Practical considerations in the statistical evaluation of biosimilarity — a laboratory perspective

EMA workshop on draft reflection paper on statistical methodology for the comparative assessment of quality attributes in drug development

May 3-4, 2018

Henriette Kuehne, Ph.D.

Coherus BIOSCIENCES

aaps®

Outline

Two types of practical difficulties associated with biosimilar comparability

I. Sources of **variability/uncertainty** that affect biosimilar comparability assessments

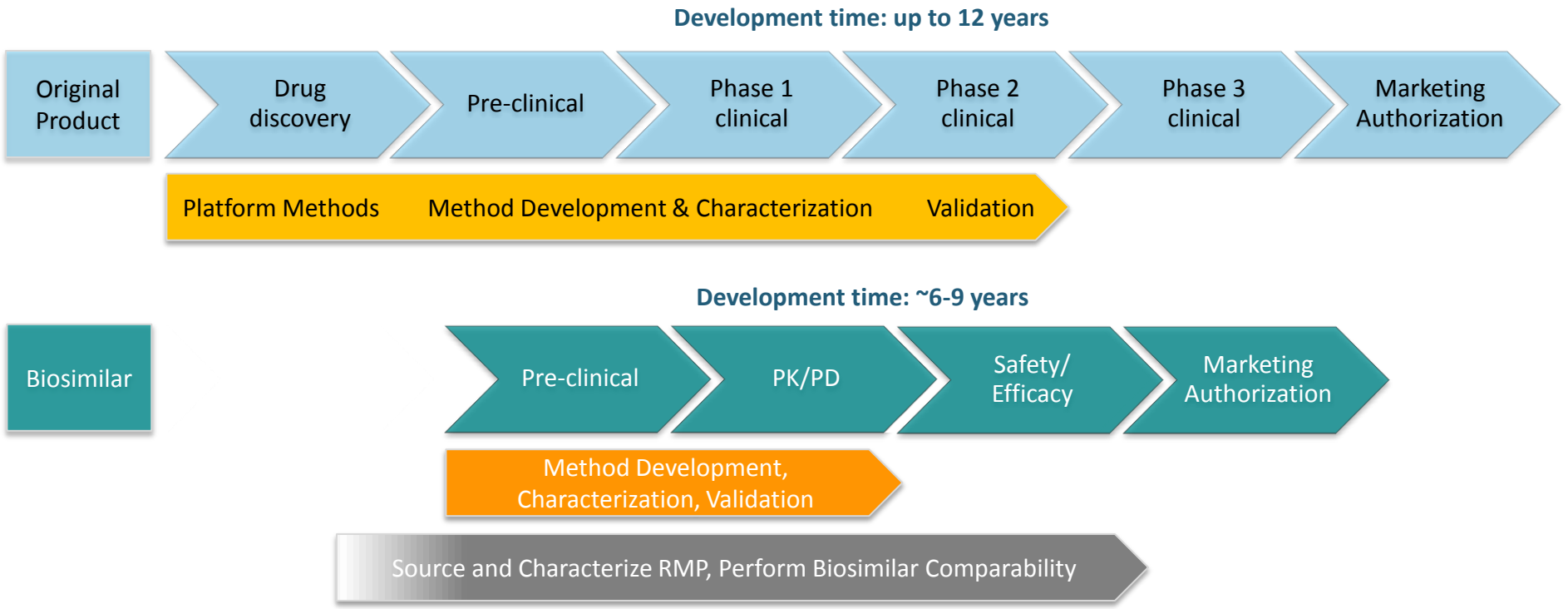
   A. Limitations in analytical experience or history due to abbreviated timelines
   B. Limitations in product knowledge for both reference medicinal product (RMP) and biosimilar product candidate (BP)

II. Anticipating and mitigating **conflicting outcomes** of mix-and-match statistical approaches

   - Incongruities in results for related QAs

> ***"Understanding the need and the options to quantify uncertainty related to decision making based on sample data is key to evaluate the capabilities statistical concepts may bring to the matter of comparing quality attributes."***
>
> – EMA draft reflection paper

PART I:
Abbreviated development timelines for biosimilars affect understanding of variability



- Estimate of uncertainty in a data set is based on historical knowledge of A) method and B) RMP
- Impact of variability depends on whether testing is conducted continuously or side-by-side

