

PSI Scientific Committee Webinar:
Statistical Challenges in Analytical Comparability and Biosimilarity
Assessment

Setting the Scene:
The regulatory landscape – a journey through time

Thomas Lang

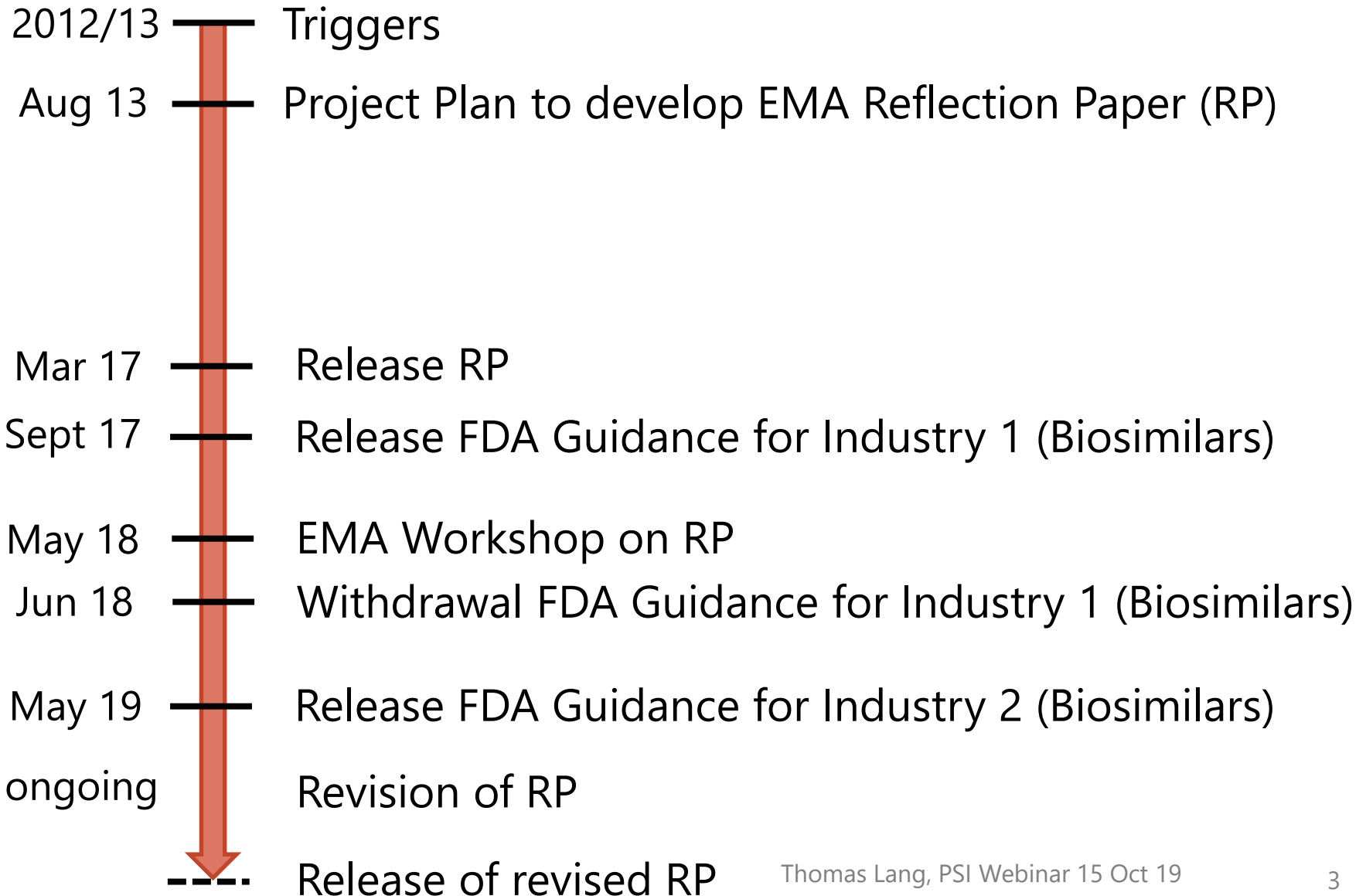
AGES Austrian Agency for Health & Food Safety

