Bundesamt für
Sicherheit im
Gesundheitswesen
BASG

EMA draft Reflection Paper* after Consultation and Workshop, Outcome and Learnings:

What if comparative statistical analysis of “non clinical“ data gets highly relevant for regulatory decision making?

*Reflection Paper on Statistical Methodology for the Comparative Assessment of Quality Attributes in Drug Development

Thomas Lang

Biostatistician, Senior Expert, BSWP-Rapporteur for the Reflection Paper*

Austrian Agency for Health and Food Safety, BASG, AGES

Disclaimer

Content of this presentation reflects personal opinion in the role as rapporteur for the Reflection Paper.

Remarks do not necessarily reflect the official view of AGES/BASG or EMA.

Content

- Triggers & History of RP development
- RP scope and content
- Public Consultation: main reactions
- EMA Workshop, brief review
- EMA Workshop learnings
- Next steps: working with the input

Abbreviations

- **RP**: Reflection Paper
- **WS**: Workshop
- **(C)QA**: (Critical) Quality Attribute
- **BS**: Biosimilar Product (candidate)
- **T/Test**: Test Product
- **RMP**: Reference Medicinal Product
- **OC**: Operating Characteristic

Triggers for the RP

Regulatory statistician and quality data

- For biosimilars: as assessors we were/are **(the only ones) involved in all areas of comparative data analysis**: Quality and Clinical
- We are **familiar with rigor of assessment** at clinical level, we sometimes see huge discrepancies in terms of rigor applied comparing QA data
- In assessment work, we have **responsibility to flag flawed QA data comparison** approaches! (This does not necessarily mean we have an alternative solution!)

Triggers for the RP

Differences in problem understanding

- What can be taken as **evidence for similarity**?
- Statistical understanding driven by quality control methods and release testing
- **“Sample-focused” way of thinking**: “If all sample points are ‘in’, there is no problem”
- Statistical inference: “I’m least interested in the samples taken” (S. Day)
- “The product is the process!” - But **what exactly is ‘the process’?**
- Is there always a way out? Allowing **justifications in case of differences** vs required discipline when applying similarity criteria;

Actions taken

Concept Paper – Reflection Paper – Workshop

- Need to write down regulators position
- Many parallels in different settings → broad scope
- Platform for interaction between quality experts and statisticians
- EMA Concept Paper development in 2013/2014
- EMA RP development 2014 – 2017
- BSWP in lead, BWP, BMWP and QWP "in loop"
- CHMP adoption Mar 2017
- Public Consultation until Mar 2018
- EMA Workshop 3/4 May 2018

  https://www.ema.europa.eu/en/events/workshop-reflection-paper-statistical-methodology-comparative-assessment-quality-attributes-drug

Further triggering observations and open questions

- **Many parallels** between ‘biosimilars case’ and ‘pre-post manufacturing change’ setting
- Often: Deal with manufacturing changes in Biosimilar developments
- The **same rigor** for demonstrating similarity required?
- Small molecules: open issues regarding alternative data analysis approaches to show similarity in dissolution (**dissolution as ‘special case**’)
- Small molecules, **special formulations**: showing similarity by other models, e.g. in-vitro-tests

Reflection Paper

Scope

Scope:

- Pre/post change comparisons
- Biosimilar vs RMP
- Other settings including special cases small molecules

Out of scope:

- Criticality assessment (CQA selection)
- Process control methodology

Reflection Paper

Content

Bundesamt für
Sicherheit im
Gesundheitswesen
BASG

Section 4: Description of settings (according scope)

Section 5: Statistical perspective: list of important aspects if statistical inference should be applied:

- parameters of interest (are there any?)
- sources of variability
- sampling, unit of observation
- distance metrics
- acceptance ranges
- quantifying uncertainty in estimation, statistical intervals

Section 6: Potential Implications for planning and assessment

Public consultation

- Comments from 15 stakeholders
- Range from individuals to consortia/organizations
- >100 pages general comments

- Concerns/Reservations
- Conflicts/Shortcomings
- Proposals

Public consultation comments

Concerns

- **“Statistical testing should not become a pass/fail criterion without reflection of context and involvement of CMC experts!”**

