ONE PART OF THE WHOLE: ANALYTICAL SIMILARITY & TOTALITY OF EVIDENCE

KATHERINE GIACOLETTI

MIDWEST BIOPHARMACEUTICAL STATISTICS WORKSHOP, MAY 14-16 2018
INDIANAPOLIS, IN
AGENDA

- Regulatory framework for biosimilarity
  - Totality of evidence, PK equivalence, Analytical Similarity (FDA), Comparability (EU)
- Statistical evaluations of comparability
  - Comparisons of means vs. comparisons of distributions
- Comparability in the context of biosimilars development
  - Potential approaches, challenges, opportunities
EU Directive (effective 2005) set up first regulatory pathway for biosimilars, but did not define them – so what is a “biosimilar”?

EMA Procedural advice (2013):
“… expected to have the same safety and efficacy profile …”

U.S. ACA (2010):
“… no clinically meaningful differences … in terms of the safety, purity, and potency …”
Both FDA and EMA use a stepwise approach to biosimilarity:

1. Analytical studies: structural & functional characterization
2. Animal studies: toxicity
3. PK/PD studies, including clinical immunogenicity
4. Additional clinical data
2012 EMA Guideline on Similar Biological Medicinal Products (proteins): quality issues –

“Quantitative ranges” for comparability, based on measured ranges of the reference product.

“… should not be wider than the range of variability of the representative reference medicinal product batches, unless otherwise justified.”

April 2015 Quality Considerations FDA GFI –

Acceptance criteria based on the totality of the analytical data and not simply on the observed range
PK profile similarity considered critical by both EMA & FDA

EMA (2012) – Similarity of PK profiles “essential part of biosimilar program”

Objective in biosimilar development is to detect possible differences in interaction with body between products, so “… observing 90% CIs of ratios of biosimilar to reference product within a pre-specified, justified acceptance range may not, in itself, be sufficient.”

FDA (2016) –

Standard BE criteria recommended: 90% CI for ratio of geometric means within acceptable limits, starting point 80 – 125%. Justify other limits if applicable.
2017 EMA “Reflection Paper on Statistical Methodology for the Comparative Assessment of Quality Attributes in Drug Development”

- Very broad scope – many comparability applications, many possible statistical approaches
- No firm guidance and not focused on biosimilarity assessments
- CHMP Guideline on Similar Biological Medicinal Products: scientific principles of biosimilar comparability are based on those applied for comparisons pre- and post-process changes (ICH Q5E)
REGULATORY FRAMEWORK FOR BIOSIMILARS

- 2017 Draft Statistical Approaches for Analytical Similarity FDA GFI –
  - Risk-based approach, rank QAs according to potential clinical impact of differences
  - Very specific statistical recommendations: 3-tiered approach
    - Tier 1: TOST with ±1.5σ_R equivalence margins, (1-2α)100% CI (assumes α=.05)
    - Tier 2: “Quality Range” – high proportion (e.g. 90%) of biosimilar results within \( \hat{\mu}_R \pm X\hat{\sigma}_R \), propose X=2
    - Tier 3: Graphical comparisons
STATISTICAL EVALUATIONS OF SIMILARITY

- **Two One-Sided Tests (TOST) procedure** (Schuirmann, 1987)
  - $H_0: \mu_T - \mu_R \leq \theta_1$ or $\mu_T - \mu_R \geq \theta_2$ vs. $H_1: \theta_1 < \mu_T - \mu_R < \theta_2$
  - Usually tested using CI approach: declare equivalence if CI within $[\theta_1, \theta_2]$
  - Original paper proposed $(1 - 2\alpha)100\%$ CI corresponds to level $\alpha$ test, but later work (Berger & Hsu, 1996) demonstrated this is often not the case and provide a method for constructing a $(1 - \alpha)100\%$ CI that corresponds to a level $\alpha$ test