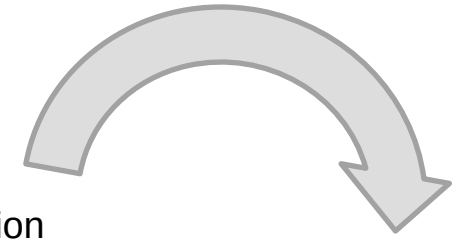
Sources of Variability A: Analytical Method Experience

Variability

a) Within sample
   - Lack of control of sample preparation
   - Sample storage (pre- and post-prep)
   - Use of different assays to measure the same (or related) QA (**see PART II**)

b) Within test
   - Measurement accuracy
   - Method precision

c) Between tests
   - Difference in test dates/intermediate precision
   - Method robustness
   - Differences in critical reagents



- Affects all products, but exacerbated in biosimilars:
  - Short historical method experience
  - Higher uncertainty re: long-term method performance

- Understanding of analytical method performance critical to
  - Similarity study design
  - Prospective definition of comparability acceptance criteria

Consequence: loss in power, increased risk to sponsor (unable to reject null hypothesis)

See also: EMA draft reflection paper, Section 5.2

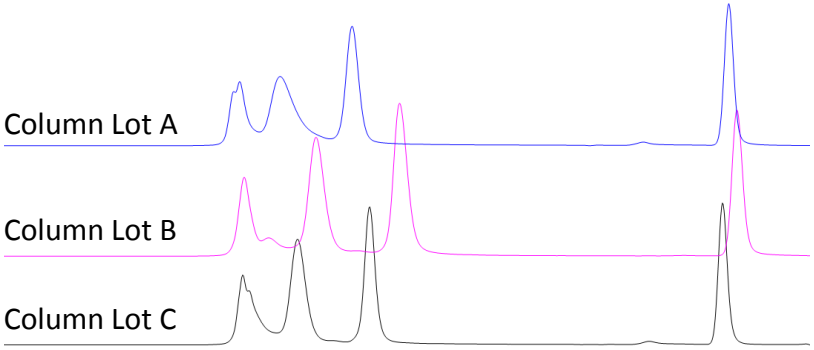
Unanticipated sources of variability: method robustness

Limited method experience can cause unforeseen issues
(can occur at any time, but may be more likely in biosimilars due to shortened development times)

- New column packing/resin lots
- New batches of reagents, esp. critical reagents (*e.g.* FBS, enzymes, cells)
- Age of reagents
- Source of water
- Drifts in “ambient” laboratory conditions
- Move to different HPLC (with different plumbing/dead volume, pumps, mixers, heaters, flow cell)
- Introduction of new analysts

→ Typically address prior to validation, but may not be fully explored in biosimilar timeline

Manufacturer calibration standard on 3 column resin lots



Unanticipated sources of variability can affect projected outcome of the study

- Overestimate variability (risk of false positive)
- Base variability estimate on known sources (risk of false negative)
- Test concurrently to minimize impact

Sources of Variability B: Product Knowledge

Variability

a) Between batch
   - Location of manufacturing
   - Scale of manufacture
   - Source of raw materials
   - Age of bulk at time of fill
   - Age of batch
   - Correlation of parent DS lots

b) Within batch
   - Time of day
   - Time since manufacturing start

Practical Difficulties

- Access to limited BP
  a) lots, manufacturing campaigns
  b) material per lot (especially if clinical)

  Note: process capability controlled by specifications

- Access to limited RMP
  a) lots
  b) material per lot

  Note: process capability/specifications unknown

  - Sourcing dictated by
    a) timing of project selection
    b) clinical and filing timeline
    c) budget
    d) market availability