Disclaimer

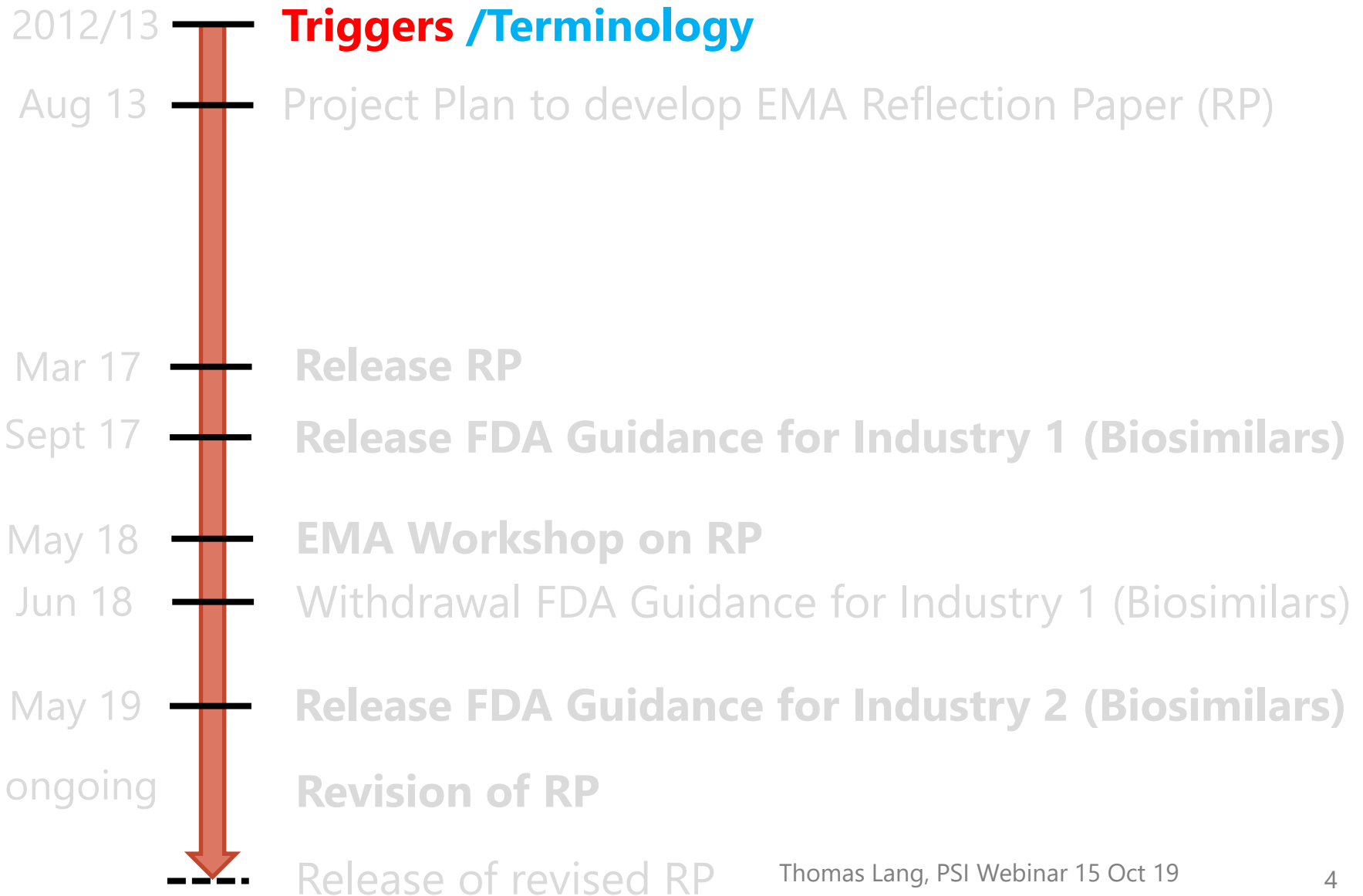
Content of this presentation reflects personal opinion in the role as Rapporteur for the EMA Reflection Paper.

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

... as time goes by



Content of talk



What is a ... ?

'Quality Attribute' (QA): any kind of physico-chemical characteristic, biological/activity characteristic, immuno-chemical property, purity/impurity characteristic, or any other in-vitro characteristic, which can be measured for drug substance/product."

e.g.: active ingredient content, protein content, binding capacity etc.

'Critical Quality Attributes' (CQAs)

- usually taken for manufacturing process control/monitoring
- strong association to clinical outcome is well known (e.g. potency), or
- association with clinical response is/remains unclear, but might be of relevance (e.g. glycosylation variants)

What is a ... ?

<https://www.ema.europa.eu/en/human-regulatory/overview/biosimilar-medicines-overview>

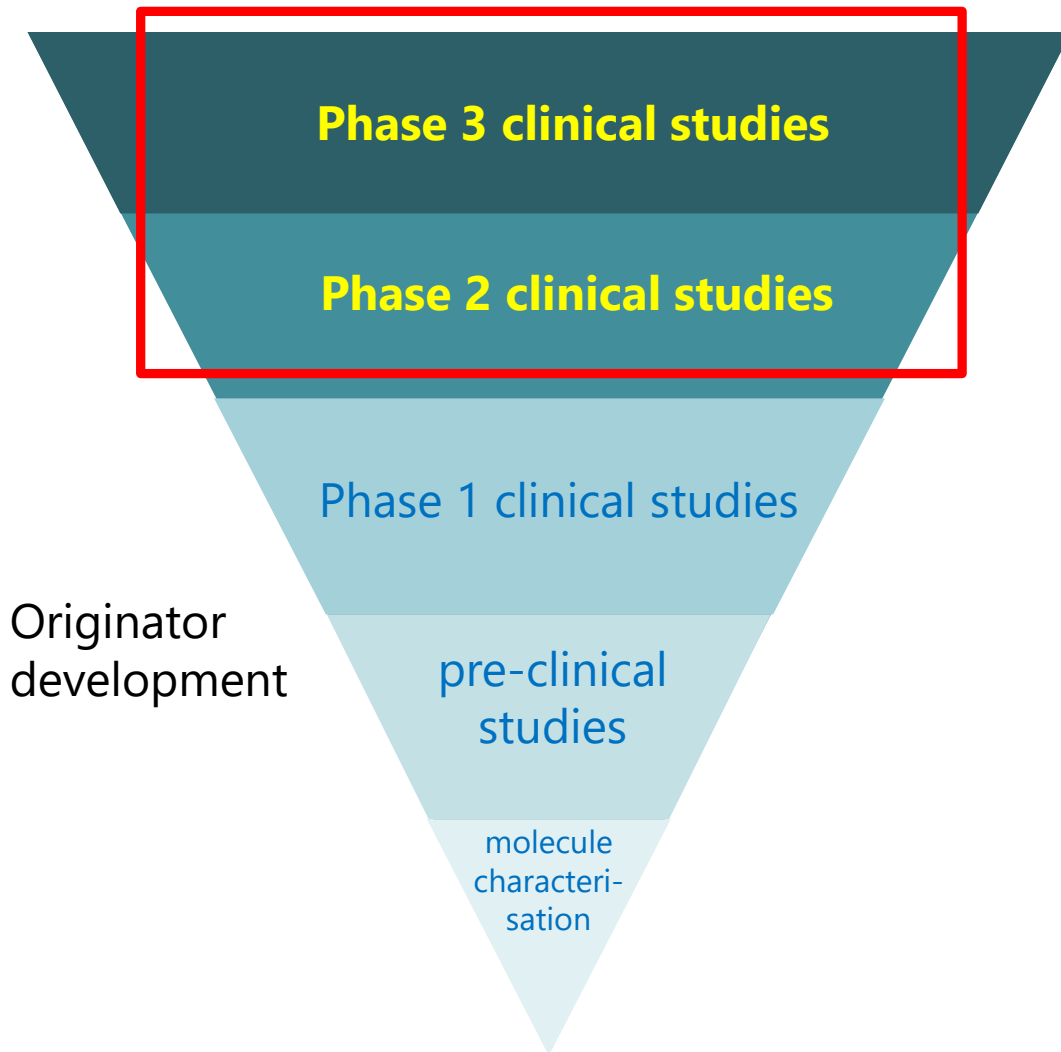
A **biosimilar** is a biological medicine highly similar to another biological medicine already approved in the EU (called 'reference medicine') in terms of structure, biological activity and efficacy, safety and immunogenicity profile;

