FDA WORKSHOP: INCREASING THE EFFICIENCY OF BIOSIMILAR DEVELOPMENT PROGRAMS

September 19, 2022

9:00 am – 4:00 pm Eastern Time

FDA WORKSHOP: INCREASING THE EFFICIENCY OF BIOSIMILAR DEVELOPMENT PROGRAMS

KEYNOTE

Robert M. Califf, M.D.
Commissioner of Food and Drugs
Food and Drug Administration

FDA WORKSHOP: INCREASING THE EFFICIENCY OF BIOSIMILAR DEVELOPMENT PROGRAMS

WORKSHOP OVERVIEW

Sarah Yim, M.D.
Director, Office of Therapeutic Biologics and Biosimilars
Food and Drug Administration

FDA U.S. FOOD & DRUG ADMINISTRATION

FDA Workshop: Increasing the Efficiency of Biosimilar Development Programs

September 19, 2022

9:00 am – 4:00 pm Eastern Time

9:00 am – 9:05 am **Welcome**

Christine Corser, PharmD, MS, Analyst, Policy Staff, Office of Therapeutic Biologics and Biosimilars (OTBB)

9:05 am – 9:15 am **Keynote**

Robert Califf, MD, Commissioner, FDA

9:15 am – 9:20 am **Introduction and Overview of the Workshop**

Sarah Yim, MD, Director, OTBB

9:20 am – 10:30 am **Session #1: The Integration of Analytical and Clinical Information to Enhance the Efficiency of Biosimilar Development Programs**

Moderator: Emanuela Lacana, PhD, Deputy Director, OTBB

Speakers/Panelists:

- Peter Stein, MD, Director, Office of New Drugs (OND)
- Steven Kozlowski, MD, Director, Office of Biotechnology Products (OBP)
- Steven Lemery, MD, Director, Division of Oncology III (DO3), OND
- Stacey Ricci, MEng, ScD, Director, Scientific Review Staff (SRS), OTBB
- Joel Welch, PhD, Associate Director of Science and Biosimilar Strategy, OBP

10:30 am – 10:45 am **Break**

10:45 am – 12:15 pm **Session #2: Innovative Statistical Methods for Integration of Data Sources Informing Biosimilar Comparative Clinical Studies**

Moderator: Thomas Gwise, PhD, Director, Division of Biostatistics (DB9), Office of Biostatistics (OB)

Speakers/Panelists:

- Shein-Chung Chow, PhD, Professor of Biostatistics & Bioinformatics, Duke University School of Medicine
- Johanna Mielke, PhD, Data Scientist, Bayer Pharma AG
- Matthew Psioda, PhD, Head of Statistical Innovation, GSK
- Danyu Lin, PhD, Professor, Department of Biostatistics, University of North Carolina
- Peter Thall, PhD, Professor, Department of Biostatistics, The University of Texas MD Anderson Cancer Center

Literature Background for Session #2:

Mielke, J., Schmidli, H., & Jones, B. (2018). Incorporating historical information in biosimilar trials: Challenges and a hybrid Bayesian-frequentist approach. *Biometrical Journal*, *60*(3), 564–582. https://doi.org/10.1002/bimj.201700152

Psioda, M. A., Hu, K., Zhang, Y., Pan, J., & Ibrahim, J. G. (2020). Bayesian design of biosimilars clinical programs involving multiple therapeutic indications. *Biometrics, 76*(2), 630–642. https://doi.org/10.1111/biom.13163

Zeng, D., Pan, J., Hu, K., Chi, E., & Lin, D. Y. (2018). Improving the power to establish clinical similarity in a Phase 3 efficacy trial by incorporating prior evidence of analytical and pharmacokinetic similarity. *Journal of Biopharmaceutical Statistics, 28*(2), 320–332. https://doi.org/10.1080/10543406.2017.1397012

12:15 pm – 1:00 pm **Lunch Break**

1:00 pm – 1:45 pm **Continuation of Session #2**

1:45 pm – 2:00 pm **Break**

2:00 pm – 3:30 pm **Session #3: Biosimilar Comparative Clinical Endpoint Study Design: Choices to Optimize Efficiency**

Moderator: Stella Grosser, PhD, Director, Division of Biostatistics 8 (DB8), OB

Speakers/Panelists:

- Carol Kim, PharmD, Scientific Reviewer, SRS, OTBB
- Kathy Fritsch, PhD, Mathematical Statistician, Division of Biostatistics 3, OB
- Jessica Kim, PhD, Mathematical Statistician Team Leader, DB8, OB
- Wanjie Sun, PhD, Mathematical Statistician Team Leader, DB8, OB
- Yow-Ming Wang, PhD, Associate Director for Biosimilars and Therapeutic Biologics, Office of Clinical Pharmacology
- Steven Lemery, MD, Director, DO3, OND
- Nikolay Nikolov, MD, Director, Division of Rheumatology and Transplant Medicine, OND

3:30 pm – 4:00 pm **Workshop Summary and Concluding Remarks**

Sarah Yim, MD, Director, OTBB

FDA Workshop: Increasing the Efficiency of Biosimilar Development Programs

THE INTEGRATION OF ANALYTICAL AND CLINICAL INFORMATION TO ENHANCE THE EFFICIENCY OF BIOSIMILAR DEVELOPMENT PROGRAMS

FDA Review of Biosimilars:
2022 Status Update

Increasing the Efficiency of Biosimilar Development Programs
FDA Public Workshop - September 19, 2022

M. Stacey Ricci, M.Eng., Sc.D.,
Director, Scientific Review Staff
Office of Therapeutic Biologics and Biosimilars | OND | CDER

U.S. FOOD & DRUG
ADMINISTRATION

You can't really know where you are going until you know where you have been.

Maya Angelou

BPD Programs and Approvals

As of **September 2, 2022**:

- CDER has received meeting requests to discuss the development of biosimilars for **47 different reference products**
- **106** Biosimilar Product Development (BPD) Programs have been enrolled
- FDA has received **60 BLAs** and approved **38 biosimilars, 3 as interchangeable**; 22 products marketed

