Federal Institute
for Drugs
and Medical Devices

Quality and Statistics: bringing two worlds together

Andreas Brandt

Disclaimer

The views expressed in this presentation are the presenter‘s personal views and not necessarily the views of BfArM or EMA

I am a statistician, not a quality expert

Introduction

- Quality experts have long experience in making comparability assessment of QAs
  - Comparability concept developed in 1990s for evaluation of manufacturing changes
  - Transferred to biosimilar concept in the 2000s
- More sophisticated ‘statistical’ methods were applied
- Increasing importance of the comparability exercise for totality of evidence of biosimilarity

  → Statisticians were increasingly involved
- Languages and traditions of quality experts and statisticians are different

Lost in translation

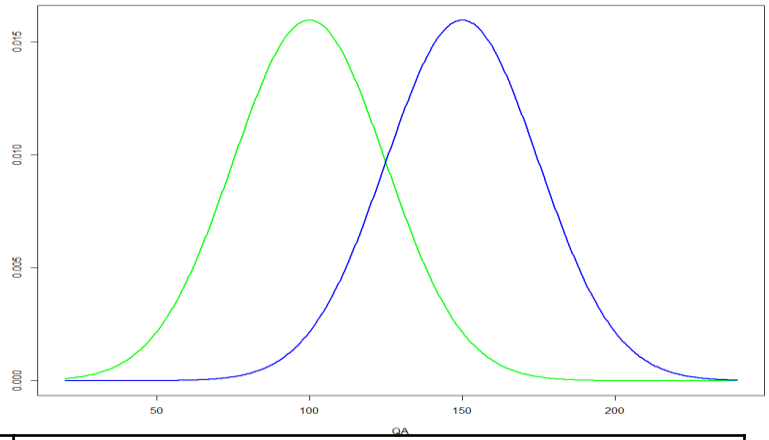
Mathematicians are a sort of Frenchmen; if you talk to them, they translate it into their own language, and then it is immediately something quite different.
(Johann Wolfgang von Goethe)

Quality tradition: Comparability ranges

- Comparability range = quantitative range established based on the measured quality attribute of reference medicinal product batches
  - Min-max, mean +/- X*SD, tolerance intervals
- Test batches should be included, differences justified

Reference range: mean ±3SD$_{ref}$

- How often (in %) will the mean ±3SD$_{ref}$ reference range contain all sampled test batches?



N_{test}=5	41.6%
N_{test}=10	25.0%
N_{test}=15	15.0%

- $N_{reference}$=30; SD_{test}=SD_{ref}=25

test product
reference product

Statistical tradition: equivalence testing

- Absence of relevant differences between two products
  - Quality: ‘Comparability‘
  - Statistics: ‘Equivalence‘
- Equivalence testing (of means)
  - demonstrating equivalence of parameter(s) describing test and reference distribution
  - First FDA draft GL (withdrawn)
  - Draft EMA RP: not explicitly recommended, but perceived in this way

Quality experts were unhappy…

- Misunderstandings of ‘equivalence‘ concept
  - Equivalence ≠ equal
- Equivalence testing of means is wrong concept
  - Fixed mean may not exist (shifts and drifts)
  - Equivalence of means is neither sufficient nor necessary for similarity – not center but range of reference distribution is important
  - No basis for justification of an equivalence margin