- **Tolerance Interval Approaches**
  - Compare individual results or statistical interval based on comparator to TI based on reference

- **Similarity of distributions**
  - OVL, PSR, non-parametric approaches (Kolmogorov-Smirnov, Mann-Whitney U, etc.)
PROPORTION OF SIMILAR RESPONSE (PSR)

- Originally referred to as the overlapping coefficient OVL (Bradley, 1985) and later referred to as proportion of similar response PSR (Rom & Hwang, 1996)
- For independent samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ with densities $f$ and $g$, respectively, PSR is defined as

$$PSR(f, g) = \int min[f(x), g(x)]dx$$

- PSR ranges from 0 to 1, 0 = entirely disjoint populations, 1 completely overlapping distributions
ESTIMATION OF PSR

- For two normal populations with means $\mu_f$ and $\mu_g$ and common variance $\sigma^2$

$$PSR(f,g) = 2\Phi \left( \frac{-|\delta|}{2} \right),$$

where $\delta = \frac{\mu_f - \mu_g}{\sigma}$ is the standardized difference.

- Maximum Likelihood Estimate of PSR replaces $\mu_f$ and $\mu_g$ with $\bar{X}$ and $\bar{Y}$ and uses pooled within-group variance $s^2$ for estimating $\sigma^2$ (Inman & Bradley, 1994).


- Non-parametric approaches have been developed (Stine and Heyse, 2001).
Rom & Hwang (1996) present the general formulation of PSR for equal or unequal variances.

For X and Y, independent random variables with $X \sim N(\mu_x, \sigma^2)$ and $Y \sim N(\mu_y, \rho^2\sigma^2)$, $\rho > 0$, PSR is the area under the minimum of the two densities, which can be estimated by summing two ($\rho = 1$, equal densities at a single point) or three ($\rho \neq 1$, equal densities at two points) areas.

PSR in this more general case can be parameterized as a function of $\theta$ and $\lambda$,

$$\theta = \frac{\mu_x - \mu_y}{\sigma \sqrt{\rho}} \quad \text{and} \quad \lambda = \frac{(\log(\rho))}{2},$$

and as $\rho \to 1$, $\text{PSR} \to 2\Phi\left(\frac{-|\theta|}{2}\right)$.
PSR WITH UNEQUAL VARIANCES

- PSR takes into account differences in distributions due to differences in the mean or the variance.
- Estimation of PSR when variances differ is more complicated due to the fact that the probability density curves of the two populations have more than one point of intersection.
PSR FOR CLINICAL COMPARISONS: LOT CONSISTENCY

- Traditional comparison based on TOST (CI for ratio of GMTs (geometric mean titers) within $[-\delta, \delta]$ (e.g. [0.67, 1.5])

- Giacoletti & Heyse (2004) demonstrated application of PSR to evaluate lot consistency: 90% CI on PSR >0.8, simulations showed:
  - PSR controls Type I error at $\sim 0.05$ under $H_0$ for comparisons of 2 or 3 lots (pairwise) with equal variances (similar performance to GMT comparison)
  - PSR has slightly lower power for PSR (0.95) vs. GMT (0.98) comparison when true GMT ratio or true PSR = 1.0

  - **PSR has decreasing power as variance ratio increases, with means equal**
Lachenbruch, et al. (2004) noted that differences in variability have potentially important implications for patients/subjects.

- A “good” procedure would, among other things, “not require assumptions regarding equality of variances, control the Type I error probability close to nominal levels, be robust to non-normality, and be relatively easy to use.”

- “Ease of understanding by non-statisticians is also desirable.”
PSR FOR ANALYTICAL SIMILARITY

How to set acceptance criteria for analytical similarity using PSR?