See also: EMA draft reflection paper, Section 5.2

Practical limitations to understanding true extent of RMP variability

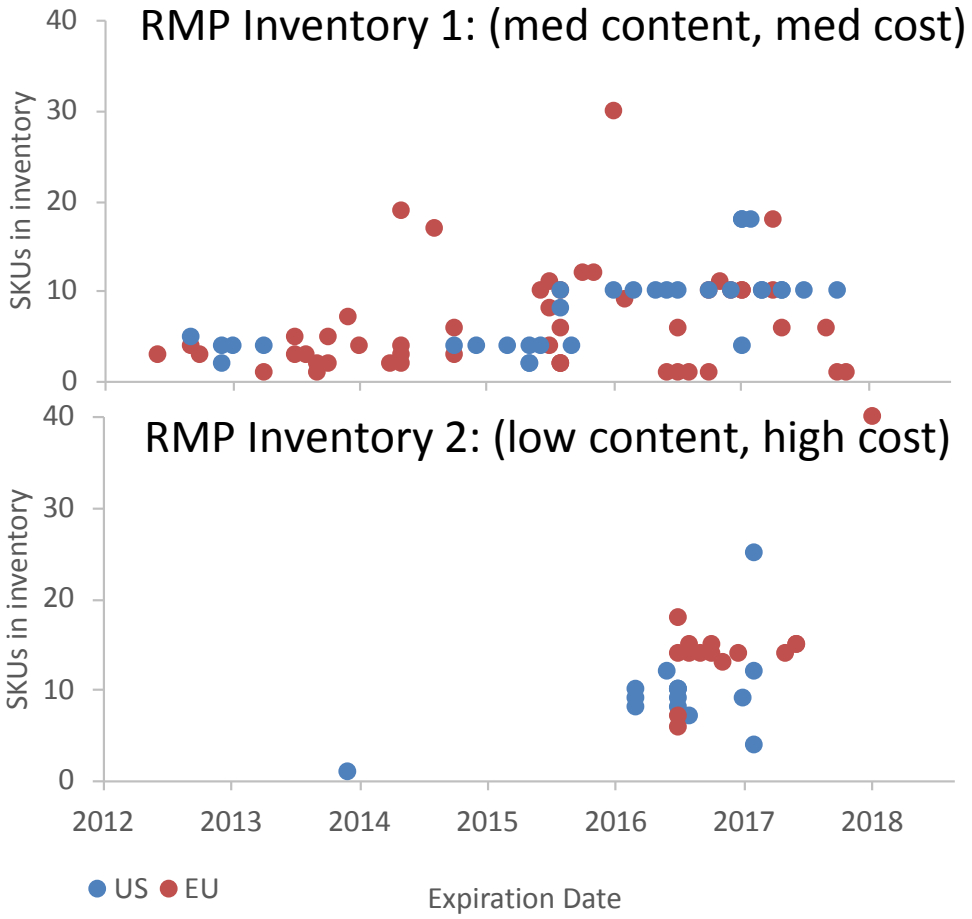
Assumption:

> ***"Whenever […] two products are compared, […] these products can be 'consistently' manufactured, guaranteed by adequate process-control measures."***
> – EMA draft reflection paper

Reality:

- → Limited set of lots
- → Potentially skewed distribution
- → May need to decide which lots to analyze for which attribute

Variability deemed acceptable in RMP is unknown and cannot always be estimated from analysis of available RMP lots



RMP variability vs. BP variability: batch selection for biosimilarity assessment

Batches specified in Advisory Committee Briefs (US FDA)	Number of lots included for MOA assays*		
	biosimilar	US originator	EU originator
Amgen ABP 501 (adalimumab)	10 (out of 10)	10 to 21 (out of 24)	10 to 17 (out of 18)
Amgen ABP 215 (bevacizumab)	13 (out of 19)	24 (out of 27)	27 (out of 29)
Celltrion CT-P13 (infliximab)	13 to 16 (out of 26)	16 to 27 (out of 45)	13 to 23 (out of 41)
Sandoz EP2006 (filgrastim)	15 to 17	10 to 15	34 to 52
Sandoz GP 2015 (etanercept)	8 to 9 (out of 19)	11 to 13 (out of 34)	11 to 12 (out of 50)

Source: FDA AdCom briefs

*Rationale/justification for inclusion of specific lots unknown

→What sampling strategy should be employed:

- Include all lots?
- Every clinical lot?
- First from from each expiration year?
- Everything within expiry?

→ How does sampling strategy impact:

- Outliers
- Manufacturing drift/shifts in RMP
- Manufacturing changes in biosimilar candidate
- Bridge clinical lots/different jurisdictions
- Include sub-commercial scale biosimilar lots?

Number of Reference Product Lots - To establish meaningful similarity acceptance criteria, sponsors should acquire a sufficient number of reference product lots. […] minimum of 10 reference product lots be sampled.

Number of Biosimilar Product Lots - To allow for meaningful comparisons, we recommend a minimum of 10 biosimilar lots be included in the analytical similarity assessment.

– US FDA draft guidance on biosimilarity, 2017

What factors are critical in batch selection to adequately model the RMP population and arrive at a valid comparison?