- **Totality of evidence-based decision making** shall not be endangered
- **Scope of reflection paper is too broad**, we need more specific reflections dependent on the setting
- Where is the **gap for biosimilars?** We have a good system in place
- Application of statistical methodology shall not paralyze the pharmaceutical industry
- Content of reflection paper expected to have **adverse implications for healthcare systems and payers**

Public consultation comments

Shortcomings

- RP gives no answer to question: **What is similarity?**
- Can „**consistent manufacturing**“ ever be compatible with shift/drift in means?
- „Equivalence testing of means” meaningful in presence of **shift/drift in means**?
- RP **promotes equivalence testing of means**, contradicts other ICH guidance
- **Statistical tools using intervals** are unnecessarily depreciated
- Considering **lower variability for the BS** a problem contradicts GL
- Systematic **within-specs changes** not addressed in RP
- Dependence drug product - drug substance not sufficiently addressed
- (Unknown) **age of RMP batches important source of variability**, not sufficiently addressed

Public consultation comments

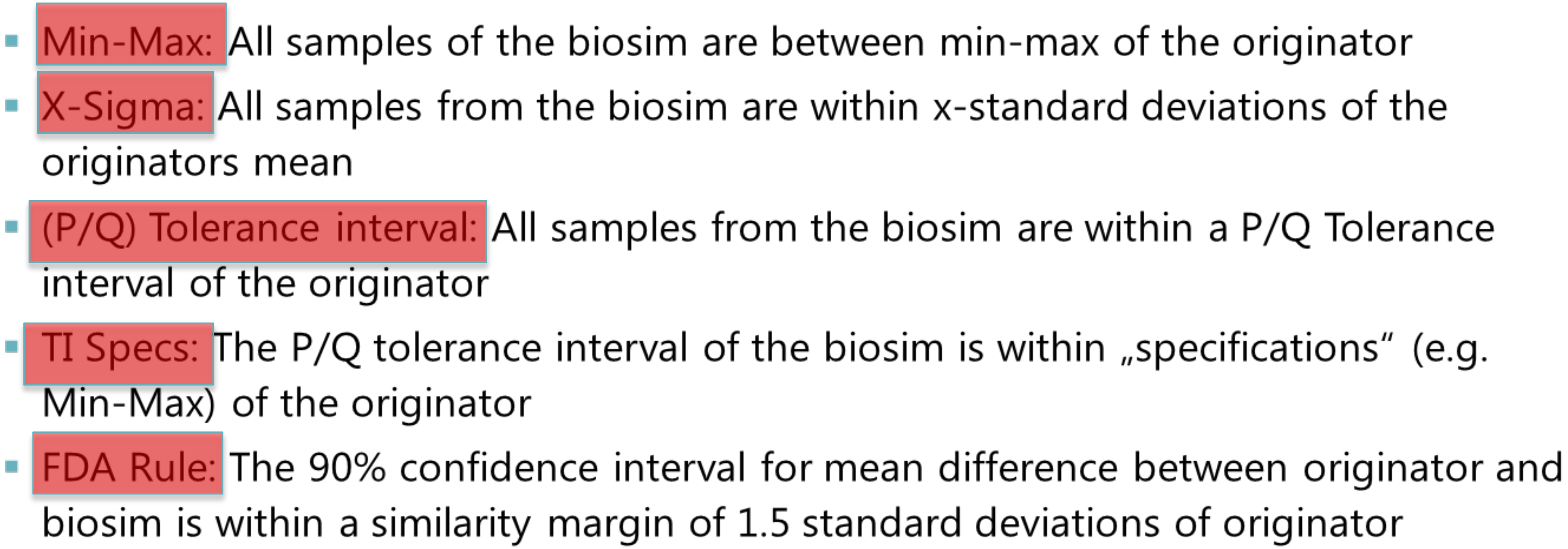
Proposals

- RP should be **transformed into Guideline(s)**
- No new concepts if not sufficiently **proven to be fit for purpose** (*understand OCs ?)*
- More in depth exploration of potential of **Bayesian methodology**
- Look at (underlying) **manufacturing process distribution**
- **Highlight connection to clinical relevance**, add aspects on criticality assessment
- **Differentiate earlier phases from later stages**, where more is known
- For pre-post-change: Focus on **post approval changes**
- Shift focus from statistical inference to **experimental design aspects**
- **Let Applicant chose objective and comparison method, also showing OCs to ensure that patient risk is adequately controlled**