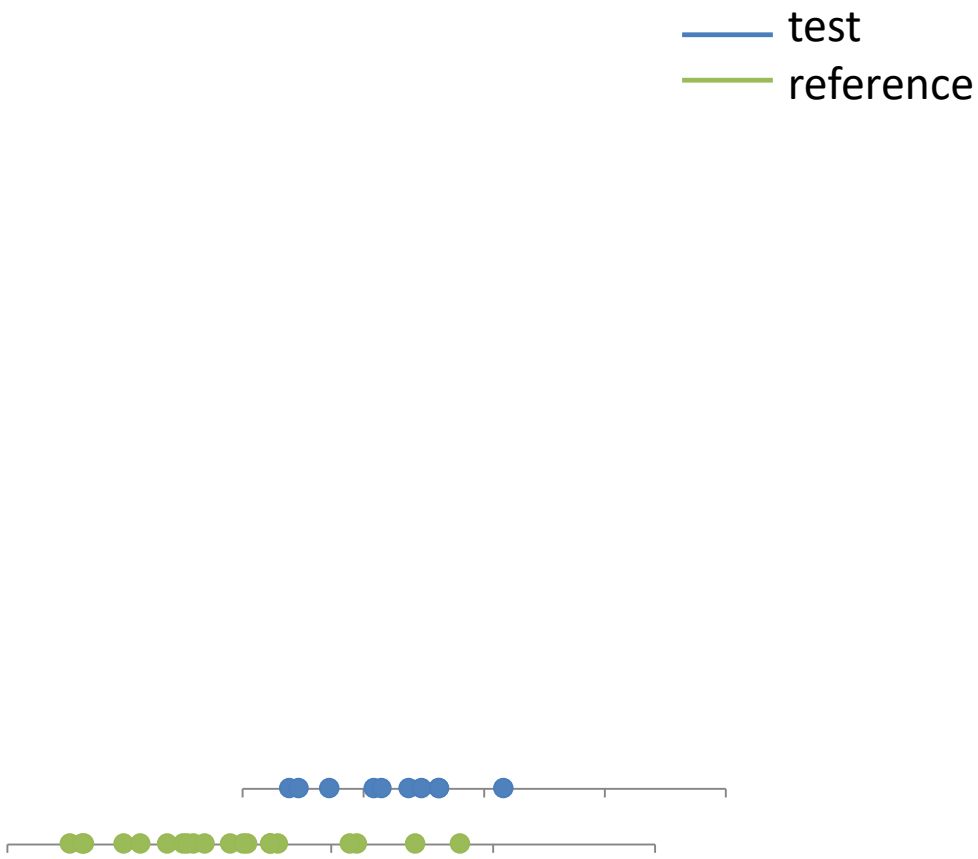
Looking for common basis

- Both sides have good arguments that simple concepts of the other side may have serious limitations
- What is the basis for coming together?
  - Achieve a common understanding of the question to be answered
  - Understand the properties of statistical approaches to answer it
  - Achieve a common understanding what role statistics can/should play for totality of evidence

What is the question?

- Overall aim: demonstration that test and reference product are comparable
  - EMA GL: The aim of the biosimilar comparability exercise is to demonstrate that the biosimilar product and the reference medicinal product chosen by the applicant are similar
- Traditional comparability assessment: strong focus on test batches at hand

Descriptive analysis



What is the question?

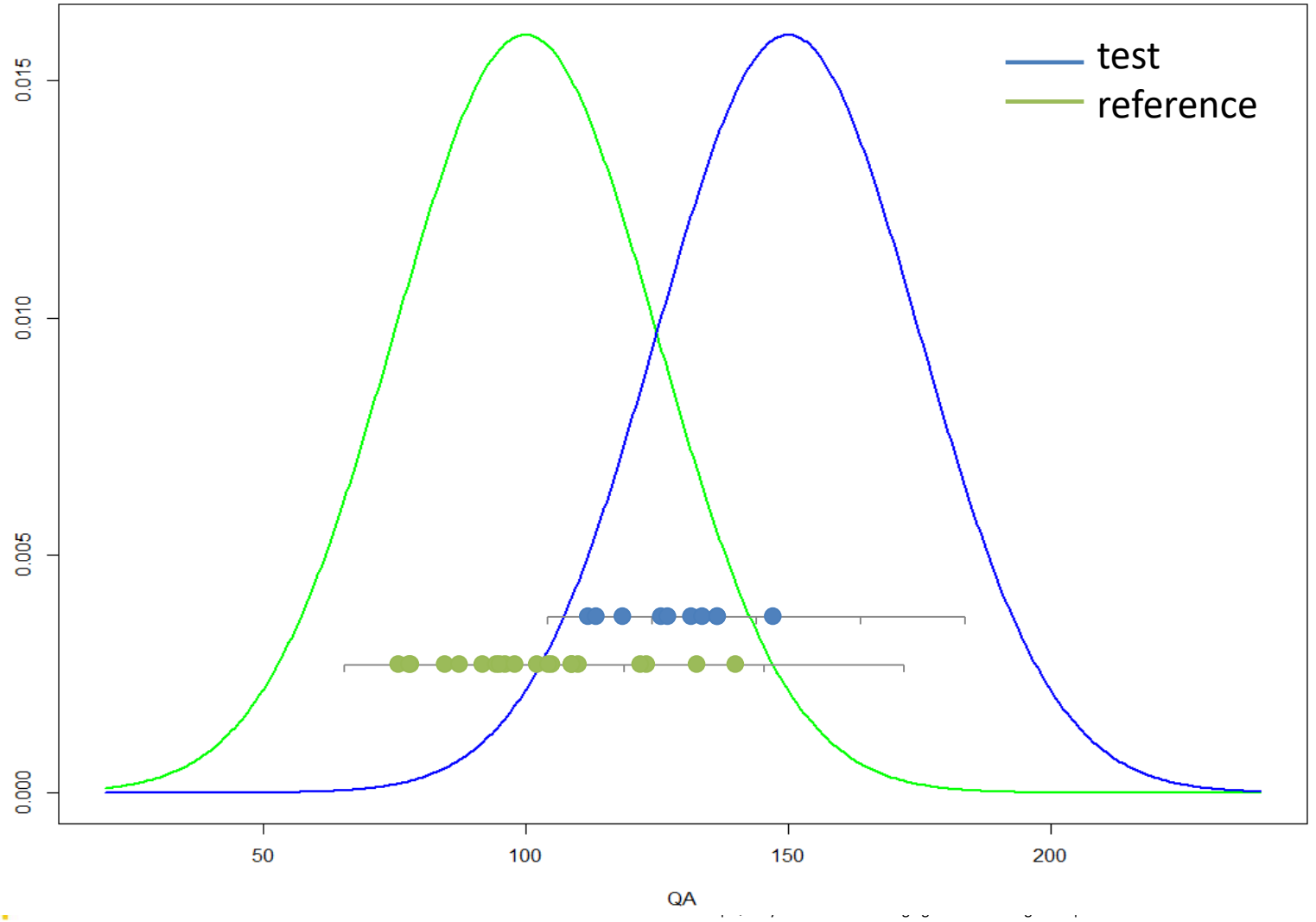
- Overall aim: demonstration that test and reference product are comparable
- Traditional similarity assessment: strong focus on batches at hand