A biosimilar is not regarded as a generic of a biological medicine. This is mostly because the natural variability and more complex manufacturing of biological medicines do not allow an exact replication of the molecular micro-heterogeneity;

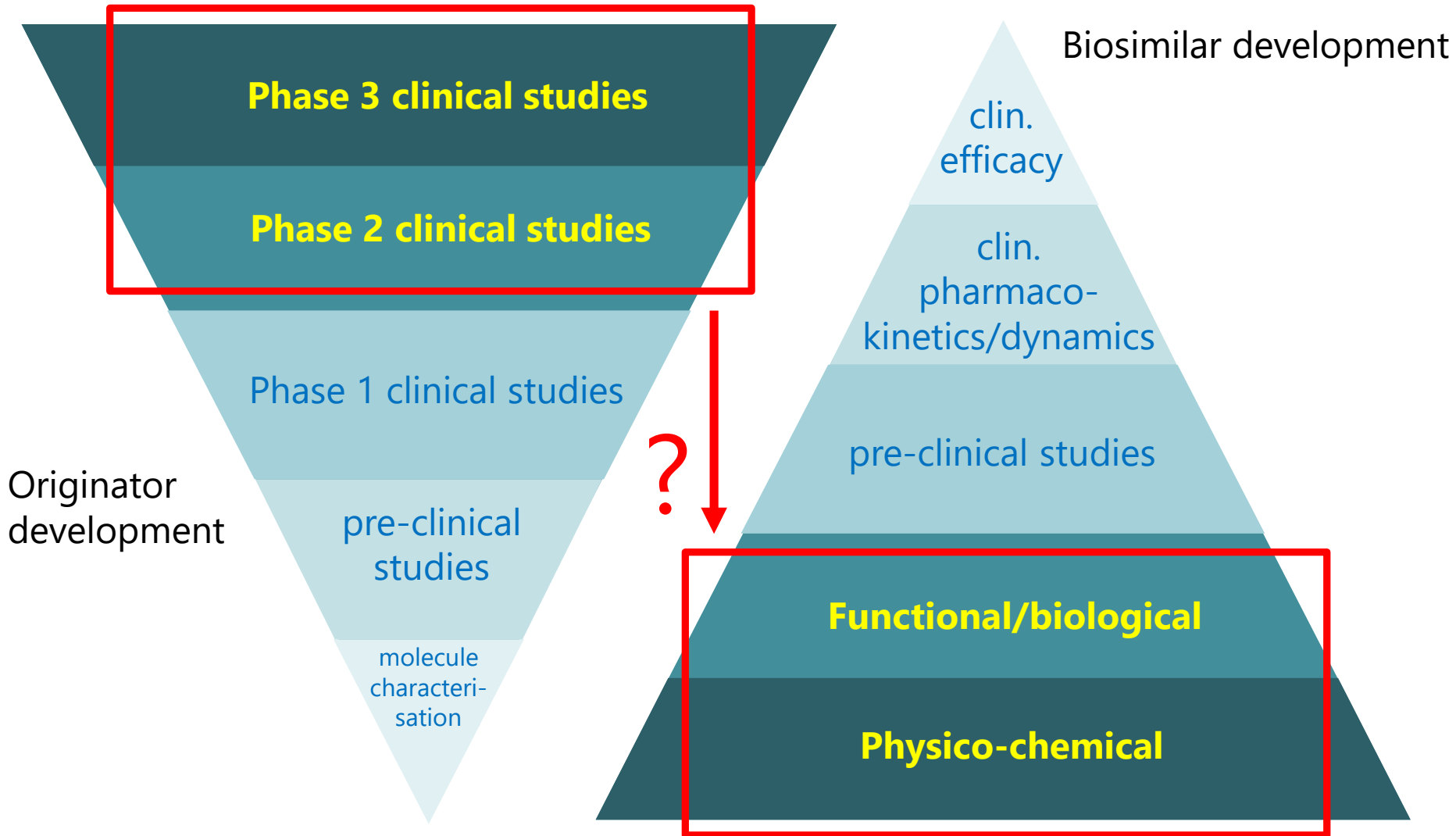
The main reasons for the initiative

- “If we want to develop a biosimilar without clinical trials evidence, what would regulators request?”