FDA Approved Biosimilar and Interchangeable Products

B Biosimilar

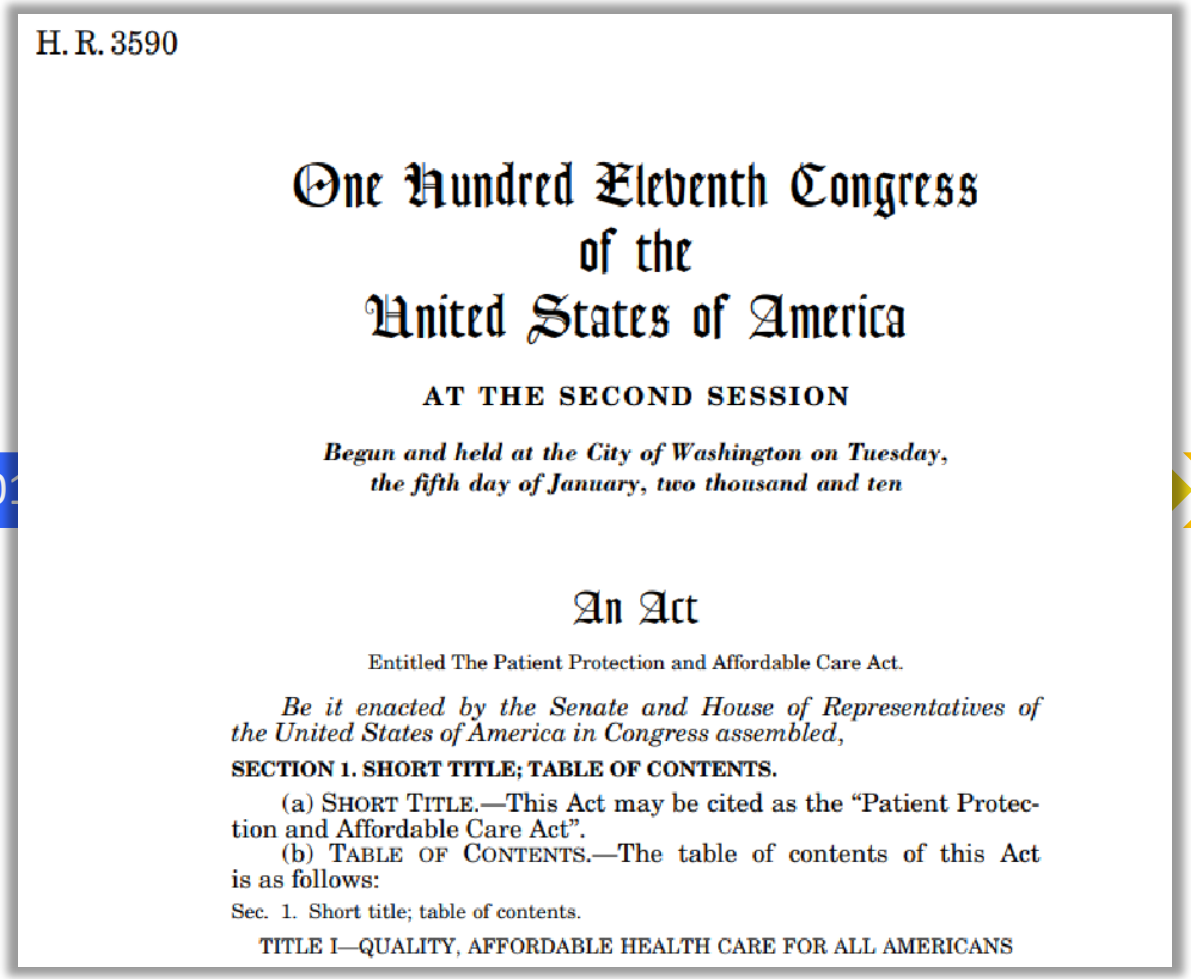
I Interchangeable Biosimilar

	Product Class	Approvals
Supportive Care	Filgrastim	B B B
	Epoetin	B
	Pegfilgrastim	B B B B B B
Oncology	Rituximab	B B B
	Bevacizumab	B B B
	Trastuzumab	B B B B B
Autoimmune	Infliximab	B B B B
	Etanercept	B B
	Adalimumab	B I B B B B B
	Insulin Glargine	I B
Ophthalmology	Ranibizumab	B I

Key Biosimilar Milestones

2010 2011 2012 20 2020 2021 2022

BPCIA grants FDA the authority to approve biosimilar and interchangeable products



H. R. 3590

One Hundred Eleventh Congress
of the
United States of America

AT THE SECOND SESSION

Begun and held at the City of Washington on Tuesday, the fifth day of January, two thousand and ten

An Act

Entitled The Patient Protection and Affordable Care Act.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "Patient Protection and Affordable Care Act".

(b) TABLE OF CONTENTS.—The table of contents of this Act is as follows:

Sec. 1. Short title; table of contents.

TITLE I—QUALITY, AFFORDABLE HEALTH CARE FOR ALL AMERICANS

General Requirements

A 351(k) application must include information demonstrating that the biological product:

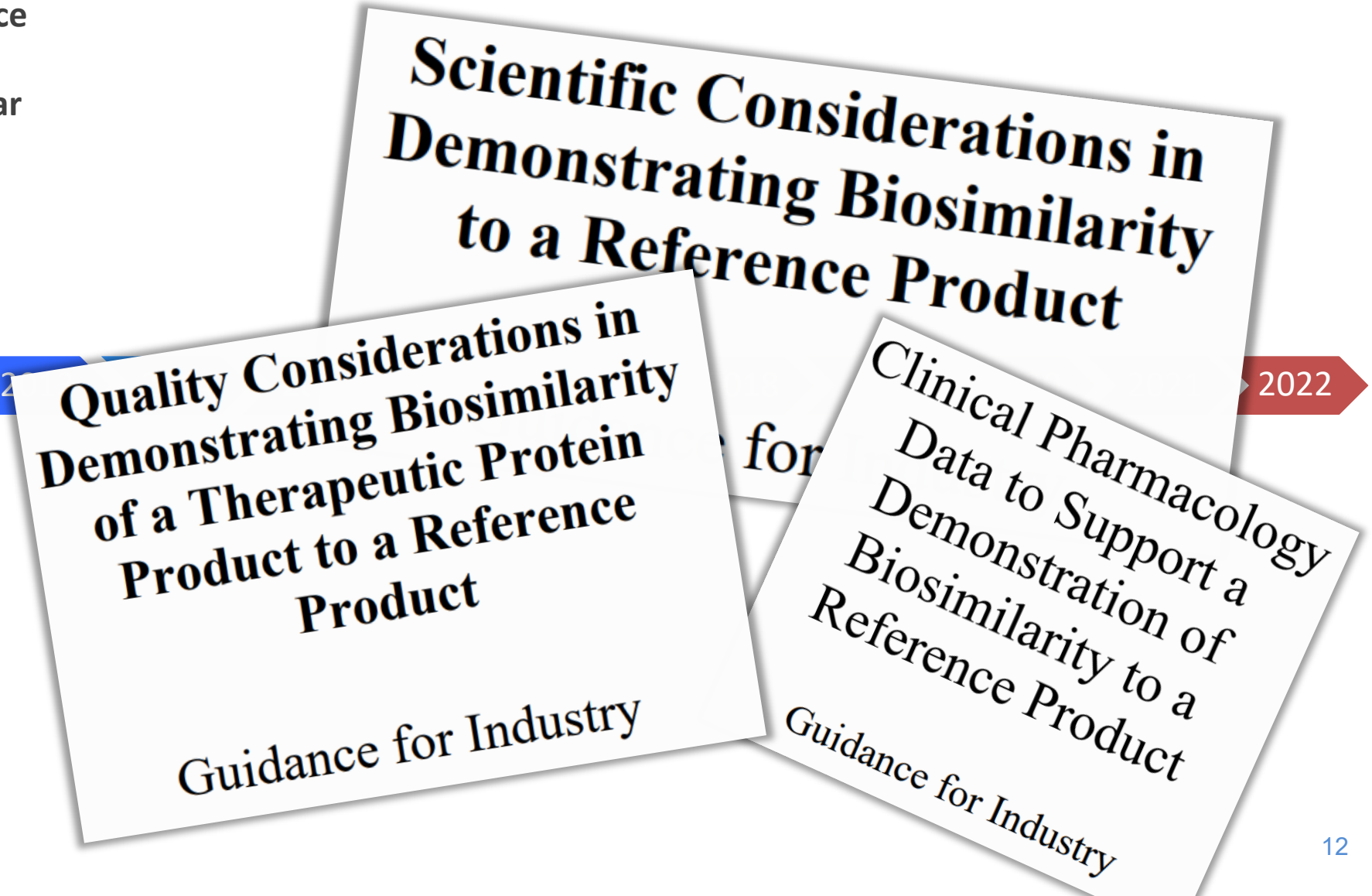
- Is **biosimilar** to a reference product
  - **Highly similar to and has no clinically meaningful differences from the** FDA-approved reference product
- Utilizes the **same mechanism(s) of action** for the proposed condition(s) of use -- but only to the extent the mechanism(s) are known for the reference product;
- **Condition(s) of use** proposed in labeling **have been previously approved** for the reference product;
- Has the **same route of administration**, **dosage form**, and **strength** as the reference product; and
- Is manufactured, processed, packed, or held in a facility that **meets standards** designed to assure that the biological product continues to be **safe, pure, and potent.**

Key Biosimilar Milestones

FDA publishes guidance on recommended approach for biosimilar development (analytical, animal, clinical studies)