- A mean difference of 0.125s with equal variances (referenced by FDA in power statement for TOST, n=10, equivalence margin of ±1.5s) corresponds to PSR = 0.95
  - with n=10 of each product, 85% probability that PSR 90% CI>0.59 or 95% CI>0.52
  - If true mean difference is 0 (more typical for power calculations) and equal variances (PSR=1), with n=10 of each product, 85% probability that 90% CI>0.60 or 95% CI>0.53
- A mean difference of 1.5s and equal variances corresponds to PSR of 45%
APPLICATIONS FOR COMPARABILITY EVALUATIONS IN PHARMACEUTICAL DEVELOPMENT/MANUFACTURING

- **Clinical:**
  - bioequivalence (generic drug approvals, etc.)
  - lot consistency for vaccines
  - Population bridging

- **Non-clinical:**
  - Manufacturing change
  - Analytical method bridging
  - Analytical similarity
CHOOSING STATISTICAL TECHNIQUES – IN GENERAL

- Evaluating the entire distribution rather than focusing on differences in the means is desirable
  - Better patient protection, more complete understanding of nature of differences and degree of risk
  - FDA Tier 2 approach does come closer to addressing the distributions rather than just the means, but without incorporating statistical uncertainty
- The choice of statistical analysis technique and acceptance criterion must also consider the context & the data:
  - What decision will be made based on the statistical conclusion?
  - What are the risks associated with making the wrong decision?
  - How are the data generated? Randomized, controlled clinical trial vs. convenience samples of commercial reference material
Critiques of the FDA approach to analytical similarity abound

Many focus on statistical deficiencies

- Tier 1 criterion less stringent than Tier 2
- Arbitrary criterion in each tier
- Failure to account for unknown, possibly unequal variances
- etc.
A key deficiency however is the lack of recognition of context of the question for analytical similarity

This is not a clinical trial!

Taking an analysis designed for a controlled, randomized clinical trial and blindly applying to most often non-random, non-independent, convenience sample manufacturing data (with unequal knowledge between about the data from reference and biosimilar) wreaks havoc with statistical properties of any analysis and thereby fails to provide either the patient or the biosimilar manufacturer with appropriate protection of risk.
HOW TO SET ACCEPTANCE CRITERIA FOR ANALYTICAL SIMILARITY?

- FDA 2017 draft Analytical Similarity GFI gives criteria for choosing equivalence margin:
  - Ensure biosimilar product values fall within reference product distribution
  - Unified basis for margin, regardless of different scales/magnitudes among QAs
  - Sufficient power for practical sample sizes
- PSR is able to satisfy these criteria, and more importantly, meet the objectives of the comparability analyses for clinical & non-clinical studies
CHOOSING STATISTICAL TECHNIQUES – ANALYTICAL SIMILARITY

- Is PSR a magic bullet for analytical similarity? No.

- The challenges for evaluating analytical similarity statistically remain the same – limited information about reference product lots, ability to link similarity criteria to patient risk, etc.

- However, PSR addresses an important objective for biosimilarity assessments

  - Where differences are found, the degree & nature of them, and associated risks, are more readily seen & assessed using PSR than tests of means.
CLINICAL RELEVANCE?

- Remember what we are trying to show:
  “… expected to have the **same safety and efficacy** profile …”
  “… no clinically meaningful differences … in terms of the **safety, purity, and potency** …”

- Are there opportunities for linking the analytical and PK (and other clinical) studies to address clinical relevance of QA differences?
  - Requires coordinated study designs & analyses – have to break down silos!
  - Analyses need to address full distribution of results, from both studies, and correlations among the non-clinical & clinical endpoints

**Totality of Evidence**
REFERENCES


ADDITIONAL REGULATORY REFERENCES & QUOTES
EU Directive (effective 2005) set up first regulatory pathway for biosimilars, but did not define them

“where a biological medicinal product which is similar to a reference biological product does not meet the conditions in the definition of generic medicinal products … the results of appropriate pre-clinical tests or clinical trials relating to these conditions must be provided.”

