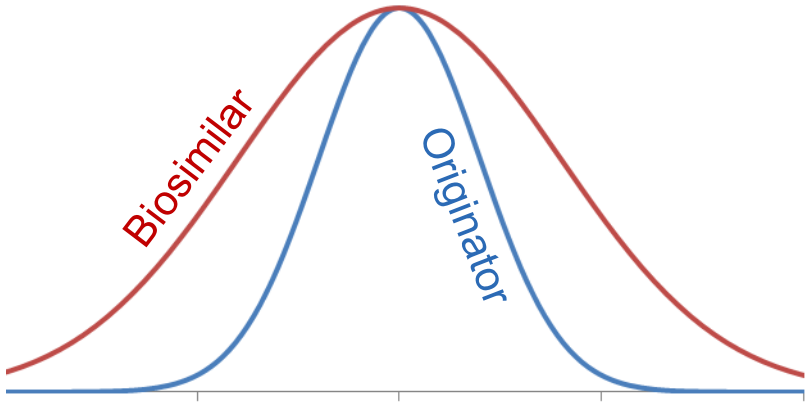
SANDOZ a Novartis company

The conceptual & theoretical implications of equivalence testing



All biosimilar batches are within variability of originator

means are different → not equivalent



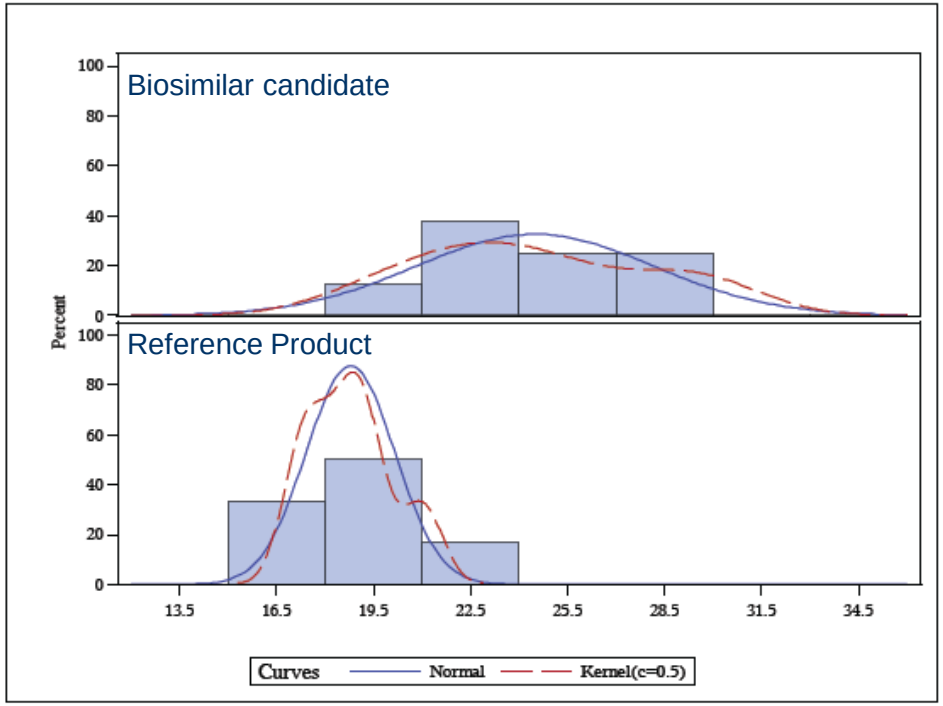
Some biosimilar batches are outside of the variability of originator