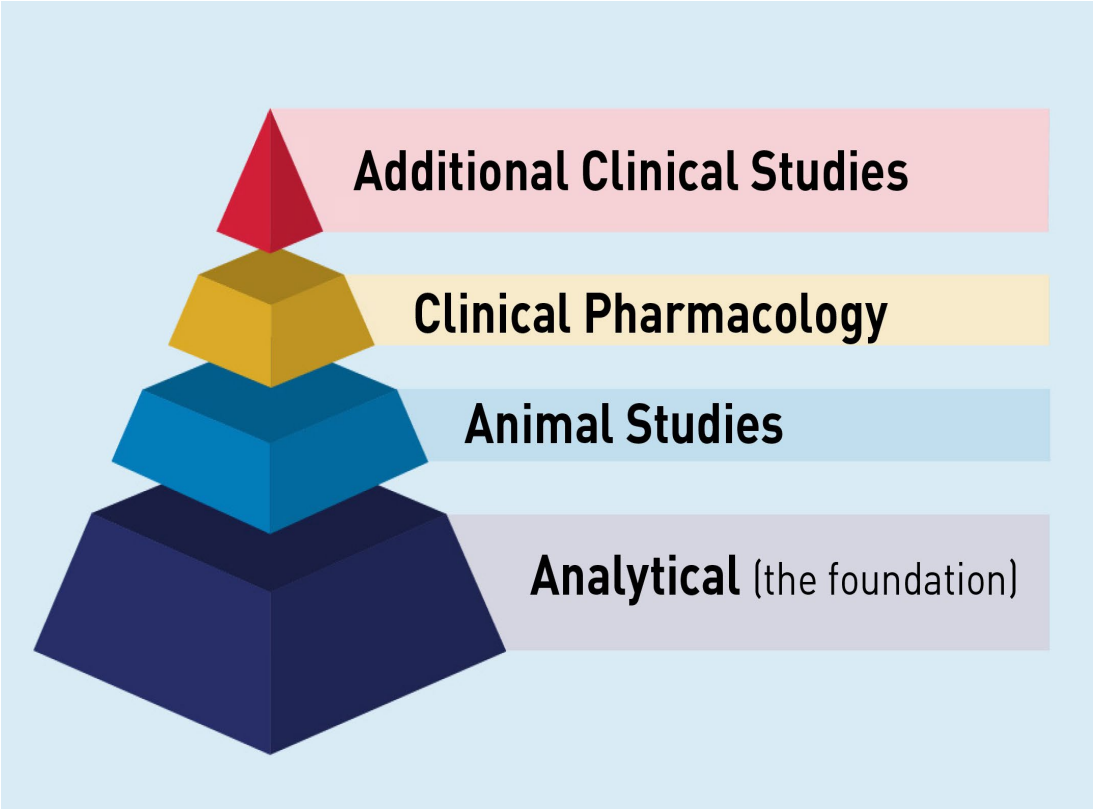
2010 | 2011 | 2012 | 2022

BPCIA grants FDA the authority to approve biosimilar and interchangeable products



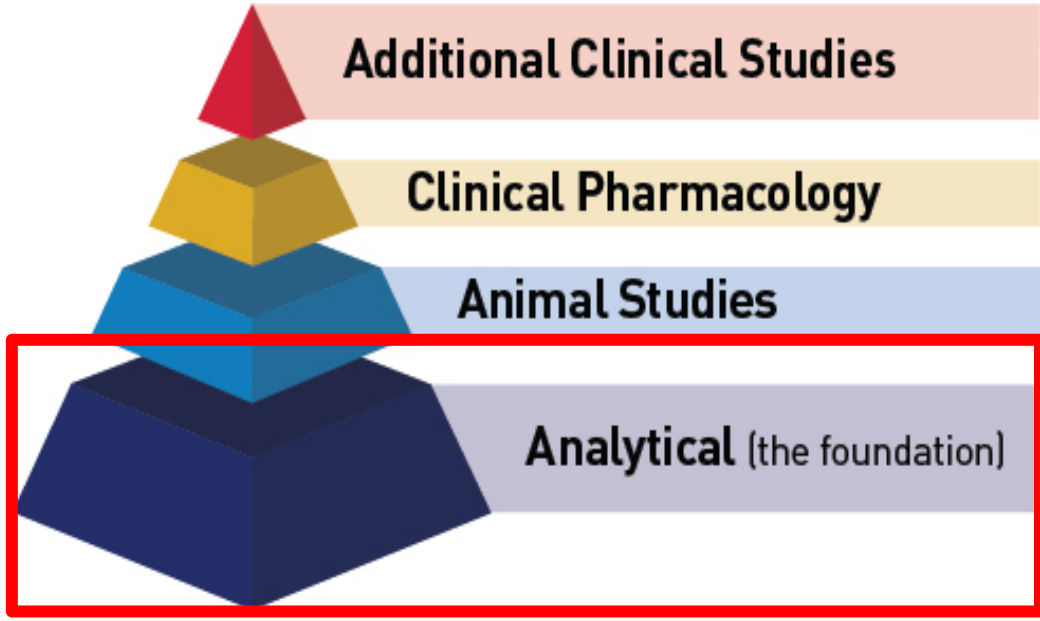
Recommendations: Demonstrating Biosimilarity

- **Goal:** to establish biosimilarity between proposed product and reference product, not to establish safety & effectiveness
- Stepwise approach to generate data in support of a demonstration of biosimilarity
- *Totality-of-the-evidence* approach to evaluating biosimilarity



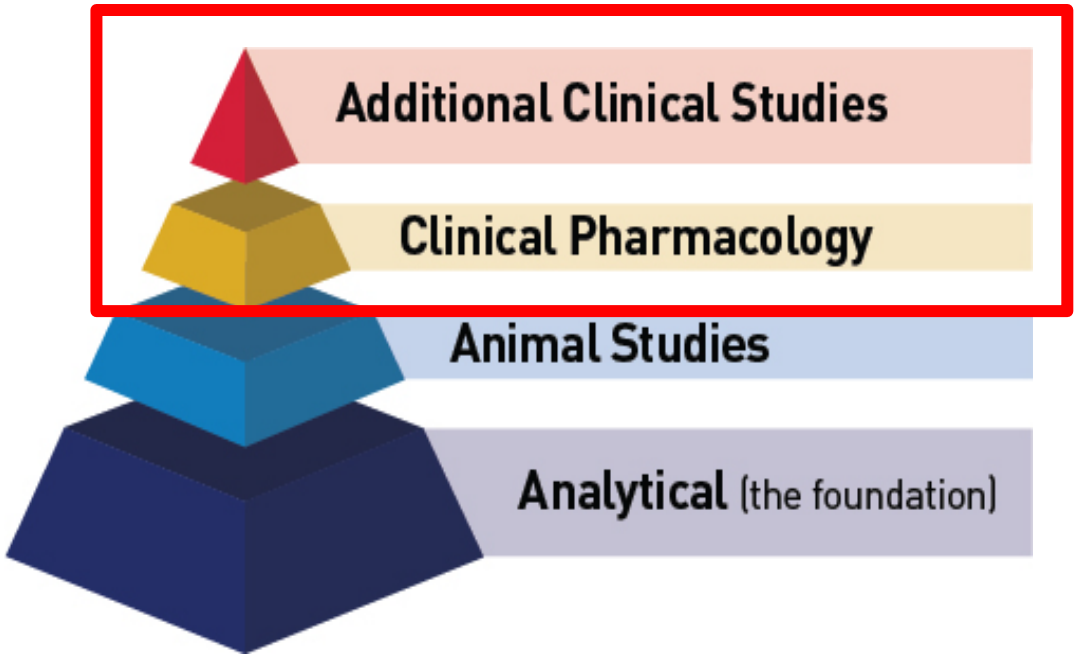
Comparative Analytical Data is the Foundation

- Compare multiple physicochemical and biological attributes of each product
  - Analytical studies are generally more **sensitive** than clinical studies in detecting differences between products, should differences exist
  - A biosimilar product with **highly similar structure and function** to the reference product should perform like the reference product (i.e., have **similar efficacy and safety** as the reference product)
- Analyze quality attributes from multiple lots of reference product and proposed biosimilar
  - Structural (physicochemical) properties
  - Functional (biological) activities



Role of Clinical Studies

- The nature and scope of clinical studies will depend on the extent of residual uncertainty about the biosimilarity of the two products after conducting structural and functional characterization, and animal studies
- FDA generally expects an adequate clinical PK comparison, and PD if relevant, between the proposed biosimilar product and reference product
- An assessment of immunogenicity is also expected