EMA Procedural advice (EMA/94051/2011, March 2013) defined biosimilar as

“… similar to a biological medicine that has already been authorized… expected to have the same safety and efficacy profile and are generally used to treat the same conditions.”

U.S. ACA (2010) definition of biosimilarity

“that the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components” and that “there are no clinically meaningful differences between the biological product and the reference product in terms of the safety, purity, and potency of the product.”
U.S. FDA – Section 351(k) of the Public Health Service Act (PHSA)

“... must contain, among other things, information demonstrating that ‘the biological product is similar to a reference product’ based upon data derived from: Analytical studies, Animal studies (including the assessment of toxicity); and ... clinical study or studies ... to demonstrate safety, purity, and potency ...”

CHMP

“Comparability studies are needed to generate evidence substantiating the similar nature, in terms of quality, safety and efficacy, of the similar biological medicinal product and the chosen reference medicinal product authorized in the EEA.”
KEY EMA GUIDANCE DOCUMENTS:

- Guideline on Similar Biological Medicinal Products (Revised 2014)
- Guideline on Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Non-Clinical and Clinical Issues (2014)
- Guideline on Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins as Active Substance: Quality Issues (2014)

REFLECTION PAPER ON STATISTICAL METHODOLOGY FOR THE COMPARATIVE ASSESSMENT OF QUALITY ATTRIBUTES IN DRUG DEVELOPMENT (2017)
REGULATORY FRAMEWORK FOR BIOSIMILARS

Key FDA Guidance Documents:

- April 2015 Scientific Considerations in Demonstrating Biosimilarity to a Reference Product
- April 2015 Quality Considerations in Demonstrating Biosimilarity of a Therapeutic Protein Product to a Reference Product
- December 2016 Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product
- September 2017 (Draft) Statistical Approaches to Evaluate Analytical Similarity
2014 EMA Guideline on Similar Biological Medicinal Products –

“The scientific principles of ... a biosimilar comparability exercise are based on those applied for evaluation of the impact of changes in the manufacturing process of a biological medicinal product (as outlined in ICH Q5E).”

Stepwise approach, starting with physiochemical & biological characterization. “Clinical data cannot be used to justify substantial differences in quality attributes.”
REGULATORY FRAMEWORK FOR BIOSIMILARS

- April 2015 Scientific Considerations FDA GFI –
  - Risk-based, totality of evidence, stepwise approach to evaluating biosimilarity
  - Emphasizes analytical characterization as foundational: minor structural differences can significantly affect safety and/or effectiveness

- April 2015 Quality Considerations FDA GFI –
  - “Acceptance criteria should be based on the totality of the analytical data and not simply on the observed range of product attributes of the reference product.”
2012 EMA Guideline on Similar Biological Medicinal Products (proteins): non-clinical & clinical issues –

- Nature & complexity of reference product affect extent of studies & analytical differences guide planning of further studies
- Similarity of PK profiles “essential part of biosimilar program”
  - Objective in biosimilar development is to detect possible differences in interaction with body between products, so “…observing 90% CIs of ratios of biosimilar to reference product within a pre-specified, justified acceptance range may not, in itself, be sufficient.”

2016 Clinical Pharmacology Data to Support Demonstration of Biosimilarity FDA GFI

- Clin Pharm studies build on comparative analytical studies, addressing residual uncertainties
- **Standard BE criteria recommended:** 90% CI for ratio of geometric means within acceptable limits, starting point 80 – 125%. Justify other limits if applicable.
2016 Clinical Pharmacology Data to Support Demonstration of Biosimilarity FDA GFI –

- Clin Pharm studies build on comparative analytical studies, addressing residual uncertainties
  - Details four possible determinations from analytical similarity & resulting implications for further studies, including clin pharm
- Objective is to evaluate similarities & differences in PK & PD profiles, important to assessing whether clinically meaningful differences exist
- Standard BE criteria recommended: 90% CI for ratio of geometric means within acceptable limits, starting point 80 – 125%. Justify other limits if applicable.