  → Descriptive analysis (=describing what we observe)

  → possible conclusion: biosimilar developer was able to produce batches with comparable quality
- Is this sufficient for drawing conclusion on comparability of test and reference **product**?
  - Product is more than limited number of batches

  → Inferential conclusion: what is observed (batches) is used for a conclusions about an 'underlying truth' (product) that cannot be (completely) observed

Inferential question



What is ‘inferential‘?

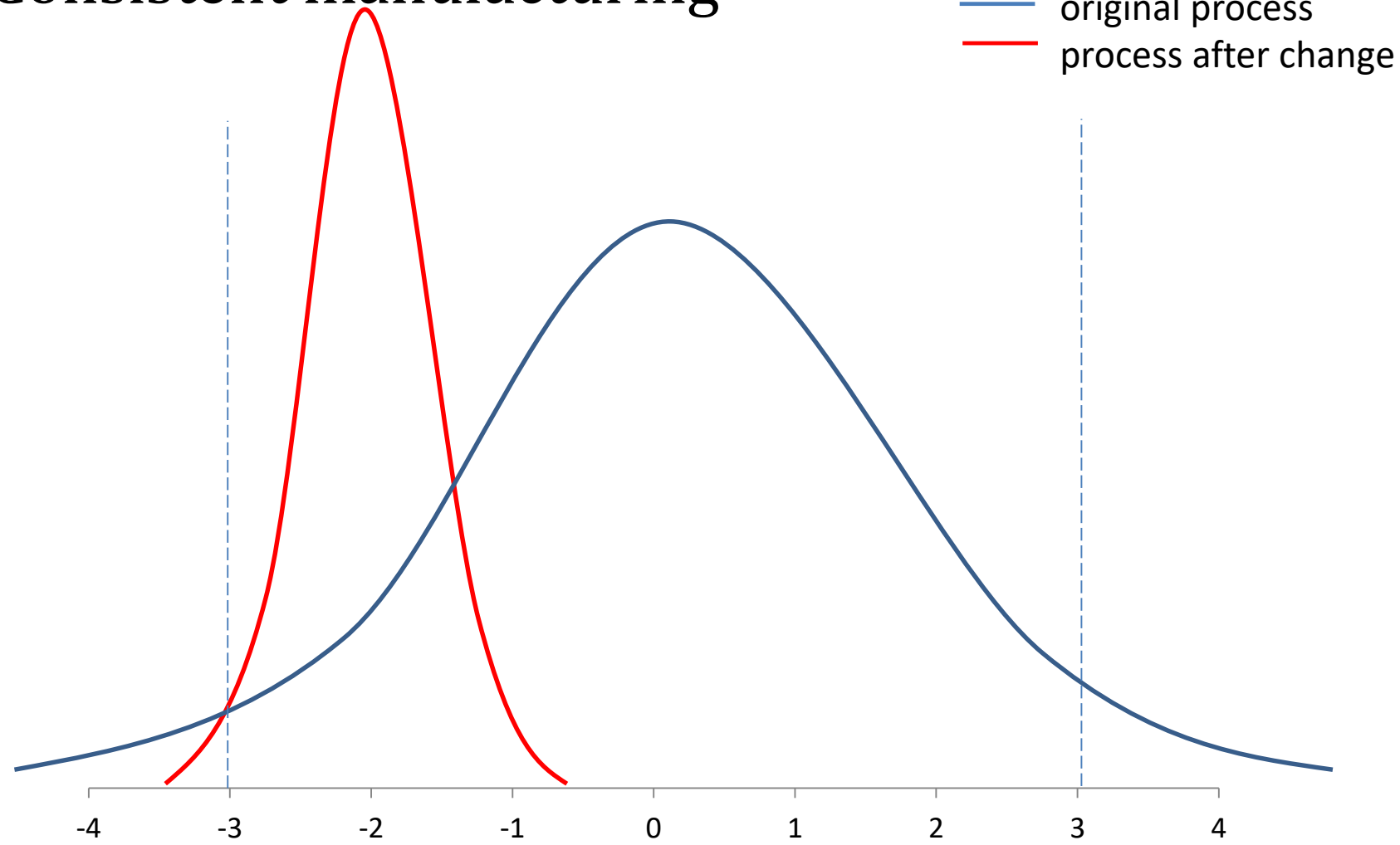
- Important learning from EMA workshop: understanding of term ‘inferential‘
  - Analysis being ‘descriptive‘ or ‘inferential‘ does not depend on applied method/similarity criterion but on conclusion that is drawn