PART I Conclusion
Sources of Variability (A and B)

Timing of similarity study matters:

Similarity Approach	PROs	CONs	Outcome
Concurrent (or batched) side-by-side on all lots	• Minimize effects of method variability/drift	• Must wait for all lots to be produced/procured • Limited within expiry lots • Age mismatch BP vs RMP	Mitigate effects of method variability (Source A)
Continuous (at similar age/age ranges, or repeatedly throughout product lifetime)	• Minimize effects of lot age • Include more lots • Allows age extrapolation and adjustments	• Methods must be robust and finalized early • Logistically challenging	Mitigate effect of product variability (Source B)

→ Trade off: reducing uncertainty in one source of variability may increase uncertainty in another
→ Take customized approach depending on attribute: methods more prone to long-term variability (*e.g.* DSC, some HPLC) side-by-side; good intermediate precision (*e.g.* content, glycans) over time; stability indicating (*e.g.* degradants) repeatedly over time

PART II:
Higher purity/specific activity can adversely affect match in potency versus posology

EMA guidance stipulates that:

1. Posology must be the same,
2. Deviations in strength must be justified
(CHMP/437/04 Rev 1 (2014))

→ Expectation that certain critical QAs be matched (especially if desire to abbreviate clinical comparisons):
- Content (concentration, volume, label claim)
- Potency (biological activity)

"Particular attention should be given to quality attributes that might have a potential impact on safety or efficacy (e.g. impact on immunogenicity or potency)" – EMA guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: quality issues

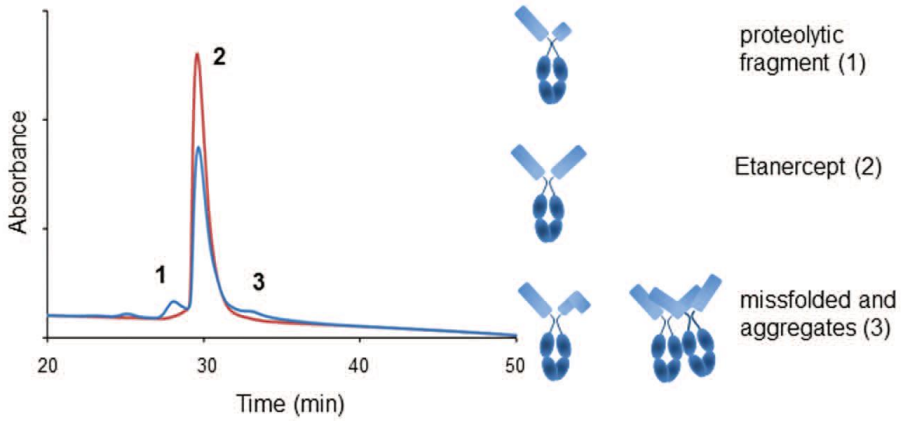
"In the biosimilar setting, any difference identified in any characteristic would need to be interpreted as a potential signal for non-similarity"

"From the general methodological point of view, the goal to demonstrate equivalence (in contrast to non-inferiority) is the focus in the biosimilar setting. […] exemptions could be potential improvements in specific QAs (e.g. impurities) which might translate to safety advantages" – EMA draft reflection paper

Current Paradigm	Content, potency and dose are critical QAs and must show high degree of similarity	Higher purity (lower immunogenicity) can be desirable
Current Approach	Match dose • Content (concentration, volume, label claim) • Potency (biological activity) • increased specific activity **(*two-sided acceptance limit*)**	Accept higher purity in the biosimilar candidate **(*one-sided acceptance limit*)**
Challenge	→ Can create conflicts between content, potency, dose	

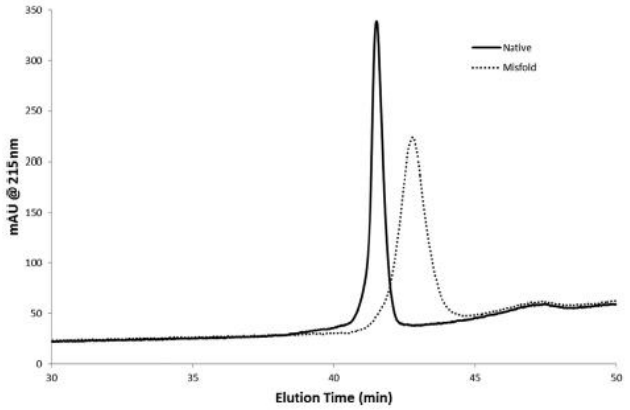
Case study: etanercept (Enbrel)