Pyramid of evidence



Pyramid of evidence is turned upside down



The main reasons for the initiative

- “If we want to develop a biosimilar without clinical trials evidence, what would regulators request?”
- Statistical thinking? No simple transfer from ‘clinics’ to ‘quality’
- Huge diversity in similarity criteria

Frequently seen similarity criteria

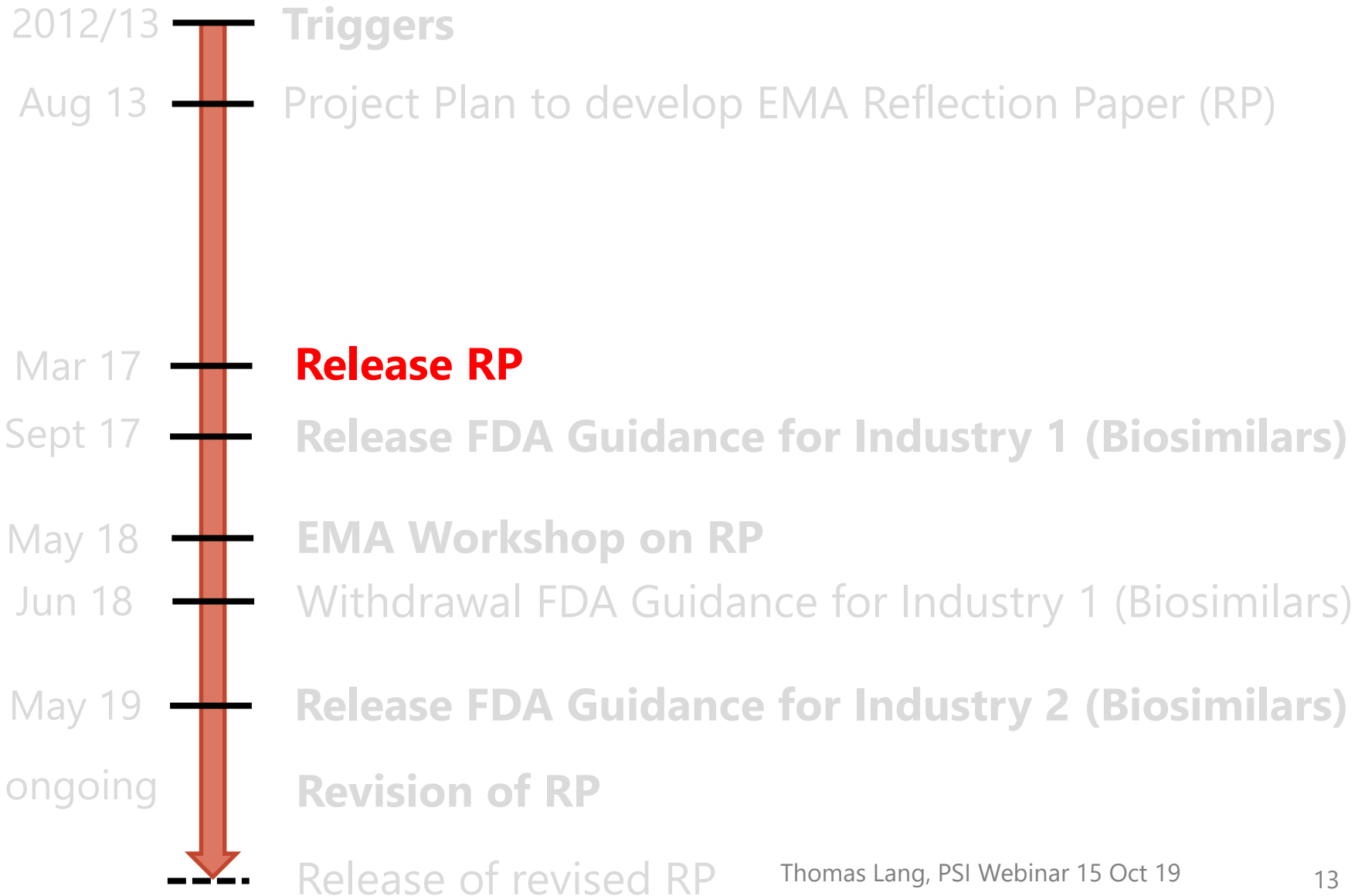
What to compare: distribution parameter estimates? ranges?

- **Min-Max:** All samples of the biosimilar are within the min-max-range of the originator
- **X-Sigma:** All samples from the biosimilar are within $\pm x$ -standard deviations from originator's mean
- **"FDA Tier-1":** The 90% confidence interval for difference in means is within ± 1.5 standard deviations of originator
- **(P/Q) Tolerance interval:** All samples from the biosimilar are within a P/Q Tolerance interval of the originator
- **"graphical"** ~ heuristic approaches
- ...