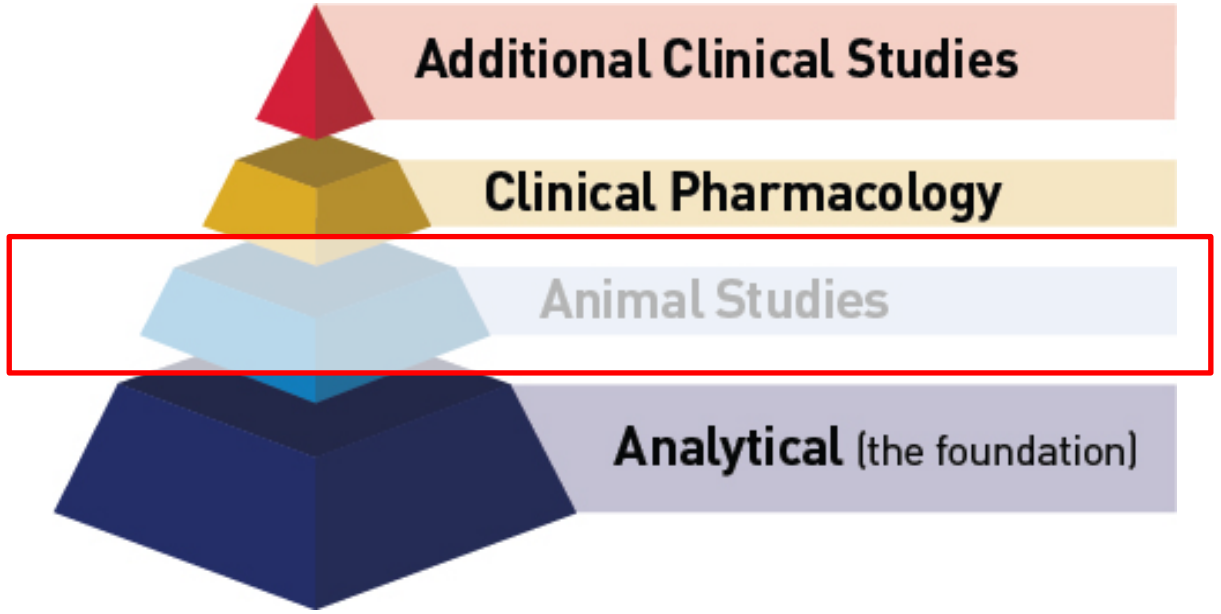


Key Biosimilar Milestones

BPCIA grants FDA the authority to approve biosimilar and interchangeable products

FDA publishes guidance on recommended approach for biosimilar development (analytical, animal, clinical studies)

First biosimilar approved in the U.S.



2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022

Key Biosimilar Milestones

BPCIA grants FDA the authority to approve biosimilar and interchangeable products

FDA publishes guidance on recommended approach for biosimilar development (analytical, animal, clinical studies)

First biosimilar approved in the U.S.

FDA publishes gui on recommended approach for interchangeabili

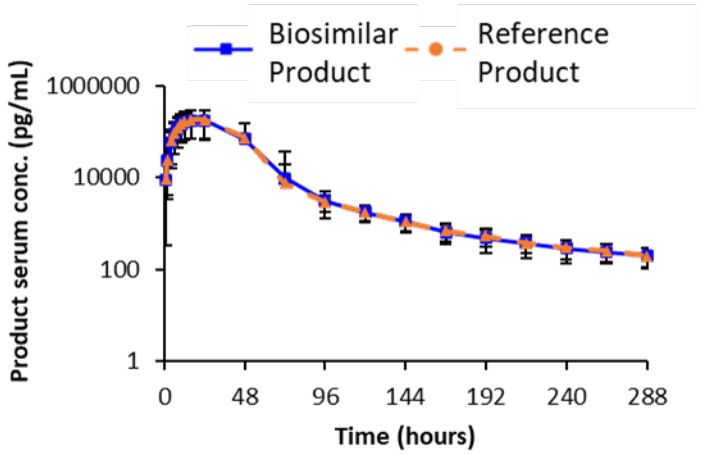
New guidance with revised recommendations on analytical data for biosimilars

2010 2011 2012 2013 2014 2015 2016 2017

B

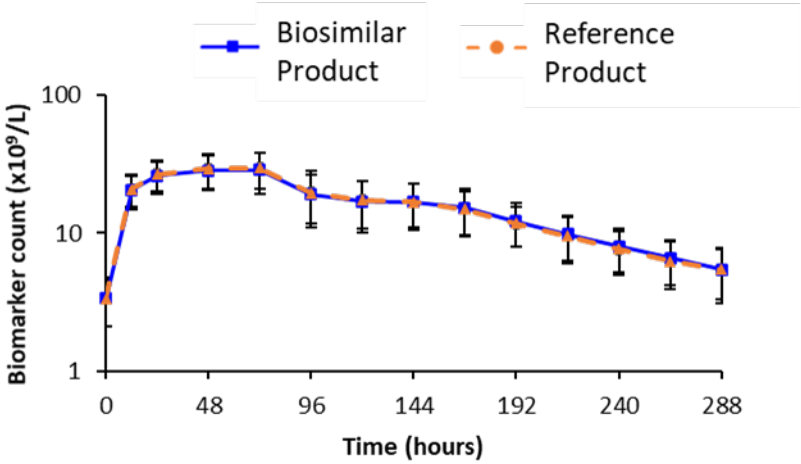
Clinical Studies to Demonstrate Similar Exposure, Efficacy, and Safety

Similar Exposure



PK Similarity

Similar Efficacy



PD Similarity

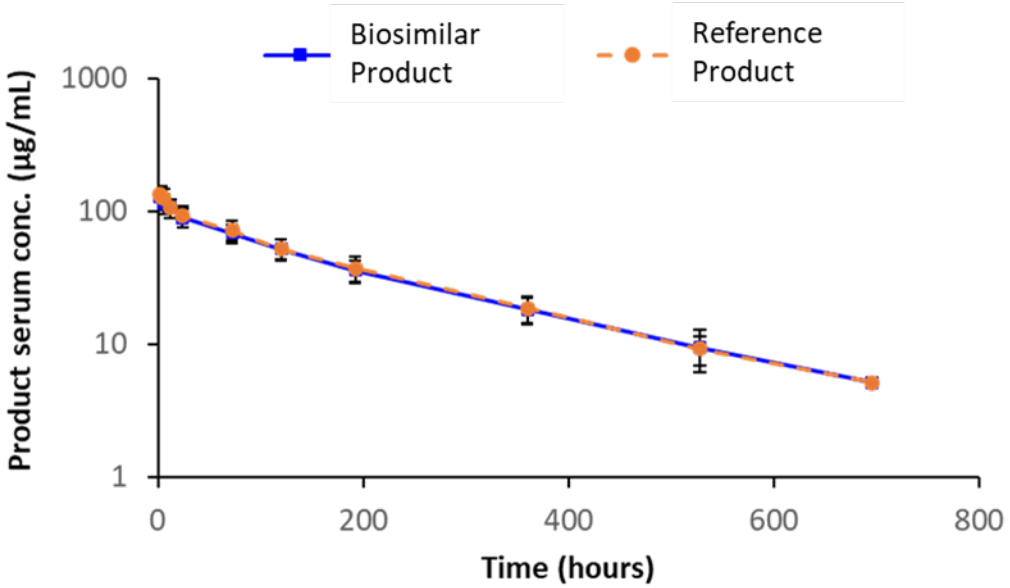
Similar Safety

	Biosimilar Product	Reference Product
Immunogenicity	Similar	
Adverse events	Similar	

Comparative Immunogenicity/Safety

Clinical Studies to Demonstrate Similar Exposure, Efficacy, and Safety

Similar Exposure



PK Similarity

Similar Efficacy

Similar Safety

	Biosimilar Product	Reference Product
Response rate, n (%)	116/248 (46.8%)	129/256 (50.4%)
90% CI for risk ratio estimate	0.7981–1.0796	
Equivalence margin	0.740–1.350	
Immunogenicity	Similar	
Adverse events	Similar	

CCS for Efficacy and safety (immunogenicity)

Addressing “residual uncertainty”

- Current paradigm uses either PD or efficacy endpoints to support the demonstration of no clinically meaningful differences
- As understanding continues to grow of comparative analytical data, and physiological role of product quality attributes (and small differences that may exist in those attributes), this will reduce the uncertainty that drives decisions about clinical data
- See draft guidance for industry: *Clinical Immunogenicity Considerations for Biosimilar and Interchangeable Insulin Products* (November 2019)