Enbrel purity by HIC; marketed Enbrel contains ~17% misfold



Source: Haverick et al., mAbs 6 (4), pp. 852–858 (2014)

Current manufacturing technology allows for effective removal of misfold, generating highly pure etanercept

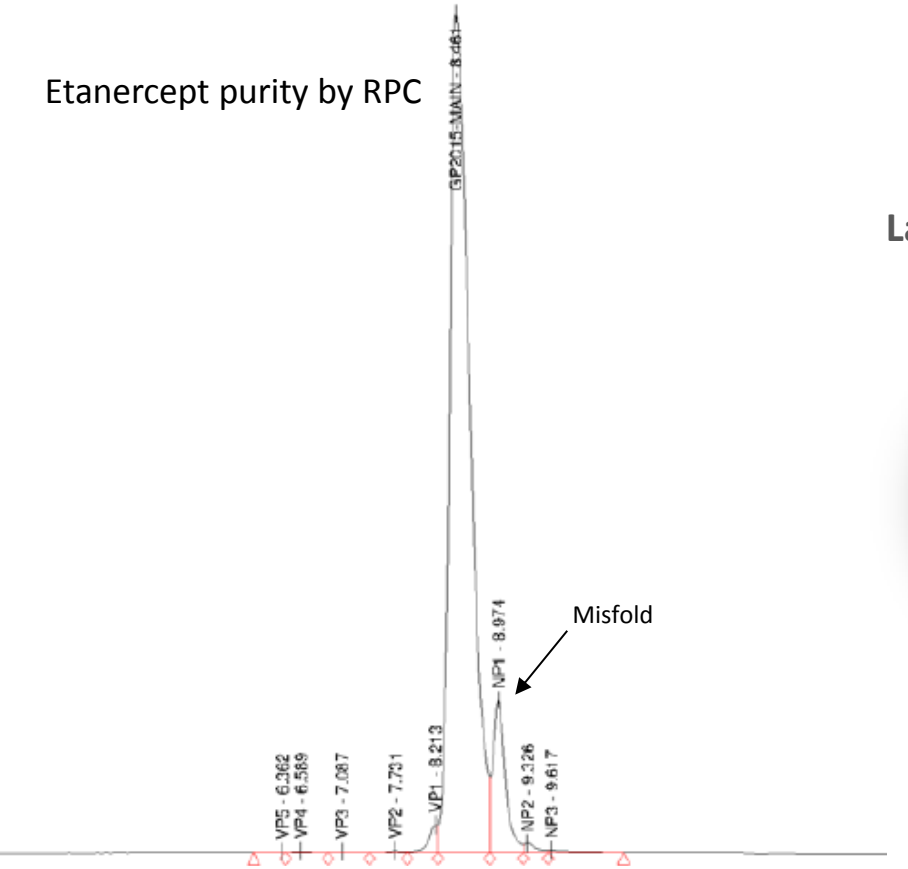


Source: Arakawa et al., Protein Expression and Purification 116, pp. 144-151 (2015)

If ~17% of Enbrel is misfold, at least some of which is inactive, how does one match both protein content and potency per dose?