The main reasons for the initiative

- “If we want to develop a biosimilar without clinical trials evidence, what would regulators request?”
- Statistical thinking? No simple transfer from ‘clinics’ to ‘quality’
- Huge diversity in similarity criteria
- As regulatory statistician →
responsibility to flag flawed QA data comparison approaches,
but often no straight forward alternative
- Differences in problem understanding and language
- Many regulatory settings potentially benefitting
 - pre-post manufacturing change
 - generics: dissolution / in-vitro comparison

Content of talk



Similar issues in various areas

All under one roof?

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

Reflections regarding framework

No single method promoted, some criticised

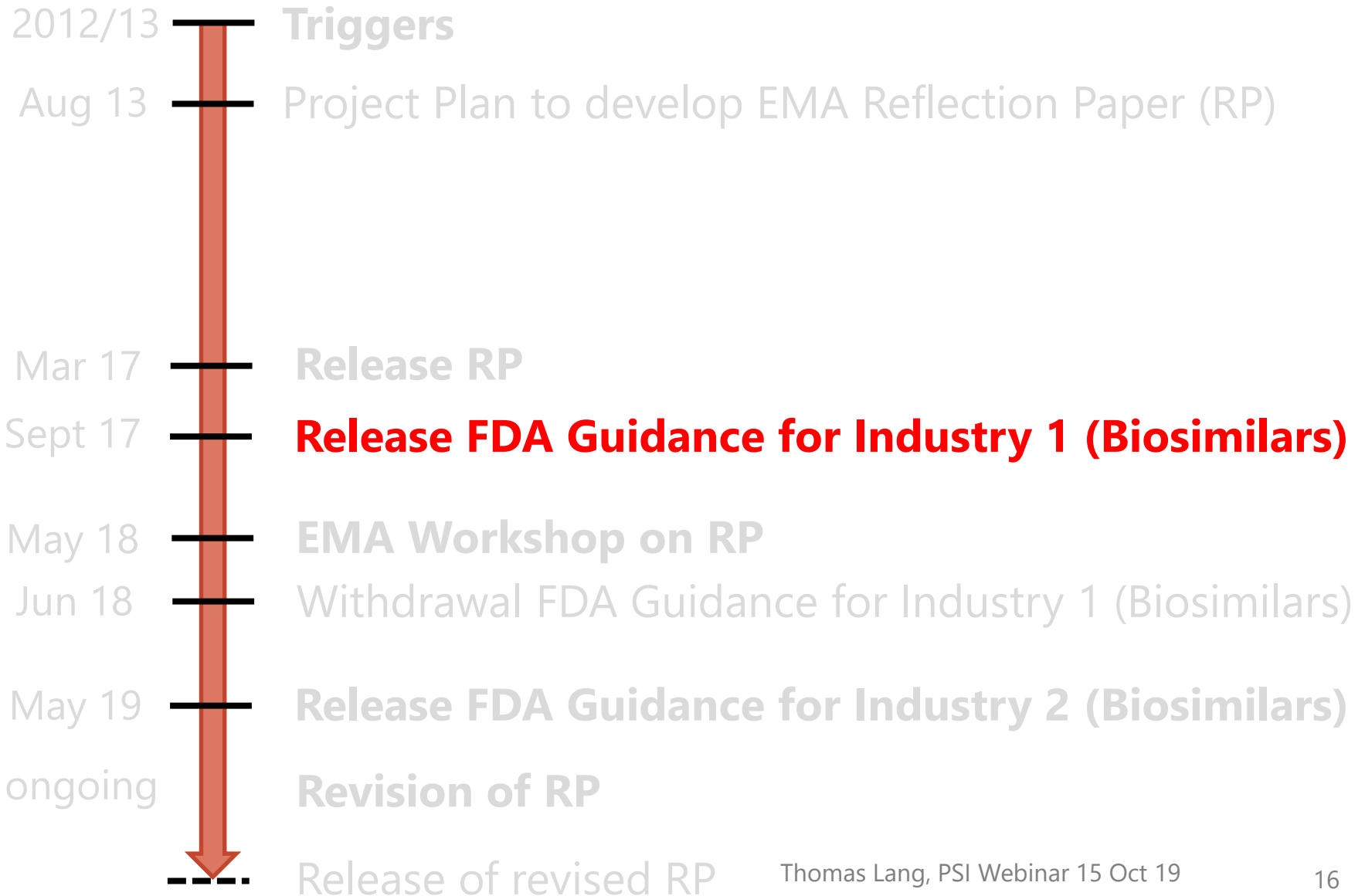
Section 4: Description of settings (according scope)

Section 5: Important aspects enabling statistical inference:

- 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

Content of talk



Statistical Approaches to Evaluate Analytical Similarity Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

3-tiered approach suggested

Tier 1: Equivalence testing of means

A test of the equivalence hypothesis can be conducted by requiring the simultaneous rejection of the following two one-sided null hypotheses:

$$H_{01} : \mu_T - \mu_R \leq -\delta \text{ vs. } H_{a1} : \mu_T - \mu_R > -\delta$$

$$H_{02} : \mu_T - \mu_R \geq \delta \text{ vs. } H_{a2} : \mu_T - \mu_R < \delta$$

With $\delta = 1.5 \sigma_R$, the test generally should support equivalence if the 90% confidence interval of the difference in means lies within the interval $(-1.5 \sigma_R, 1.5 \sigma_R)$ (i.e., the lower limit of the 90% confidence interval for the difference in means is greater than $-1.5 \sigma_R$ and the upper limit is less than $1.5 \sigma_R$). Use of this multiplier in computing the equivalence margin results in a test with reasonable properties under what we feel are realistic conditions.