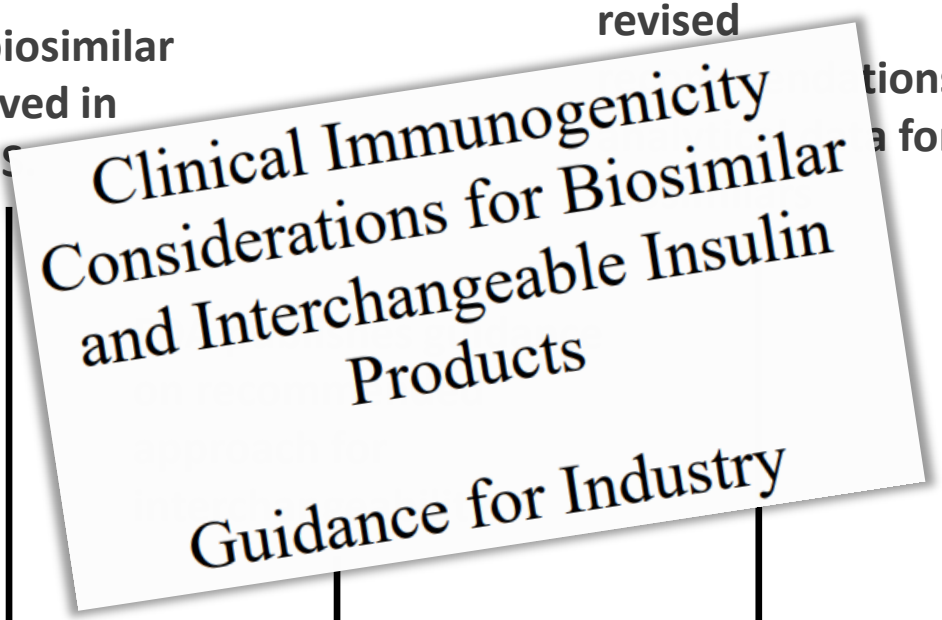
Key Biosimilar Milestones

BPCIA grants FDA the authority to approve biosimilar and interchangeable products

FDA publishes guidance on recommended approach for biosimilar development (analytical, animal, clinical studies)

First biosimilar approved in the U.S

New guidance with revised ...tions on ...a for



First interchangeable biosimilar approved in the U.S.

2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022


FDA
U.S. FOOD & DRUG
ADMINISTRATION

A Firm Foundation – The Power of Analytics in Biosimilar Development

Joel Welch, Ph.D.
Associate Director for Science & Biosimilar Strategy
Chair of Emerging Technology Team
Office of Biotechnology Products, Office of Pharmaceutical Quality
CDER | US FDA

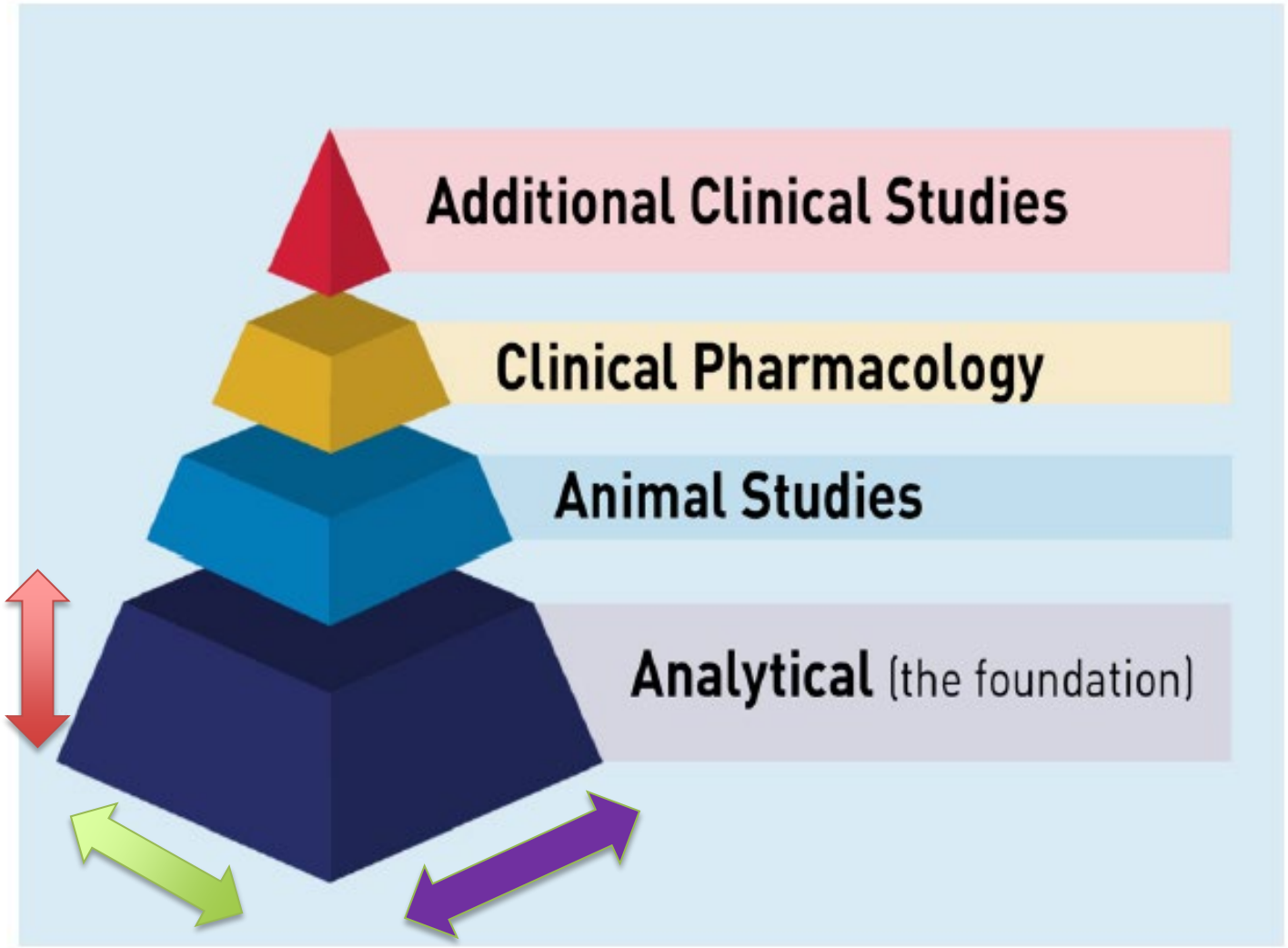
FDA Public Workshop: Increasing the Efficiency of Biosimilar Development Programs

Introduction

Analytics are the base of the pyramid in biosimilar development.

A “base” is defined by:

- The Breadth of Analytics (how wide)
- The Reproducibility of Analytics (how thick)
- The Sensitivity of Analytics (how deep)



The Breadth of Analytics

A Hypothetical Release Specification of NME Drug Substance (n=12)
Appearance
pH
HCP
Osmolality
Protein Concentration
Identity
Glycan Profile
Aggregates
Fragments
Charge
Potency
Bioburden/Endotoxin