EMA Workshop

Some snapshots from 'Methods sessions'

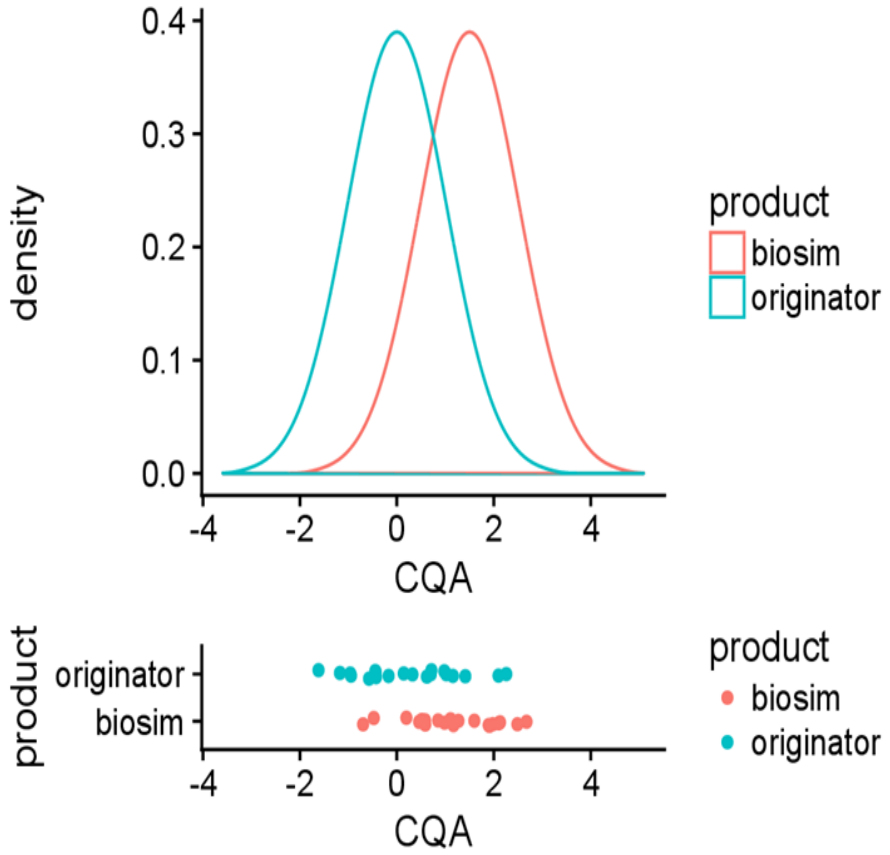
Klinglmüller et al.: Simulation study of OCs for 5 similarity criteria