means are the same → equivalent

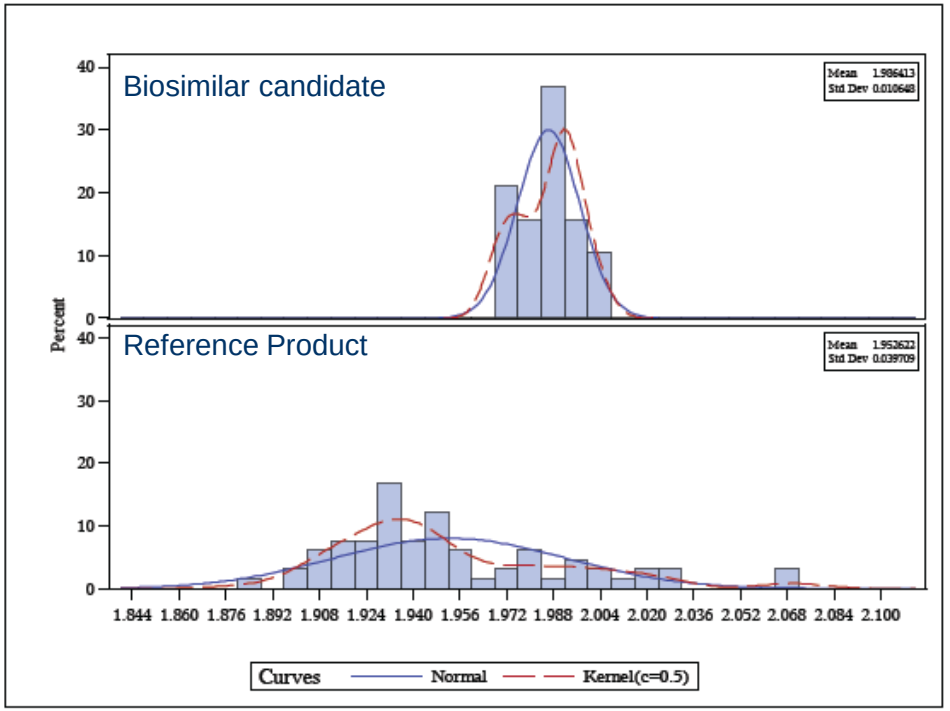
SANDOZ
a Novartis company

The practical obstacles for statistics

Very low sample sizes
& analytical variabiltiy



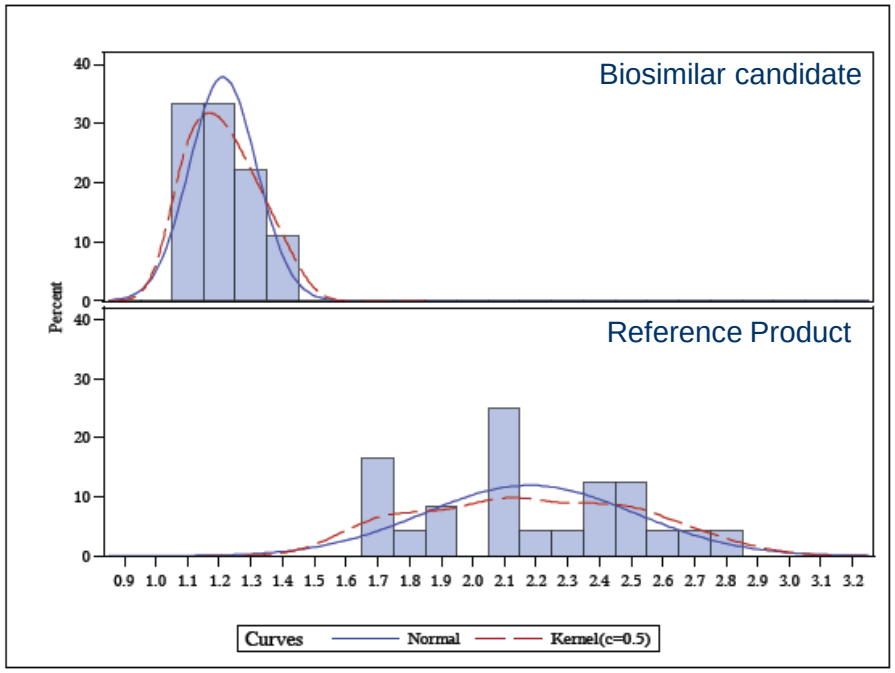
Non-normal distributions



SANDOZ
a Novartis company