  → Making conclusion about product going beyond batches at hand = inferential
- Similarity exercise being inferential unavoidable?
  - Manufacturing control system in place

What is a “product“?

- “The product is the process“?
    - Process can change over time
- Product = all material produced by consistent manufacturing process
- Consistent manufacturing process:
    - Process that is allowed to vary (randomly and systematically) but is controlled to stay within acceptable limits

Consistent manufacturing



Federal Institute
for Drugs
and Medical Devices

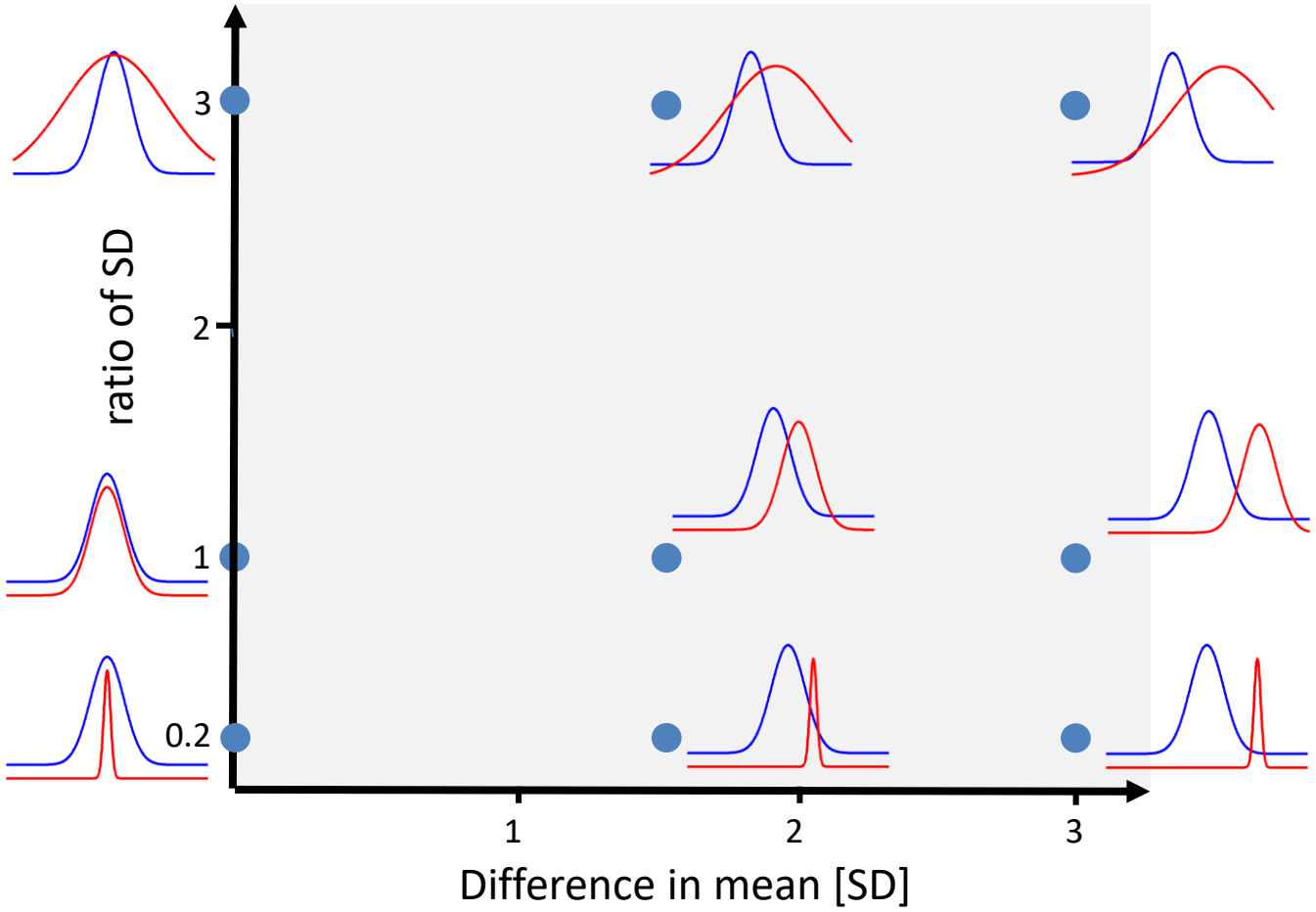
Importance of operating characteristics

- Several criteria/methods to decide on similarity of products – how to compare?
- No criterion will make ‘right’ or ‘wrong’ decisions in all situations
- Understanding operating characteristics of a decision criterion is key
    - Probability to conclude similarity for non-similar products
    - Probability to conclude similarity for similar products