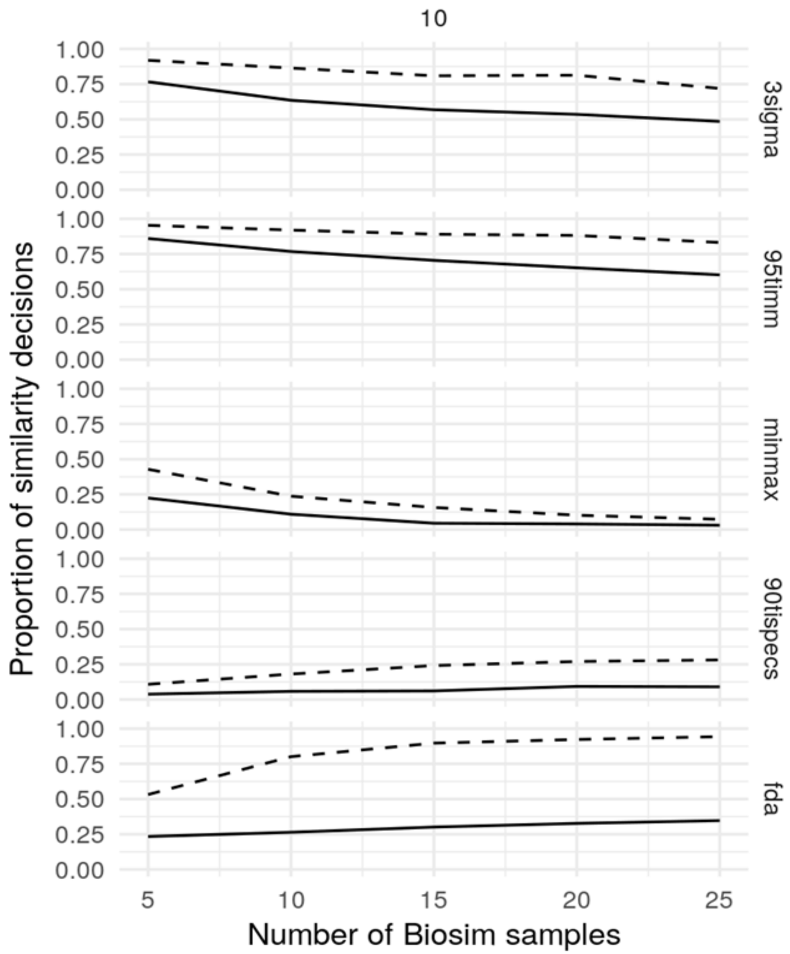


- Min-Max: All samples of the biosim are between min-max of the originator
- X-Sigma: All samples from the biosim are within x-standard deviations of the originators mean
- (P/Q) Tolerance interval: All samples from the biosim are within a P/Q Tolerance interval of the originator
- TI Specs: The P/Q tolerance interval of the biosim is within „specifications" (e.g. Min-Max) of the originator
- FDA Rule: The 90% confidence interval for mean difference between originator and biosim is within a similarity margin of 1.5 standard deviations of originator

EMA Workshop

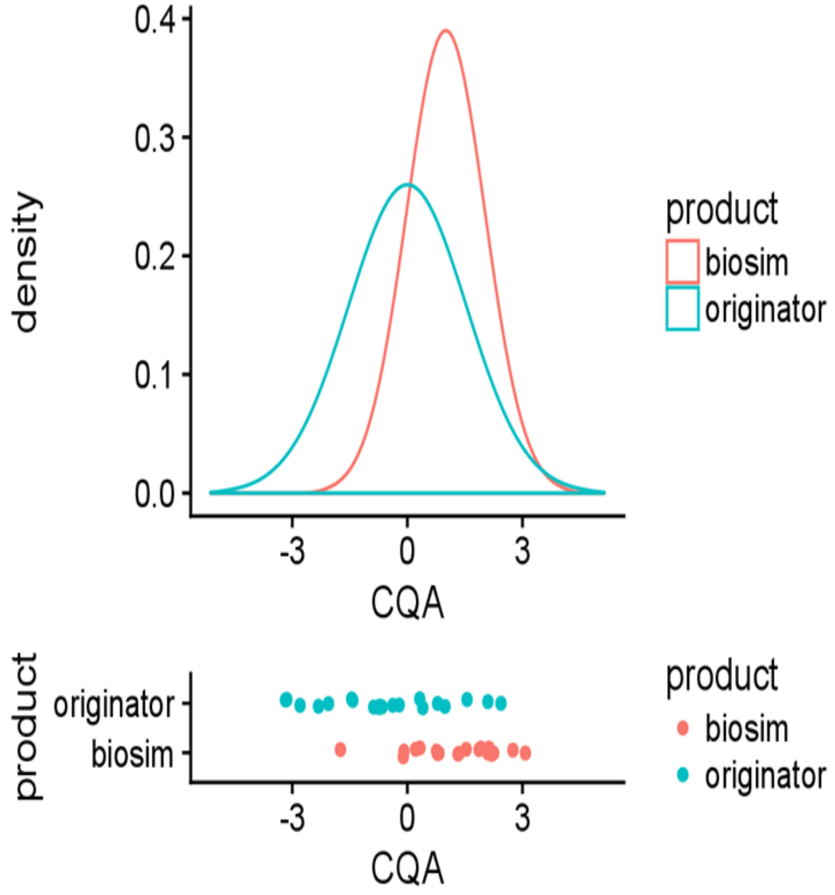
Some snapshots from ‘Methods sessions’