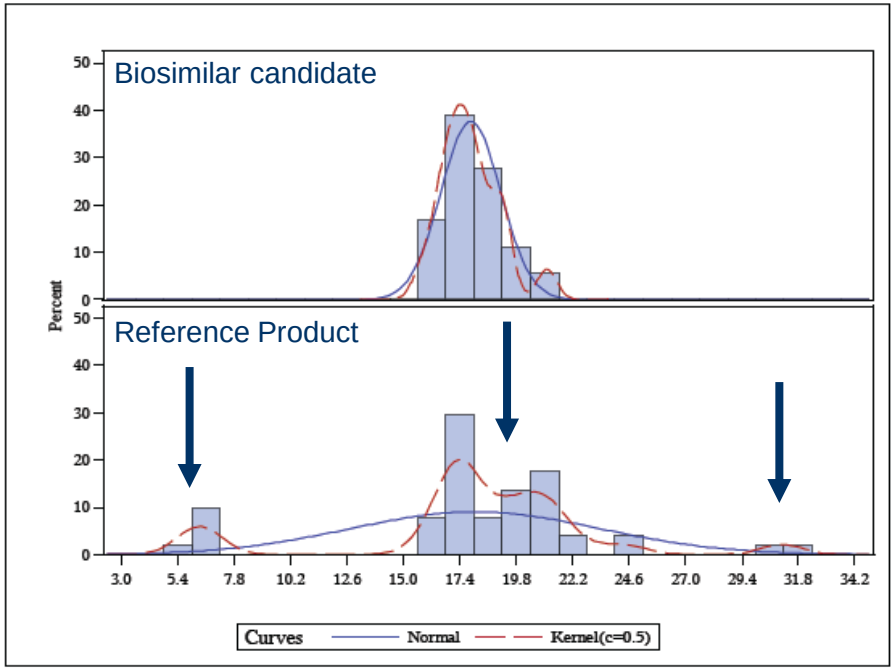
The practical obstacles for statistics

Undesirable quality attributes (less is better)



More than one reference product population & „outliers“

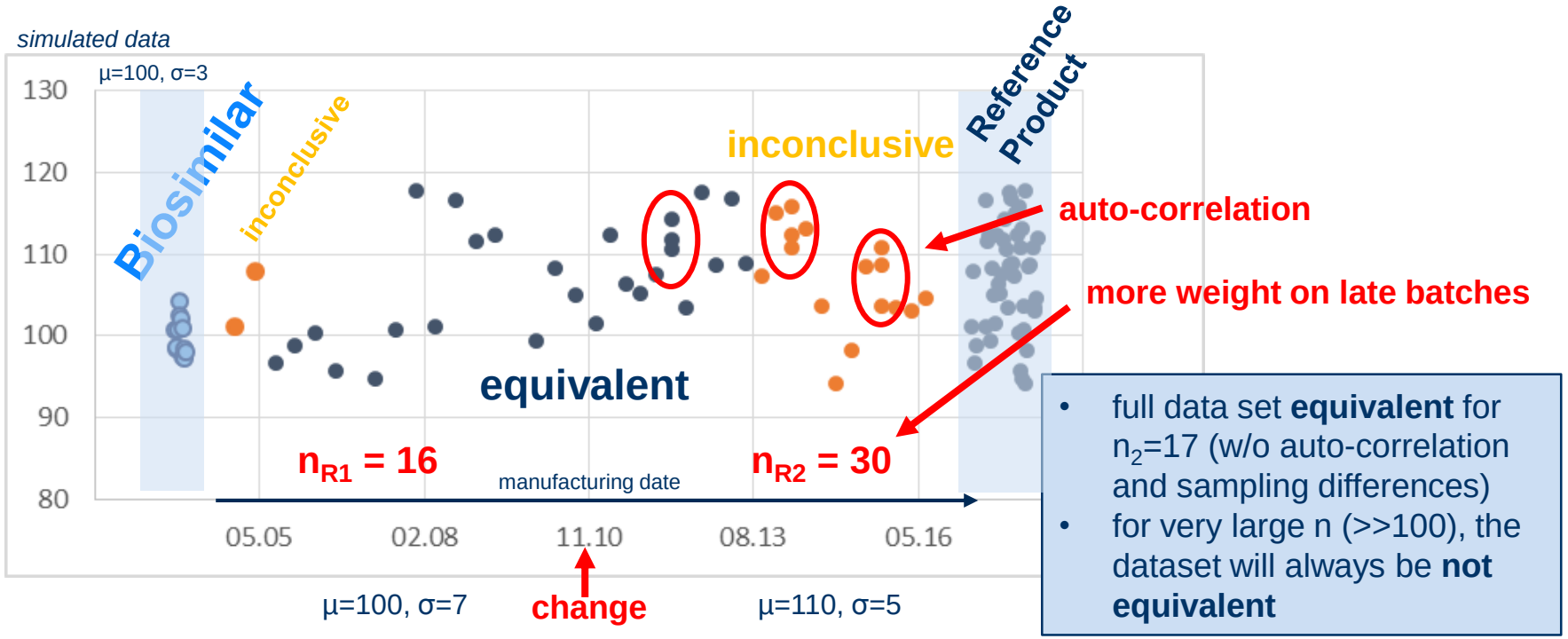
due to manufacturing changes, long-term common cause variability, or special cause variability