Framework for comparing OCs of similarity criteria

- Defining range of relevant underlying scenarios
    - Underlying distributions for products to be compared
- Quantitative definition of true similarity
- Sample size (number of test and reference batches)
- Summary measure(s) to describe OCs

Framework: Range of scenarios of interest



modified from Stangler, 2018

Framework: Definition of true comparability

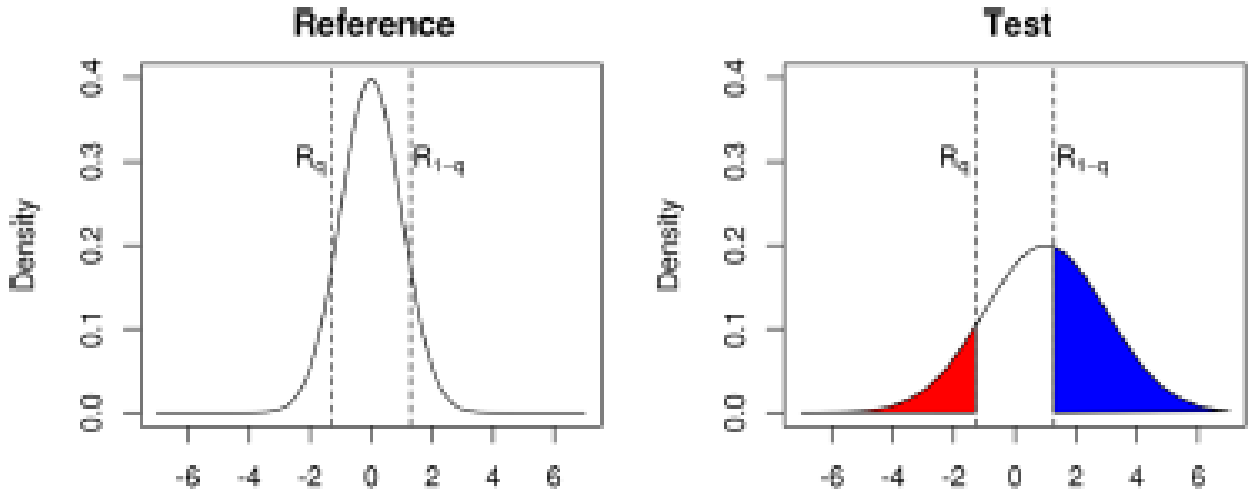
- Inferential decision on comparability requires quantitative definition of ‘true comparability‘