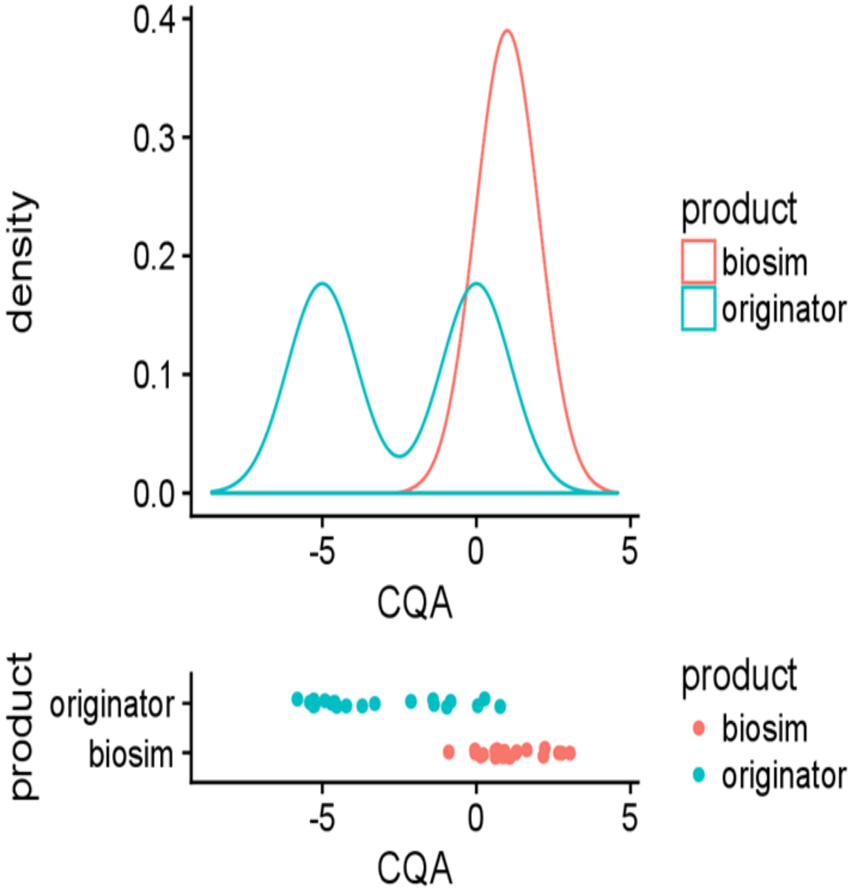




EMA Workshop

Some snapshots from ‘Methods sessions’