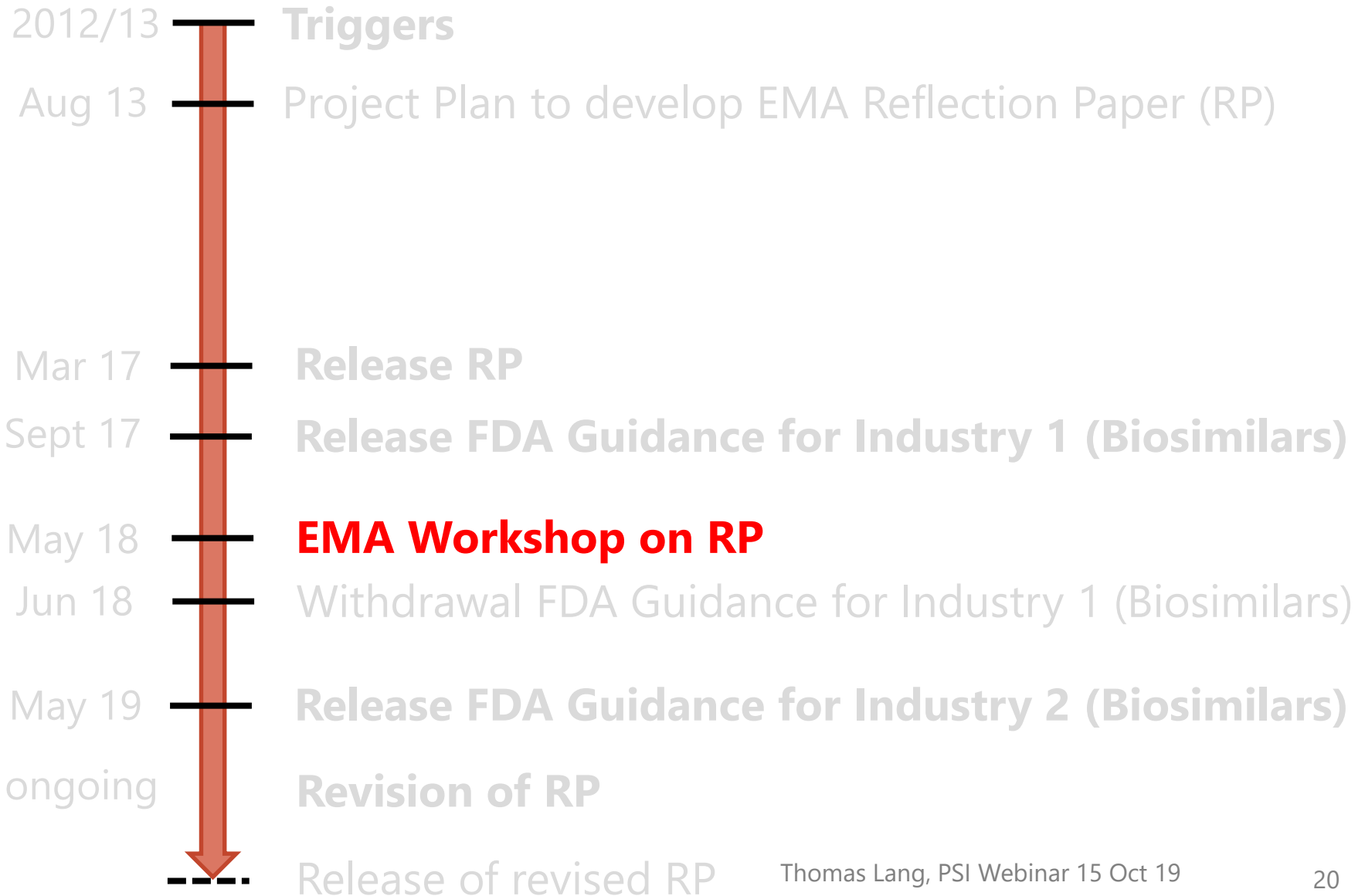
3-tiered approach suggested

Tier 2: Quality range approach (Tier 3: Visual displays)

For Tier 2, the similarity acceptance criteria based on reference product results for a specific quality attribute should be defined as $(\hat{\mu}_R - X\hat{\sigma}_R, \hat{\mu}_R + X\hat{\sigma}_R)$, where $\hat{\mu}_R$ is the sample mean and $\hat{\sigma}_R$ is the sample standard deviation based on the reference product lots. The multiplier (X) should be scientifically justified for that attribute and discussed with the Agency.

Analytical similarity generally should be demonstrated for a quality attribute if a sufficient percentage of test lot values (e.g., 90%) fall within the quality range defined above for that attribute. The lots used for Tier 2 testing should, if possible, be the same as those used for Tier 1 testing.

Content of talk



Platform for new ideas

From:

“If we want to develop a biosimilar without clinical trials evidence,
what would regulators request?””

To:

“If we want to develop a biosimilar without clinical trials evidence,
what would developers suggest?””