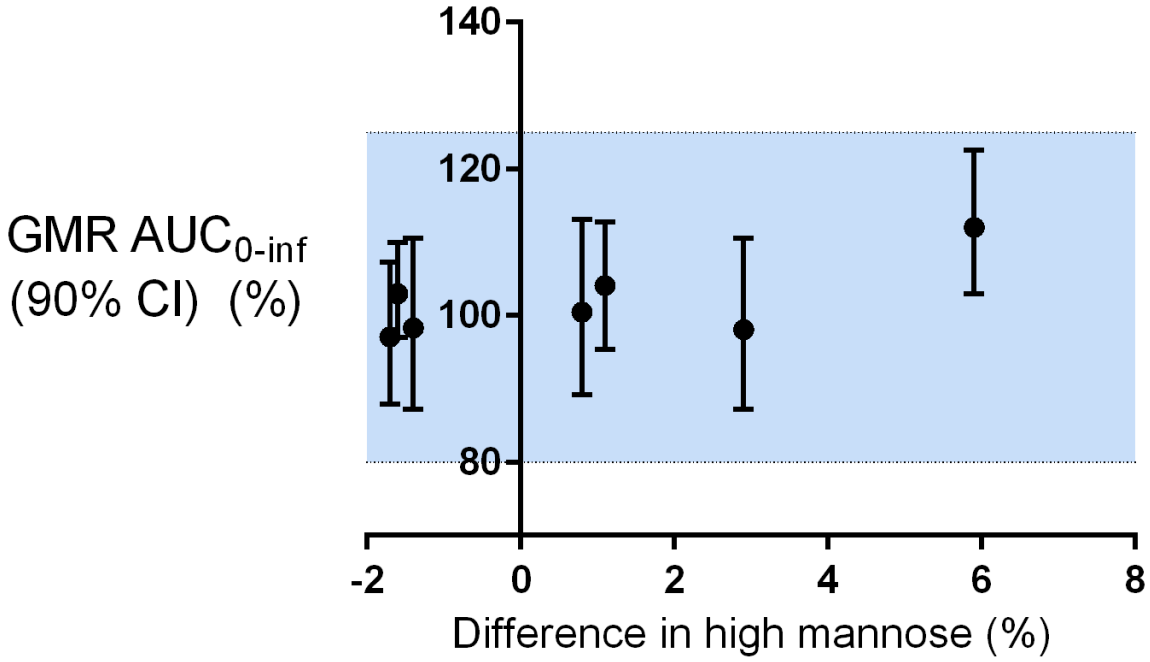
New Attributes →

Orthogonal Techniques →

Additional Comparisons →

Hypothetical Biosimilar Candidate (Release + Comparative Analytical Assessment) 40+ Tests	
Appearance	pH
Osmolality	Protein Concentration
HCP	Bioburden/Endotoxin
Identity	Glycan Profile (Individual structural comparisons)
Aggregates (SE-HPLC), (AUC), (FFF)	Fragments
Charge (iCIEF and CEX)	Target Binding – soluble, membrane), (Iso_1, Iso_2, Iso_x)
Potency (MoA_1, MoA_2, MoA_3, MoA_4, MoA_x), (Absence of Unexpected MOA) (Receptor $Binding_1$, Receptor $Binding_X$)	
Higher Order Structure (CD), Disulfide Analysis, Free Thiol	Thermodynamic Stability (DSC)
Stability Profile Comparison (Oxidation, Multiple Temperatures)	Primary Sequence (Glycosylated vs. Not), (100% Coverage)

The Reproducibility of Analytics



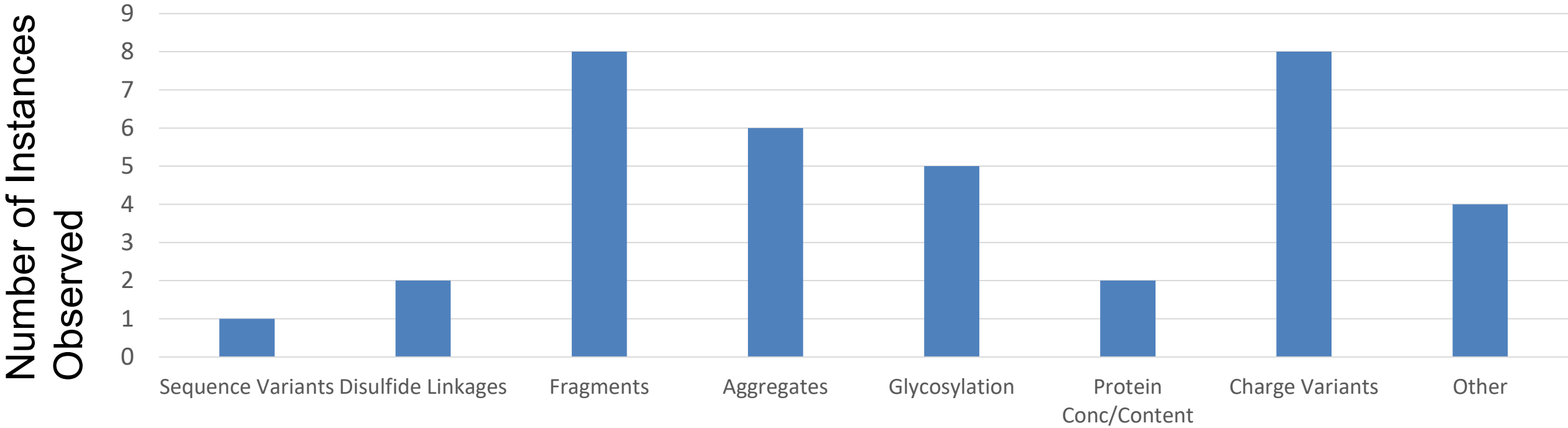
We gathered PK similarity study results for 7 biosimilar monoclonal antibody programs where the biosimilar and reference product had non-overlapping ranges for high mannose content.

analytical data can detect subtle differences in high mannose content that are not distinguishable by pharmacokinetic profile

From J. Welch et. al., "The Mannose in the Mirror: A Reflection on the Pharmacokinetic Impact of High Mannose Glycans of Monoclonal Antibodies in Biosimilar Development." Submitted.

The Sensitivity of Analytics

Subset of approved applications selected over a fixed period of time (n=12)



Information and risk assessment provided were able to address differences

Differences Observed

Conclusions

- Biosimilar development shows robustness of analytics:
  - To deeply and thoroughly interrogate all of molecule
  - To generate highly reproducible data
  - To identify differences that matter and those too small to impact clinical performance


FDA
U.S. FOOD & DRUG
ADMINISTRATION

Comparative Clinical Trials in Biosimilar Development

Steven Lemery, MD, MHS

Director, Division of Oncology 3

U.S. FOOD & DRUG ADMINISTRATION

Drug	Biosimilars Approved
Avastin	3
Enbrel	2
Epogen	1
Herceptin	5
Humira	7
Lantus	2
Lucentis	2
Neupogen	3
Neulasta	6
Remicade	4
Rituxan	3

*As of September 2, 2022 (does not differentiate interchangeable biosimilar drugs)

Comparative Clinical Studies – to a Reference Bevacizumab Product

Drug	Disease	N	ORR ratio (90% CI)	ORR diff% (95% CI)
MYL-14020 (B) (Mylan)	NSCLC	671	0.96 (0.83, 1.12)	-1.6 (-9, 5.9)
FKB238 (B) (Centus)	NSCLC	731	0.96 (0.86, 1.08)	-2 (-9, 0. 6)*
ABP215 (B) (Amgen)	NSCLC	642	0.93 (0.80, 1.09)	-2.7 (-9.3, 3.5)
PF-06439535 (B) (Pfizer)	NSCLC	719	1.02 (0.89, 1.16)	0.7 (-6.6, 7.9)
MB02 (B) (Amneal)	NSCLC	627	0.91 (0.78,1.06)	-4.0 (-11.8. 3.7)
SB8 (B) (Samsung)	NSCLC	763	1.11 (0.98, 1.27)	5.3 (-2.2, 12.9)
LY01008 (B) (Luye Pharma)	NSCLC	649	0.91 (0.80, 1.04)	-4.5 (NS)
BI695502 (B) (Boehringer Ingelheim)	NSCLC	671	0.86 (0.77, 0.95)	-9.1 (-16.6,-1.4)
CT-P16 (B) (Celltrion)	NSCLC	689	1.01 (0.88, 1.17)	0.4 (-7.0, 7.8)
BCD-021 (Biocad)	NSCLC	357	1.02 (.80, 1.32)	0.81 (-9.5, 11.1)
HLX04 (Shanghai Henlius)	CRC	677	0.92 (0.80, 1.05)^	-4.2 (-10.6, 2.1)^
Total N		7196		

+ ORR diff means a higher ORR point estimate for biosimilar product; *90%CI;

Comparative Clinical Studies – to a reference Rituximab product

Drug	Disease	N	ORR diff % (95% CI)
CT-P10 (Celltrion)	LTBFL	258	1.8 (-6.2, 10.0)
CT-P10 (Celltrion)	FL	140	5.7 (-1.7, 14.7)
IBI301 (Innovent)	DLBCL	419	-3.9 (-9.1, 1.3)*
ABP798 (Amgen)	FL	256	7.7 (-1.4, 16.8)
HLX01 (Shanghai Henlius)	DLBCL	407	1.4 (-3.6, 6.3)
PF-05280586 (Pfizer)	LTBFL	394	4.7 (-4.2, 13.5)
DRL_RI (Dr. Reddy)	DLBCL	151	-8.9 (-24.86, 6.7)#
BCD-020 (Biocad)	iNHL	174	2.8 (-12.6, 18.2)
RTXM83 (mAbxience)	DLBCL	272	0.7 (-8.8, 10.2)
GP2013 (Sandoz)	FL	629	−0.4% (−5·9 to 5·1)
5 Comparative RA studies Celltrion; Amgen; Sandoz; Pfizer; Dr. Reddy	RA	1485	
Total N		4585	