→Concept of comparability needs to be translated in quantitative terms

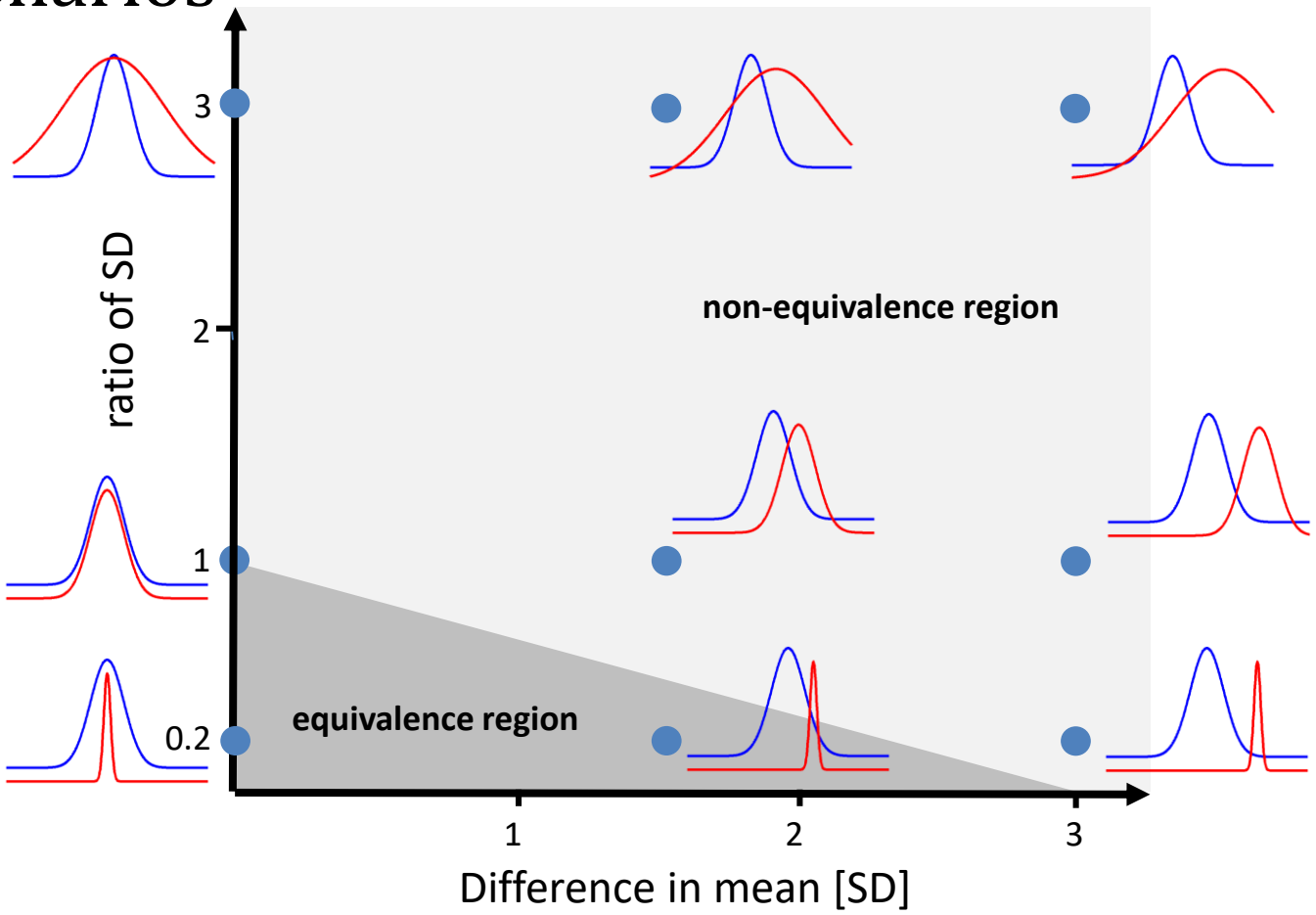
- Translation will never be perfect
  - Simplifying assumptions needed
  - ‘Statistical comparability‘ is not necessarily the same as ‘comparable’ in quality GLs

Proposals for formalisation

- Similarity of distribution parameters (mean, SD)
- "Population within population"
  - $Mean_{Test}$ +/- X*SD_{Test} within $Mean_{Ref}$ +/- X*SD_{Ref} (Stangler, 2018)
  - Tail criterion: $2q - (P(X_T < R_q)+P(X_T>R_{1-q})) \geq c$ (Mielke et al., 2019)