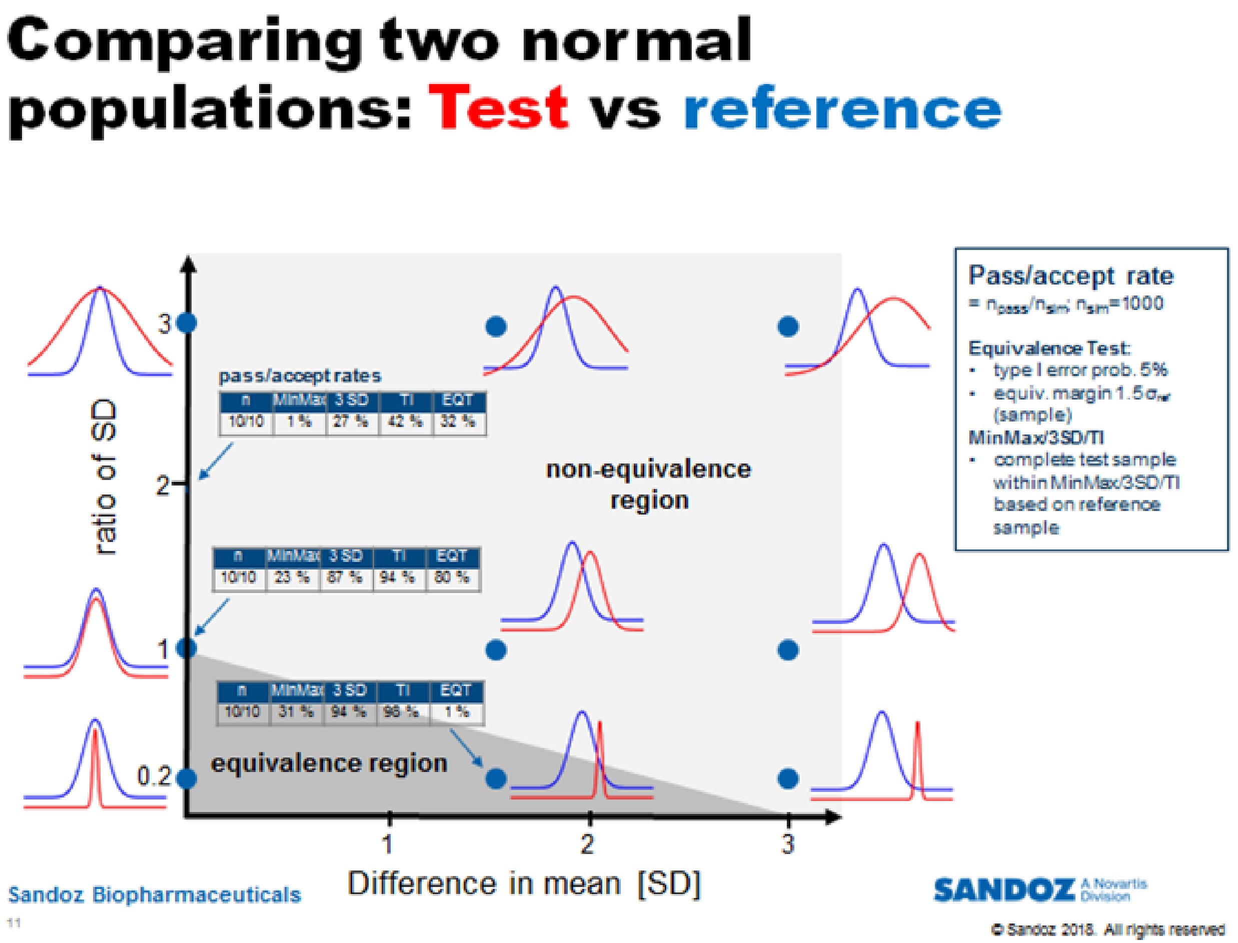




EMA Workshop

Some snapshots from ‘Methods sessions’

Stangler: Simulation study based on equivalence region



Comparing two normal populations: Test vs reference

pass/accept rates

n	MinMax	3 SD	TI	EQT
10/10	1 %	27 %	42 %	32 %

n	MinMax	3 SD	TI	EQT
10/10	23 %	87 %	94 %	80 %

n	MinMax	3 SD	TI	EQT
10/10	31 %	94 %	96 %	1 %

Pass/accept rate

= n_{pass}/n_{sim}; n_{sim}=1000

Equivalence Test:

- type I error prob. 5%
- equiv. margin 1.5 σ_{ref} (sample)

MinMax/3SD/TI

- complete test sample within MinMax/3SD/TI based on reference sample

Sandoz Biopharmaceuticals

11

SANDOZ A Novartis Division

© Sandoz 2018. All rights reserved

EMA Workshop

Some snapshots from ‘Methods sessions’

- Equivalence testing and quality range criteria aim at **different underlying similarity definitions**: equivalence of a parameter vs. population overlap
- Frameworks for exploring, comparing and visualizing OCs were shown, allowing for evaluation of the **dependency of OCs on different parameters**
- In many situations **tolerance intervals and k-SD ranges have undesirable properties**: power often decreases with increasing sample size, probability for a conclusion of similarity increases for shifts of reference away from test distribution
- Equivalence testing of means poorly performing to control a ‘population in a population’

EMA Workshop

Some snapshots from ‘Methods sessions’

- Incorporating knowledge about **sources of variation at the experimental design stage**: e.g. cyclical drifts in the production process → reduce variation, improve sensitivity and avoid potential biases of subsequent statistical analyses.
- ‘**Process based biosimilarity**’: Acceptance range is established on the basis of an interval estimated from the RMP. Range (or proportion thereof) of the test product’s QA distribution is estimated. Similarity concluded if the latter interval is included in the former.
  Interval estimates were proposed on the basis of frequentist and Bayesian tolerance and prediction intervals.
- Potential **advantages of Bayesian inference**: Additional assumptions about the distribution of underlying model parameters, eventually permitting conclusions about the probability of certain statements (e.g. similarity) being true, conditional on the observed data.
  This was considered to be closer to intuition than frequentist inference.