SANDOZ a Novartis company

The pitfalls of diligently testing the reference product

Additional batches	Sampling differences	Auto-correlation
manufacturing change, long-term common cause variability	Different weight on different times	campaign production, 1 DS in several DP batches,...
→ shift in mean	→ shift in mean	→ decrease of stdev



- full data set **equivalent** for n_2=17 (w/o auto-correlation and sampling differences)
- for very large n (>>100), the dataset will always be **not equivalent**

SANDOZ a Novartis company

Implications of sampling reference product lots across many years

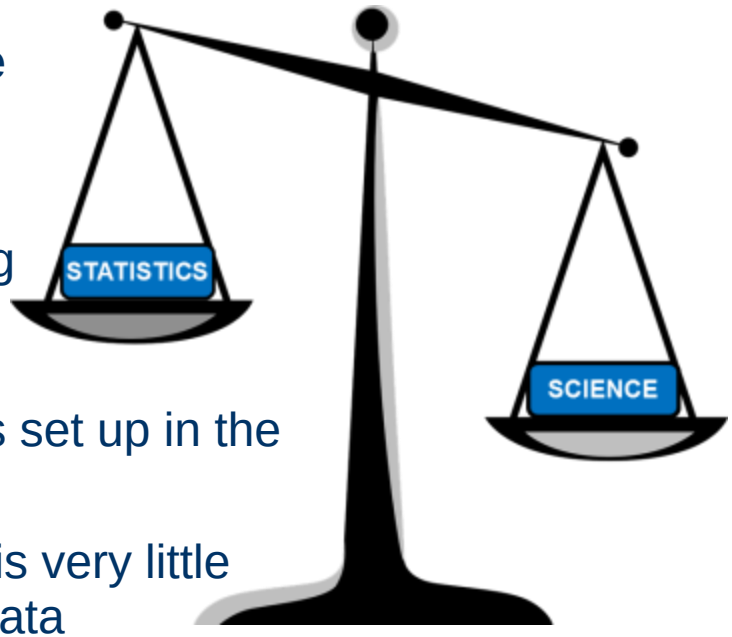
- Batch purchasing schedule (sampling) may impact the equivalence test
- If a biosimilar tests as equivalent – be cautious not to test it into inconclusive by additional reference product batches
- Equivalence testing does not allow for the definition of a useful development target

An inspirational comparison of the use of statistical tools for clinical studies vs. comparability

Clinical studies	**CMC comparability**
One primary endpoint	Multiple endpoints: quality attributes
Measure the physiological reaction after drug application	Measure the quality attributes of a given drug
Variability of the physiological processing Stratified random sampling	Variability of the manufacturing process Difficult to assure independent data
Acceptable margin for the primary endpoint based on clinical relevance	Acceptable margin based on scientific rationale -> Different for each quality attribute
▪ In the clinical evaluation, the predefinition of the endpoint and its related statistical evaluation is inevitable to mitigate the risk for bias ▪ In a comparability exercise, the endpoints are already set by the CQA assessment -> no risk for bias in selecting the „wrong endpoint“	
Statistics required for final judgment	Statistics merely facilitator to describe the level of residual uncertainty and thus the level of justification needed in case of differences

Final thoughts...

- First be clear about your scientific question, then choose the statistical tool, **and be aware of the limitations**
- With carefully chosen statistical test parameters, all described tools are able to flag those quality attributes which need further evaluation
  - May speed up final evaluation if statistics is set up in the right way
  - However, the incremental knowledge gain is very little compared to a descriptive but critical raw data comparison



- Failure to pass a statistical test does not preclude similarity. It is a tool to decide which parameters have to be discussed/investigated in more detail.
- Statistics should not be a self-contained claim for biosimilarity on the quality level, it always should be just a contributor to the totality of evidence

SANDOZ
a Novartis company

Thanks for listening!

Thanks for contributing!

Colleagues at Mengeš, Schaftenau & Kundl, Oberhaching & Holzkirchen

All registered trademarks ® are the property of their respective owners

© 2016 Sandoz. All rights reserved.

SANDOZ
a Novartis company