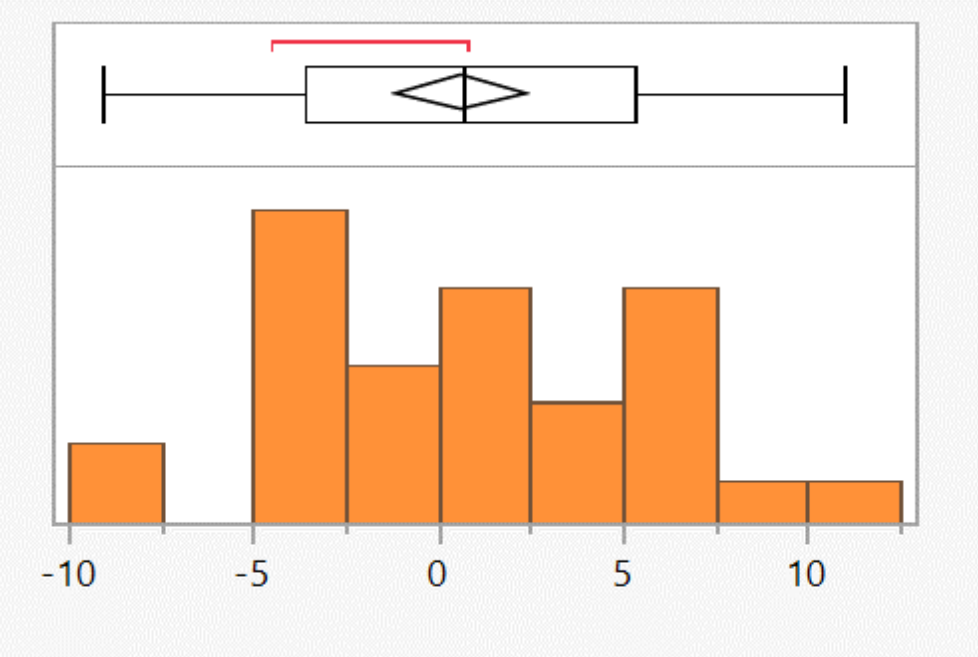
LTBFL=low tumor burden follicular lymphoma; DLBLC = Diffuse large B cell lympmhoma; NHL = non-Hodgkin lymphoma; RA = Rheumatoid Arthritis
*95% CI

Comparative Clinical Studies – to a Reference Trastuzumab Product

Drug	Breast Cancer Setting	Endpoint	N	ORR diff (9%% CI)
TA4415V (Orchid)	esBC	pCR	92 (NI)	5.6* (-0.23,0.12)
HD201 (Prestige)	esBC	tpCR	502	-3.8 (-12.8,5.4)
HLX02 (Shanghai Henlius)	mBC	ORR_{24}	649	-0.1 (-7,6.9)
BCD-022 (JSC BIOCAD)	mBC	ORR	225	6.0 (-8.05,19.9)
TX05 (Tanvex)	esBC	pCR	809	3.5
Pf-05280014 (Pfizer)	mBC	ORR	707	-4.0 (-11.0, 3.1)
Pf-05280014 (Pfizer)	esBC	pCR	226	-2.8 (-16.58,10.96)
ABP980 (Amgen)	esBC	pCR	725	7.3 (1.2,13·4)
SB3 (Samsung)	esBC	pCR	800	11 (4.1,17.3)^
CT-P6 (Celltrion)	esBC	pCR	549	-3.6 (-12 to 5)
MYL-1401O (Mylan)	mBC	ORR	500	5.6 (-3.08 to 14.04)
Total N			5784	

Higher number favors trastuzumab
*ITT analysis listed but was not primary
^upper limit outside margin

Histogram of point estimates of ORR Differences



May not include all studies presented at meetings but not published

- At least 32 trials
- 16,080 patients (*not including Rituxan RA or PK studies*)
- 11 US approvals of 3 drugs

Takeaways

- These CCS are “blunt instruments”
  - Trial-to-trial variation exists (even if Avastin/Herceptin/Rituxan were compared in hypothetical trials to themselves)
  - If a result does not fall within a margin
    - The products were different, or
    - Type II error
  - Some (hypothetical) products could demonstrate comparable clinical results without being highly similar (e.g., for products dosed above saturation)
- Overall resource intensive
- Are there other ways to assess residual uncertainties???

References

1. Shi Y, et al., Cancer Commun. 2021; 41(9):889-903
2. Qin S, et al., BioDrugs. 2021; 35(4): 445-458
3. Reck, et al., Lung Cancer. 2020; 146:12-18
4. Reinmuth N, et al., BioDrugs. 2019; 33(5):555-570
5. Thatcher N, et al., Clin Cancer Res. 2019; 25(7):2088-2095.
6. Socinski MA, et al., Ther Adv Med Onc. 2021; 13.
7. Syrigos K, et al., BioDrugs. 2021; 34(4):417-428
8. Trukhin D, et al., BioDrugs. 2021; 35(4):429-444.
9. Kim ES, et al., JTO Clin Res Rep. 2021; 3(1)
10. Ohe Y et al., Cancer Res. 2022; 82 (12 Supp): CT551.
11. Stroyakovskiy DL, et al., BMC Cancer; 22: Article 129.
12. Candelaria M, Leuk Lymphoma, 2019
13. Jurczak W, The Lancet, 2017.
14. Kim WS, Lancet Haematol, 2017
15. Kwak L, Clin Lymphoma Myeloma Leuk, 2022 (results from FDA review at drugs@fda) – hard to find in paper
16. Song Y, Adv Ther, 2021
17. Burmester G, Clin Rheum, 2020
18. Haridas VM, BioDrugs, 2020
19. Park W, MAbs, 2018
20. Poddubnaya IV, Hematol Oncol, 2020
21. Niederwieser D, Target Oncol, 2020
22. Sharman JP, BioDrugs, 2020
23. Shi Y, J Hematol Oncol, 2020
24. Smolen JS, Ann Rheum Dis, 2017
25. Viswabandya A, J Glob Oncol, 2019
26. Williams JH, Br J Clin Pharmacol, 2016
27. Nodehi RS et al., BMC Pharmacol Toxicol. 2022; 23(1):57
28. Pivot X et al., JAMA Oncol. 2022; 8(5).698-705
29. Xu B et al., BioDrugs. 2021; 35(3):337-350
30. Alexeev SM, BMC Cancer. 2020; 20:783
31. Pegram MD, Br J Cancer. 2019; 120(2):172-182.
32. Lammers PE, Br J Cancer. 2018; 119(3):266-273.
33. Von Minckwitz G, et al., Lancet Oncol. 2018; 19(7):987-998.
34. Pivot X et al., JCO. 2018; 36(10); 968-974.
35. Stebbing J et al., Lancet Oncology. 2017; 18(7): 917-928.
36. Rugo HS, et al., JAMA. 317(1); 37-47
37. Annals of Oncology (2021) 32 (suppl_5): S407-S446. 10.1016/annonc/annonc68


FDA
U.S. FOOD & DRUG
ADMINISTRATION

THE INTEGRATION OF ANALYTICAL AND CLINICAL INFORMATION TO ENHANCE THE EFFICIENCY OF BIOSIMILAR DEVELOPMENT PROGRAMS

PANEL DISCUSSION

FDA WORKSHOP: INCREASING THE EFFICIENCY OF BIOSIMILAR DEVELOPMENT PROGRAMS

BREAK

We will return and start Session #2 at 10:45 am Eastern Time

FDA Workshop: Increasing the Efficiency of Biosimilar Development Programs

INNOVATIVE STATISTICAL METHODS FOR INTEGRATION OF DATA SOURCES INFORMING BIOSIMILAR COMPARATIVE CLINICAL STUDIES

Current Critical/Challenging Issues and Possible Solutions in Biosimilar Development

Shein-Chung Chow, PhD
Department of Biostatistics and Bioinformatics
Duke University School of Medicine
Durham, North Carolina, USA

Presented at
FDA Scientific Workshop, Silver Spring, Maryland
September 19, 2022

DukeMedicine

Outline

- Regulatory perspectives
  - FDA guidance
  - Stepwise approach
- Current critical/challenging issues and possible solutions
  - Analytical similarity assessment
  - Non-inferiority/similarity margin selection
  - Interchangeable biosimilar
  - Multiple indications (extrapolation)
  - Non-medical switch
- Concluding remarks