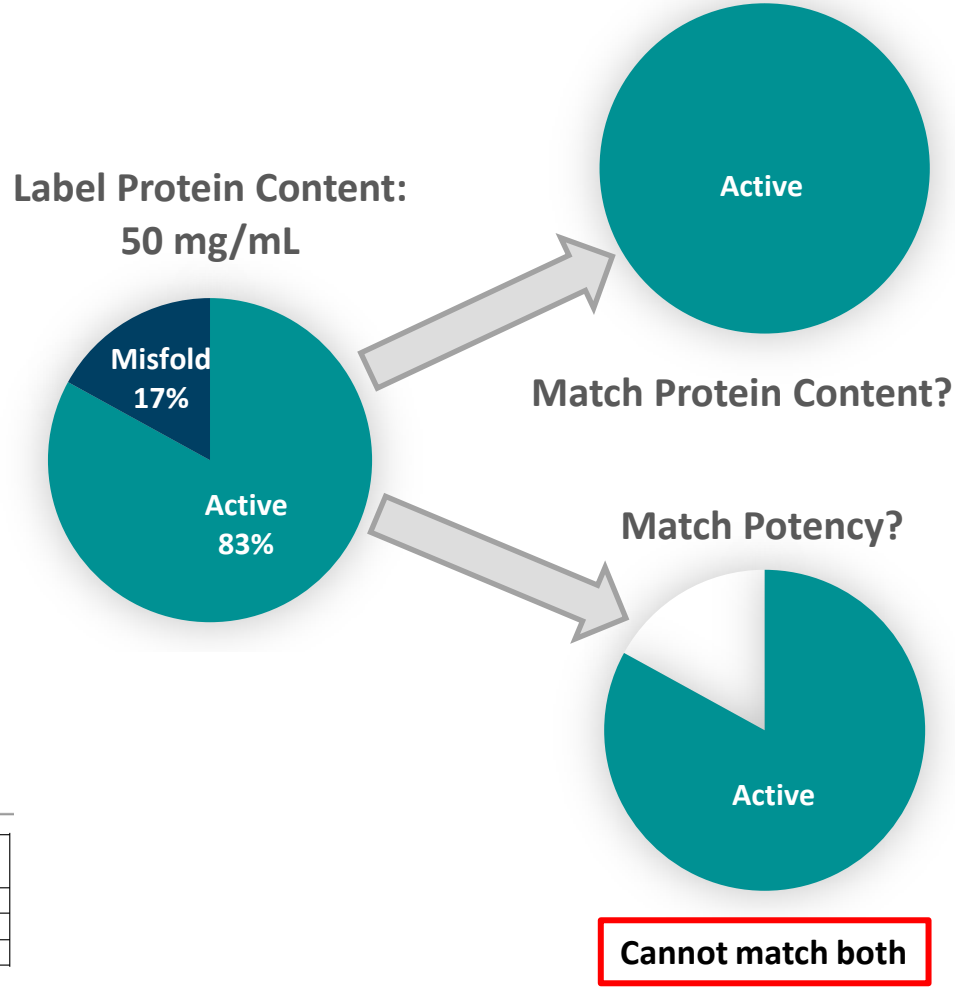
Match protein content or potency?

Etanercept purity by RPC



Product	# of lots	Sample mean, %	Sample standard deviation, %	Min, %	Max, %
GP2015	19	10.73	0.62	9.6	11.8
US-licensed Enbrel	21	16.16	1.91	10.2	17.4
EU-approved Enbrel	26	17.54	2.01	12.3	19.8

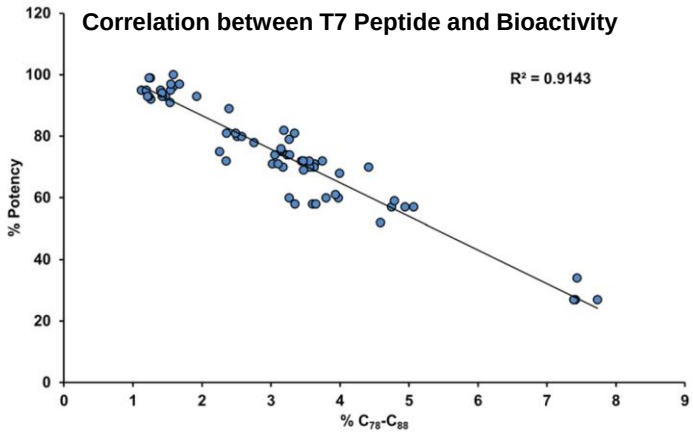
Source: FDA brief on BLA 761042, Sandoz biosimilar to Enbrel (etanercept-szzs)



Sandoz Erelzi (US vs EU)

US: content passes, potency fails equivalence

→ developed computed potency model



Source: Source: FDA brief on BLA 761042, Sandoz biosimilar to Enbrel (etanercept-szzs); Lamanna et al., Nature Scientific Reports, volume 7, Article number 3951

→ Use T7 as accurate surrogate for inactive content (1% T7 corresponds to loss of 10% potency) and adjust potency

→ Note that misfolds revert to active under physiological conditions

Adjusted potency values pass equivalence criteria

EU: content, potency pass

Attribute	Method	Key Findings
Content	UV/Vis	Equivalent
TNF-a neutralization	reporter gene	Comparable potency
Hydrophobic Variants	RPC	Lower post-peak variants in Erelzi

"Submitted data suggest that these misfolded variants may refold to the active variant correctly."