One year Public Consultation Phase

Many comments received

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

- Concerns/Reservations
- Conflicts/Shortcomings
- Proposals

Workshop sessions

- **Problem statements** and challenges
- Case studies with focus on **pre-post manufacturing changes**
- Case studies with focus on **Biosimilars**
- **Operating characteristics** of currently/frequently used similarity criteria
- **New Strategies** and alternative methodological approaches

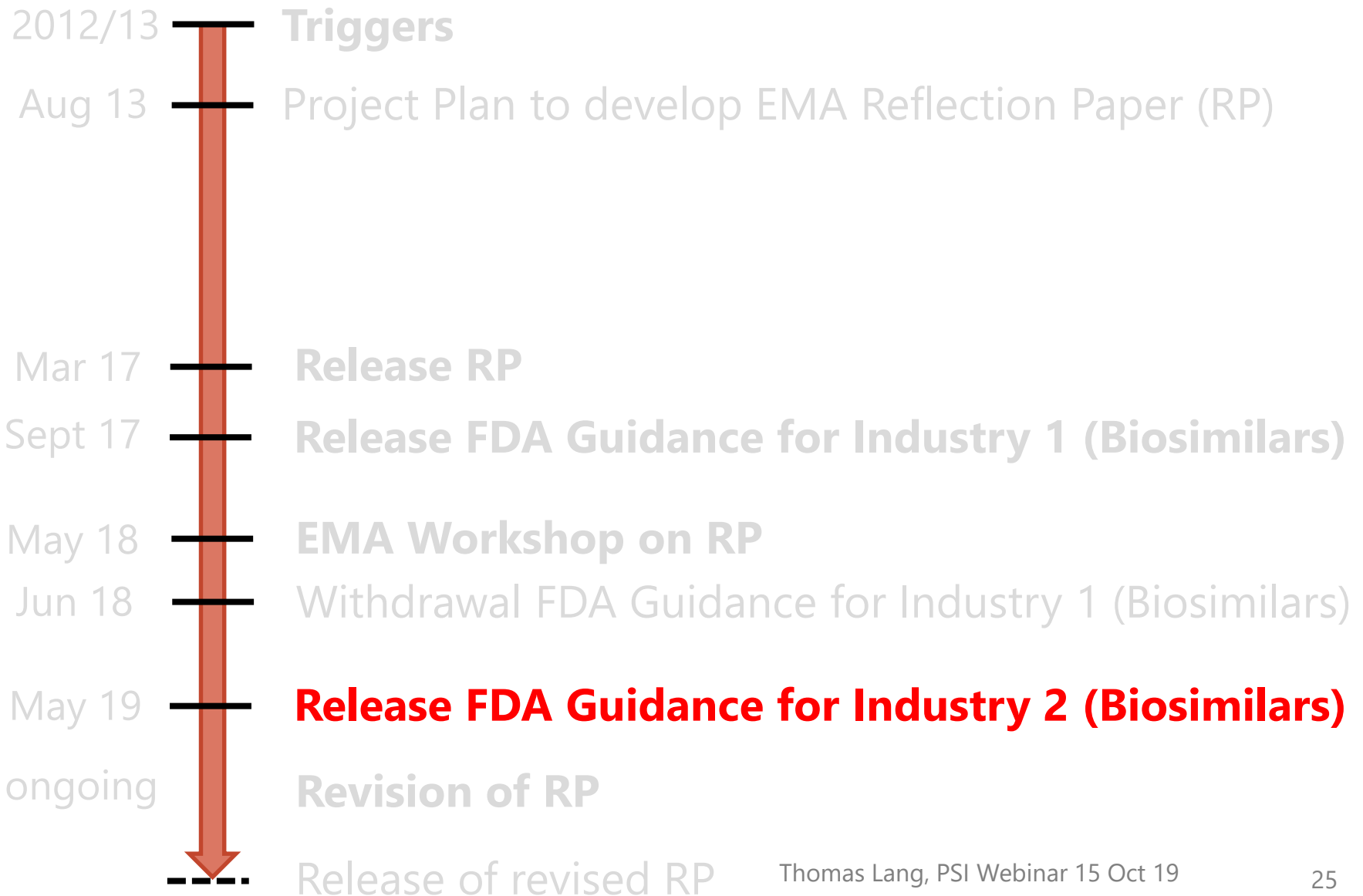
WS-Presentations available on:

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

Learnings and open question

- No 'one-size-fits-all'
- Association QA \leftrightarrow clinical outcome rarely sufficiently clear ...
- Importance to account for shifts and drifts in QAs over time
- 'Non-parametric' way of thinking: 'population within population'
- Bayesian approaches
- How good is a similarity criterion: operating characteristics

Content of talk



**Development of Therapeutic Protein
Biosimilars: Comparative Analytical
Assessment and Other Quality-Related
Considerations**

Guidance for Industry

DRAFT GUIDANCE

Parametric way of thinking

“ ... the Agency recommends that sponsors develop the manufacturing process to **target the centers of distribution** of the quality attributes of the reference product as closely as possible.”

“The objective of the comparative analytical assessment is to verify that each attribute, as observed in the proposed biosimilar and the reference product, has a **similar population mean and similar population standard deviation.**”

“ ... the QR, which **assumes that the population mean and standard deviation are similar,** is an appropriate approach to demonstrate that the proposed product is highly similar to the reference product.”