WP Learnings

Different contexts require separate considerations

- Arguments exist that **different settings** as mentioned in the scope of the RP (pre/post change, biosimilars, special small molecules) need **to be kept separated**
- Although there is no 'one-size-fits-all' solution, it might well be that in different comparison contexts **the same rigor for evidence** for similarity is desirable
- Need to **acknowledge differences** (in limitations/ options)
- Need to agree that many underlying **methodological issues remain the same**
- **Common principles** vs **required flexibility**

WP Learnings

Most likely approach to redefine the scope

From:

- pre/post change
- biosimilars
- ‘special cases’ small molecules

To:

From the regulatory perspective, it would be the **high impact of a false positive conclusion** on similarity which brings a certain data comparison of QAs **in scope**

WP Learnings

Importance to explain /agree upon "Statistical Inference"

If QA data comparison is done:
Interested 'exclusively' in the material/batches we have data for?

Example 1:

Batch release: comparison of actual batch QA data vs. specifications → interested only in the one batch to be released

Example 2:

Pre-/Post-change comparison: (batch) samples' QA data are compared → intention to 'claim' similarity beyond the samples!?

Example 3:

Biosimilar comparison: (batch) samples' QA data are compared → intention to 'claim' similarity beyond the samples!

WP Learnings

Importance to explain /agree upon "Statistical Inference"

<u>If interest goes beyond samples/batches we have data for:</u>

- For any similarity criterion applied: the **decision** drawn (similar/not similar) can be **right or wrong**;
- **False negative:** decision against 'similarity', although products/processes are sufficiently similar
- **False positive:** decision in favour of 'similarity', although products/processes are not sufficiently similar
- Adequate control of risk for false decisions desirable
- Properties of risk control → **Operating characteristics**

WP Learnings

Understanding Operating Characteristics is essential

- … because only then is it possible to quantify the risks in the regulatory decision making process;
- Best possible **knowledge of OCs** for the application of a specific similarity criterion is **key to justify its use** in the context at hand;
- Well understood **frameworks to visualize OCs** will be important to identify suitable similarity criteria;

WP Learnings

There is no unique optimal similarity criterion / statistical test

- For each specific similarity criterion, **OCs may vary considerably**, depending on the actual circumstances of data collection (i.e. number of batches, underlying data distributions, existence of shifts/drifts, differences in variability, etc.)
- It is generally **not meaningful to categorize similarity criteria** on a high level into ‘conservative’ and ‘liberal’, without context of the actual setting where the criterion is planned to be applied
- However, there is **knowledge re performance of frequently used criteria** in realistic data settings
- Equivalence testing of means suggested by FDA for tier 1 /biosimilarity: data settings /circumstances identified where approach not meaningful

Implications for giving guidance

Some suggestions reinforced / Working towards agreeable standards

- **Clarity about objective** of comparative QA data comparison:
  - Inferential approach or 'description only'?
  - What are the implications if similarity is shown at Q-level?
  - What are the risks if wrong decisions are made?
- **Think in advance:** what can be approached prospectively by making a (written) plan?
- What will be looked at? Batches? Is there a chance to **identify important sources of variability** in advance?
- Can sources of variability be **accounted for in sampling/analysis?**

Implications for giving guidance

Some suggestions reinforced / Working towards agreeable standards

- Write down planned/actual **sampling approach?** Why are some batches (not) chosen? → **transparency!**
- Is it possible to narrow down the **range of potentially important differences** in specific CQAs?
- Describe limitations
- In inferential setting: Choice of similarity criterion/statistical test based on (best possible) **knowledge about OCs**

What happens next?

RP revision, given EMAs Business Continuity Plan (BCP)

- Multidisciplinary drafting group formed after WS
(members from BSWP, BWP, BMWP, QWP)
- RP revision is not directly ‘product’-related → low priority during BCP phase 3
- Required time/support → no exemption for the RP
- Formally not on BSWP work-plan 2019
- Ways to continue currently explored

thomas.lang1@ages.at

Foto: Thomas Lang