Source: Erelzi assessment report, EMEA/H/C/004192/0000

Potential Concerns:

- Complexity of approach increases uncertainty (add error in fit to noise of bioassay)
- Approach highly product-specific, relies partly on other inactive species reverting to active in concert with T7
- May not work for other products

Post administration potency is now matched, but total potency in the administered dose is not

Remaining Questions

> ***"[…] objective for each specific QA's comparison: e.g. if means are compared, is it sufficient to rule out marked differences in one direction only (e.g. rule out increase in impurity, or decrease in potency), or is it the goal to protect against differences in either direction?***
>
> ***For example, in one and the same pre-/post-manufacturing change comparison, it may well be that that e.g. a reduction in mean post-change impurity could be acceptable (one-sided comparison), whereas for other QAs (e.g. potency) marked differences in pre-/post-change means in either direction need to be excluded (two-sided comparison), as such differences - depending on the direction - might relate to expected negative impact either on clinical efficacy or on safety."***
> – EMA draft reflection paper

WHAT IF

- Unidentified inactive misfolds didn't track with surrogate (offset in potency)?
- Reversion to active form/absence of clinical relevance could not successfully be shown?
- Lack of impurity decreases immunogenic risk?

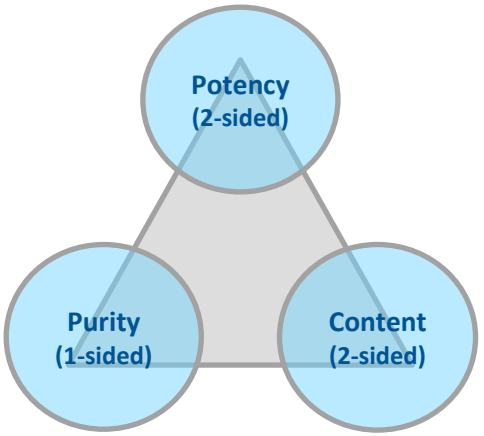
→ How do we allow higher purity even if it corresponds with increased specific activity?

PART II Conclusion
Alternative Approaches

Goal

Clear guidance on how to prioritize purity (non-stability indicating) versus dose

> ***"it is expected that the comparison will involve more than one QA. […] all the methodological considerations […] need to be applied separately for each QA selected for the comparison task. [it] is unreasonable to assume that one and the same statistical concept will be suitable for comparative evaluation of all the QAs"*** – EMA draft reflection paper



Premise

Must match potency per dose to meet bioequivalence

Possible Approaches	Caveats
Include impurities in BP to match both content and potency	May affect safety/immunogenicity
Prioritize certain QAs over others	Which QAs? With what approaches?
Allow decreased BP protein content based on purity ratios (lower concentration or fill volume)	How to match label claim?
Report protein content based on comparison to RMP active content (e.g. by HPLC)	How to define concentration?
Allow label concentration to be based on matched potency (active concentration) by using empirical "extinction coefficient" adjusted for RMP purity	Requires new paradigm