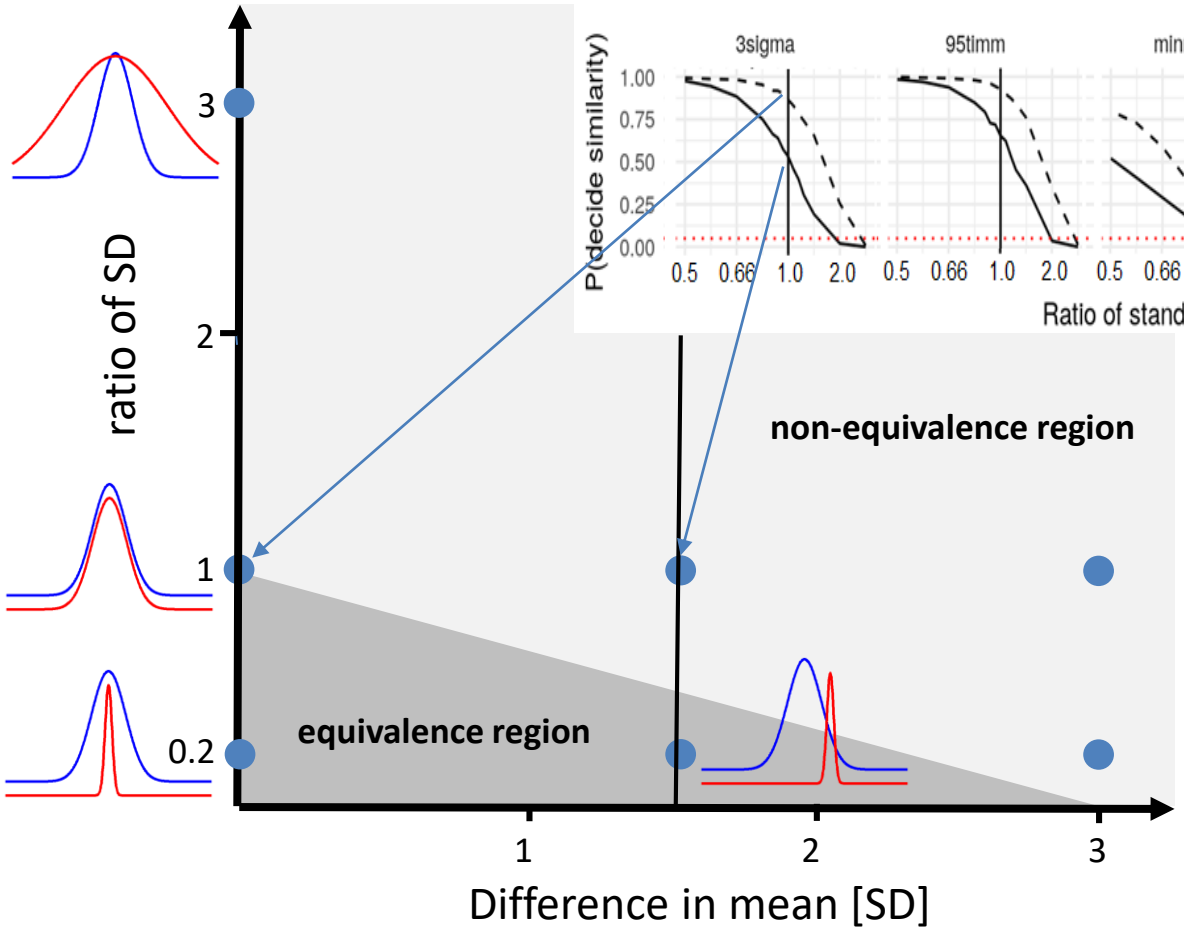


“True“ comparability for relevant scenarios



modified from Stangler, 2018

Comparing OCs



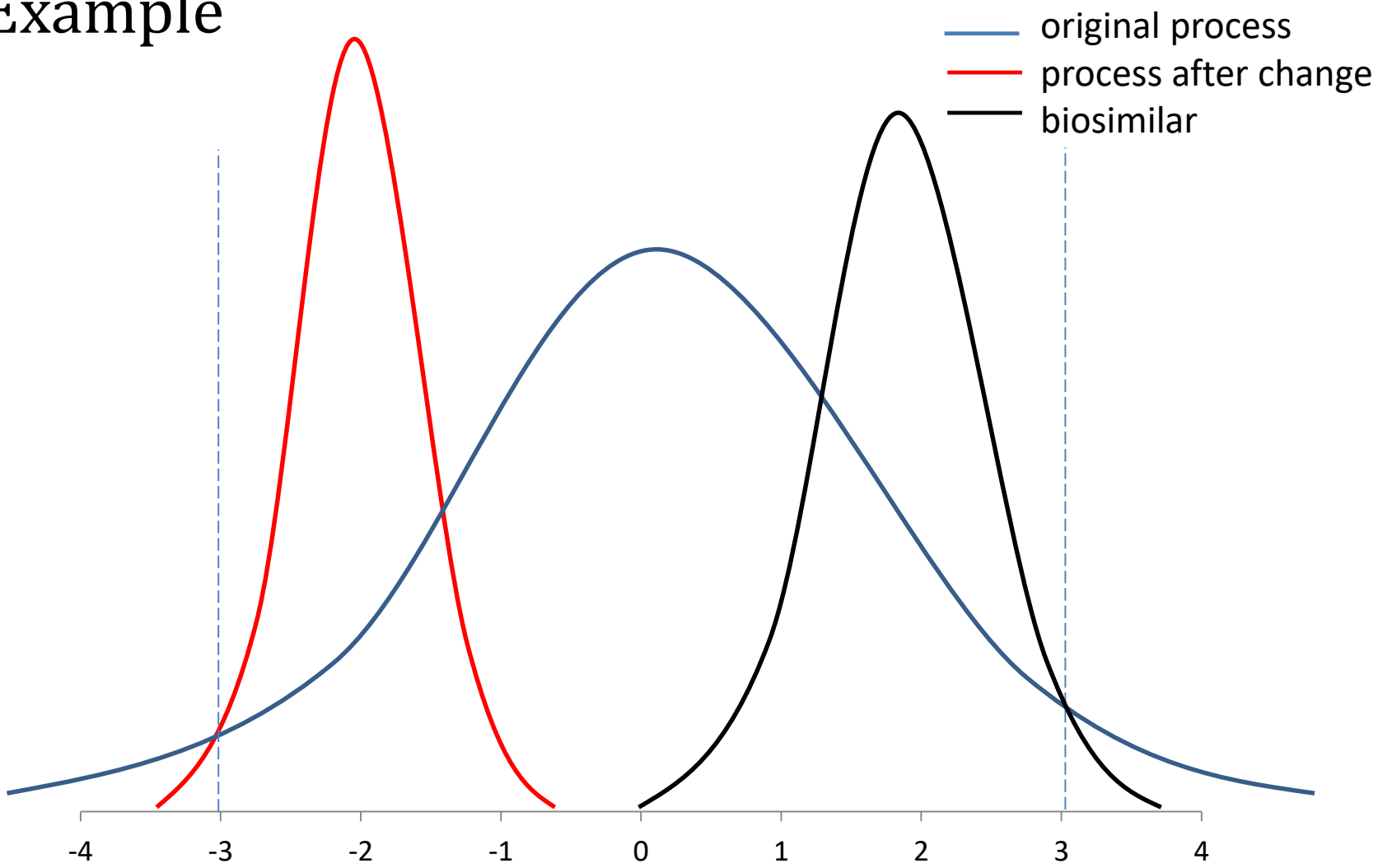
From: Klinglmueller, 2018

Shifts and drifts: what is truly comparable?

- So far: relevant scenarios assume ONE underlying NORMAL distribution for test and reference
- Starting point: Quality experts insist that mean and variance can vary over time

→ Shifts and drifts in reference (and in test) add to complexity

Example



Federal Institute
for Drugs
and Medical Devices

Shifts and drifts: operating characteristics

- Shifts and drifts may not be detectable from the data
- Shifts and drifts will influence operating characteristics
- How many shifts/drifts?
    - Any shift adds 1-2 distribution parameters
- Number of batches sampled before/after shift?
- Possible to define meaningful restrictions?

Role of statistics

- Overall aim: making decision on similarity
  - Based on totality of evidence
  - Statistical methods contribute to totality of evidence
  - Statistical decision should not be a pass/fail criterion
- But what role could statistics actually play?
- What are the consequences of passing/failing
  - Passing: statistical decision reliable for accepting similarity, for reducing non-clinical/clinical data requirements?
  - Failing: What kind of justification can be accepted to accept similarity in spite of failing, when are additional data required?

Summary

- Continued collaboration of quality experts and statisticians required
- Improve understanding of question
- Develop frameworks for comparing OCs of similarity criteria
- Agree on role of statistics for descision making

Acknowledgements

Reflection paper revision task force
Thomas Lang (AGES)
Ina-Christine Rondak (EMA)
Kit Roes (Radboud UMC Nijmegen)
Florian Klinglmüller (AGES)
Niklas Ekman (FIMEA)
Martijn van der Plas (MEB)
Tiina Hakonen (FIMEA)
Steffen Groß (PEI)
Sean Barry (HPRA)
Jobst Limberg (BfArM)

BfArM
Norbert Benda
Ann-Kristin Leuchs
Jörg Zinserling
Astrid Schäfer
Brigitte Brake
Ute Fischer
Katrin Buss
Cornelia Lipperheide
Birgit Schmauser