Comeback of Tier-2 approach

One approach to data analysis would be the use of descriptive **quality ranges for assessing quantitative quality attributes** of high and moderate risk, and the use of raw data/graphical comparisons for quality attributes with the lowest risk ranking or for those quality attributes that cannot be quantitatively measured (e.g., primary sequence). The acceptance criteria for the quality ranges (QR) method in the comparative analytical assessment should be based on the results of the sponsor's own analysis of the reference product for a specific quality attribute. **The QR should be defined as $(\hat{\mu}_R - X \hat{\sigma}_R, \hat{\mu}_R + X \hat{\sigma}_R)$** , where $\hat{\mu}_R$ is the sample mean, and $\hat{\sigma}_R$ is the sample standard deviation based on the reference product lots. **The multiplier (X) should be scientifically justified for that attribute and discussed with the Agency.**

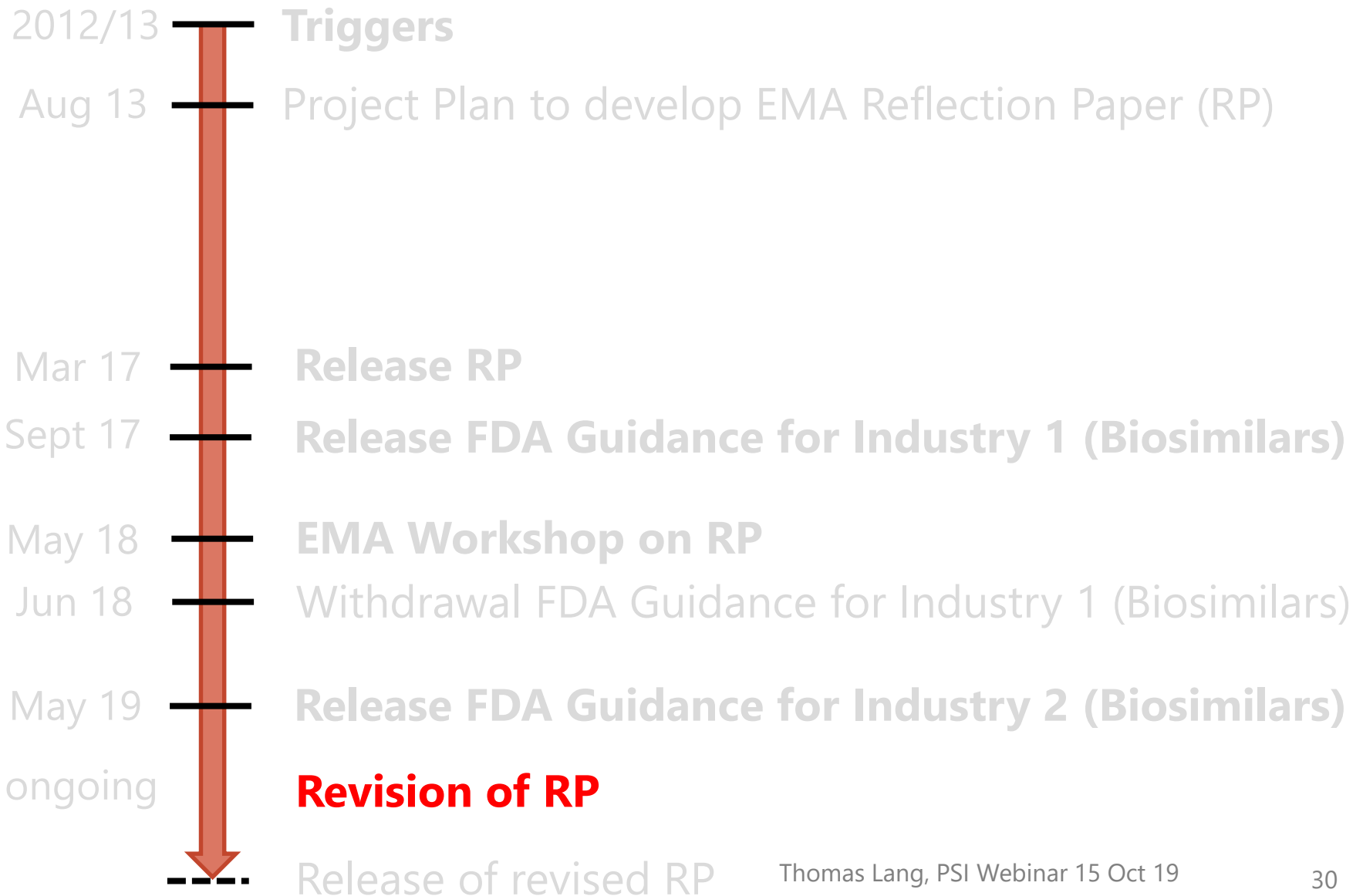
"... Comparative analysis of a quality attribute would generally support a finding that the proposed product is highly similar to the reference product when **a sufficient percentage of biosimilar lot values (e.g., 90%)** fall within the QR defined for that attribute."

Prospective planning is recommended

“ ... The Agency recommends development of a comparative analytical assessment plan ...”

“The final comparative analytical assessment plan should include the risk ranking of attributes, the type of data evaluation to be used for each attribute/assay, and the final data analysis plan. The plan should specify the anticipated availability of both proposed biosimilar and reference product lots for evaluation of each attribute/assay and should include a rationale for why the proposed number of lots should be considered sufficient for the evaluation.”

Content of talk



RP revision shifts the focus

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**

With the revision it is planned to ...

- Keep nature of reflection - "no Guideline"
- Focus on QA-comparison highly relevant for regulatory decision making - "responsibility"
- Describe a framework for decision making - "flexibility":
 - elaboration concerning risk for false conclusion on similarity
 - evaluation of similarity criteria via operating characteristics
- Propose minimum set of aspects addressed prospectively - "Plan":
 - importance of certain QAs ("criticality")
 - discussion of relevance of differences in selected QAs
 - clarity re similarity criterion, analysis plan
 - sampling approach (feasibility /limitations ...)



Austrian
Federal Office for
Safety in Health Care

BASG

Thomas Lang

Biostatistician, Senior Assessor

BASG -

Austrian Federal Office for Safety in Health Care

Traisengasse 5

1200 Vienna

thomas.lang1@ages.at

www.basg.gv.at