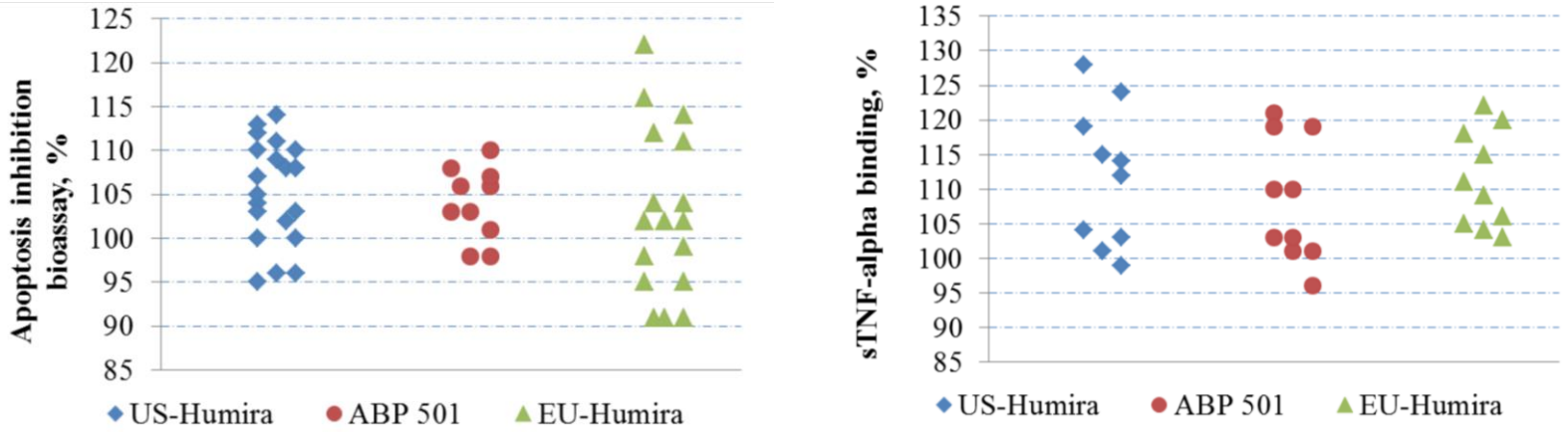
Acknowledgments

AAPS Biosimilar Focus Group and CMC subcommittee members
Rick Burdick, PhD

Karen Miller, PhD
Janice Davis, PhD
LeeAnne Merewether
Matt McQueen
Alan Carr

Backup slides

Known sources of variability: method precision



Source: FDA brief BLA 761024, Amgen biosimilar to Humira (adalimumab-atto)

- Wide range of results (cell-based: 90 – 120%, ELISA: 95 – 130%)
- Variance may be due to assay variability or true lot variability (raw data needed to distinguish)
- How small a difference can be detected?

→ For assays that are inherently variable need to keep variability sufficiently low to detect meaningful differences within the noise and arrive at a reasonable equivalence margin

- Ensure assay fully optimized (e.g. dilution step)
- Include sufficiently sample set
- Test additional replicates of single samples
- Test side by side in single assay

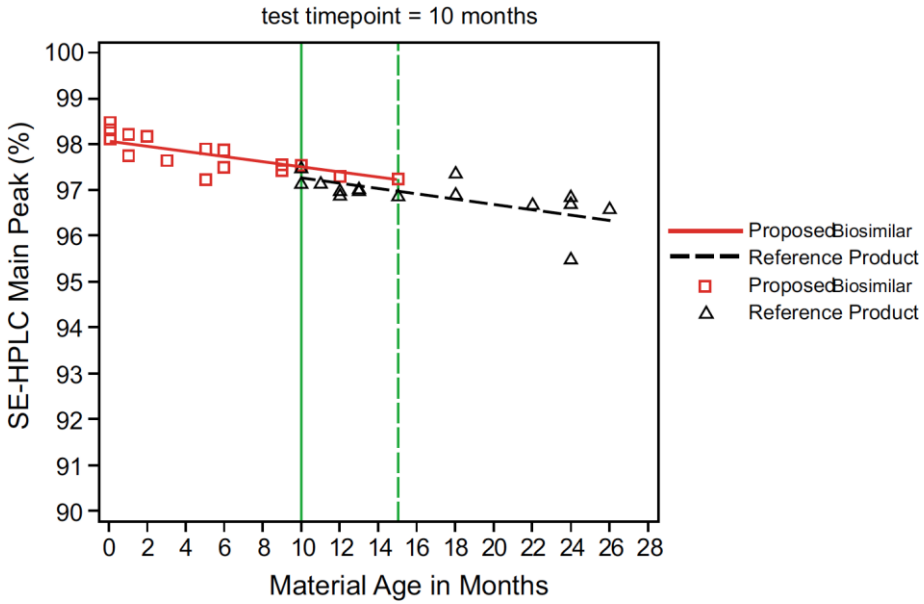
Keep variability sufficiently low to detect meaningful differences; due to time constraints can more difficult to validate methods with high precision

Stability indicating attributes (age of batch): a moving target

Awareness of lot age at the time of testing and adjustment for age:

- RMP sourced ~6 to 12 M into its lifetime
- Biosimilar tested from time zero

→ Lower impurities in biosimilar



Source: Markus et al., BioDrugs 31:175–187 (2017)

Caveats

- Initially limited understanding of assay consistency and robustness
- RMP ages or expires prior to method finalization/optimization
- QA requiring equivalence test may be stability indicating *(e.g.* MOA potency)

→How to design and **time** the similarity assessment to ensure materials are matched in age and prevent false conclusion of equivalence (*e.g.* of young BP to aged RMP)?

→Consider impact of uncertainty in extrapolation/time correction on overall uncertainty

Effects of lot age may increase overall uncertainty in equivalence assessments