Members/observers of EMA BSWP
+ BSWP secreteriat

References

CHMP. Draft Reflection paper on statistical methodology for the comparative assessment of quality attributes in drug development. 2017

CHMP. Meeting Report: Workshop on the draft reflection paper on statistical methodology for the comparative assessment of quality attributes in drug development. 2018

CHMP. Similar biological medicinal products containing biotechnology-derived proteins as active substance: quality issues. 2014

ICH Q5E Biotechnological/biological products subject to changes in their manufacturing process: comparability of biotechnological/biological products. 2005

Mielke J et al.. The Assessment of Quality Attributes for Biosimilars: a Statistical Perspective on Current Practice and a Proposal. AAPS J 2019

Stangler T, Schiestl M. Similarity assessment of quality attributes of biological medicines: the calculation of operating characteristics to compare different statistical approaches. AAPS Open 2019

Stangler T. Performance characteristics of quality range methods and equivalence testing in the comparative assessment of quality attributes. Presentation at Workshop on the draft reflection paper on statistical methodology for the comparative assessment of quality attributes in drug development, https://www.ema.europa.eu/en/events/workshop-reflection-paper-statistical-methodology-comparative-assessment-quality-attributes-drug

Klinglmueller F. Operating characteristics of frequently used similarity rules. Presentation at Workshop on the draft reflection paper on statistical methodology for the comparative assessment of quality attributes in drug development, https://www.ema.europa.eu/en/events/workshop-reflection-paper-statistical-methodology-comparative-assessment-quality-attributes-drug

Thank you very much for your attention!

Contact

Federal Institute for Drugs and Medical Devices
Division Research, Unit Biostatistics and Special Pharmacokinetics
Kurt-Georg-Kiesinger-Allee 3
D-53175 Bonn

Contact person
Dr. Andreas Brandt
andreas.brandt@bfarm.de
www.bfarm.de
Tel. +49 (0)228 99 307-3797

Inferential question