Definition of biosimilarity

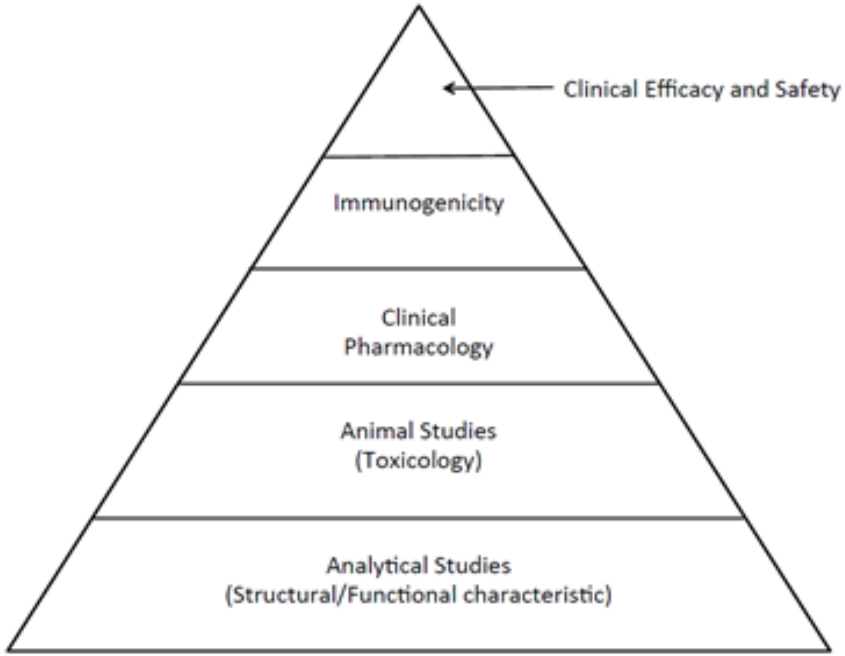
A biosimilar product:

Is highly similar to the reference product notwithstanding minor differences in clinically inactive components

There are no clinically meaningful differences in terms of safety, purity and potency.

DukeMedicine

Stepwise approach for obtaining totality-of-the-evidence

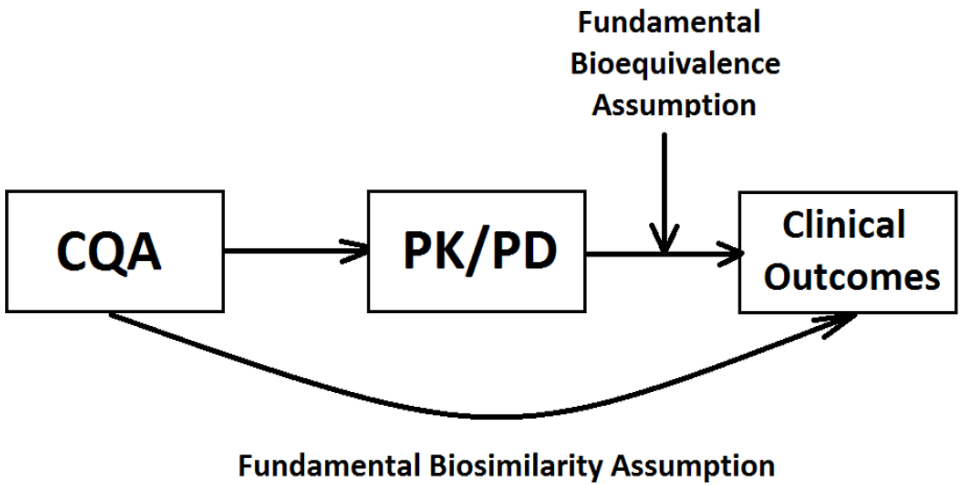


Stepwise approach

Stepwise approach consists of analytical, PD/PD and clinical similarity

- Analytical similarity
  - Critical quality attributes at various stages of manufacturing process
- PK/PD similarity
  - Pharmacokinetics (PK)
  - Pharmacodynamics (PD)
- Clinical similarity
  - The assessment of immunogenicity
  - Safety/tolerability
  - Efficacy

Relationship between *in vitro* testing and *in vivo* testing



Analytical similarity assessment

- FDA recommended 3-tier approach
- Step 1
  - Identify critical quality attributes (CQAs) that are relevant to clinical outcomes
- Step 2
  - Classification of CQAs into three tiers according to their criticality or risk ranking relevant to clinical outcomes
- Step 3
  - Similarity assessment at each tier

Analytical similarity assessment

- Tier 1 CQAs
  - Most relevant to clinical outcomes
  - Equivalence test
- Tier 2 CQAs
  - Mild-to-moderate relevant to clinical outcomes
  - Quality range approach
- Tier 3 CQAs
  - Least relevant to clinical outcomes
  - Raw data and graphical comparison

FDA's perspectives

- The equivalence test has been criticized
  - Data-dependent and inflexible
  - Need an efficient test procedure with flexible margin
- Possible solutions
  - Equivalence test with flexible margin

    (Lee, Oh, and Chow, 2019, *Enliven: Biosimilars Bioavailab.*, 3(2): 5-11)
  - Bayesian approach

    (Chiu, Liu, and Chow, 2014, *JBS*, 24: 1254-1263).

DukeMedicine

Non-inferiority/similarity margin selection

- FDA 2016 revised guidance is often considered for selection of non-inferiority margin
- Concerns
  - There is often a disagreement between the margin proposed by the sponsor and the margin suggested by the FDA
  - The guidance is based on statistical reasoning rather than clinical judgement
- ICH guideline indicates that both statistical reasoning and clinical judgement should be taken into consideration

FDA's perspectives

- Current practice
  - Estimate the margin based on a systematic review and meta-analysis which include several previous studies by retaining 50%-80% of the treatment effect (e.g., the half width of the CI obtained from a meta analysis)
- Statistical challenges
  - Selection bias
  - Study difference and possible treatment-by-study interaction (validity of poolability)
- Possible solutions
  - Bayesian approach with an appropriate prior

Recent development

- When there is a disagreement, it is suggested risk/benefit assessment be performed if the estimated margin is deviated from the truth
- The risk/benefit assessment can be performed in both ways:
  - Assuming the margin proposed by the FDA is true
  - Assuming the margin proposed by the sponsor is true
- Then, find a balance point which is acceptable to both the sponsor and the regulatory agency

Issues of interchangeability

The biological product to be interchangeable with the reference product if

(A) the biological product

(i) is biosimilar to the reference product; and

(ii) can be expected to produce the same clinical result in any given patient; and

(B) for a biological product that is administered more than once to an individual, the risk in terms of safety or diminished efficacy of alternating or switching between use of the biological product and the reference product is not greater than the risk of using the reference product without such alternation or switch.

DukeMedicine

Issue of interchangeability

- FDA draft guidance *Considerations in Demonstrating Interchangeability With a Reference Product* (2017, 2020)
- An interchangeable product can be expected to produce the same clinical result as the reference in any given patient
  - In practice, it is not possible to demonstrate this in any given patient
  - However, it is possible to demonstrate this in any given patient with certain assurance, e.g., the assessment of individual bioequivalence.

Issue of interchangeability

- Possible solution
  - Consider the following probability

$$p = Prob\left(\begin{array}{c}\text{Test product can produce the same}\\ \text{clinical resutls as that of the reference}\\ \text{product in any given patient}\end{array}\right)$$

  - We can then test the following hypotheses using historical data or real-world data in conjunction with Bayesian approach

$$H_0: p \leq p_0 \quad \text{versus} \quad H_a: p > p_0$$

DukeMedicine

Biosimilar with multiple indications

- The applicants need to provide sufficient scientific justification for extrapolation, which should be able to address the following issues
  - Mechanism of action (MOA)
  - Pharmacokinetics (PK)
  - Immunogenicity
  - Difference in expected toxicities
  - Any other factors that may affect the safety and efficacy of the product in each condition of use

DukeMedicine

Extrapolation across indications

- In 2017 ODAC meeting, the ODAC indicated that they feel uncomfortable in support of biosimilar claim without seeing clinical data from other indications not studied.
- Possible solution
  - Bayesian design involving multiple indications

    (Psioda et al. 2020, *Biometrics*. 76(2): 630–642)

DukeMedicine

Published in final edited form as:
Biometrics. 2020 June ; 76(2): 630–642. doi:10.1111/biom.13163.

Bayesian Design of Biosimilars Clinical Programs Involving Multiple Therapeutic Indications

Matthew A. Psioda[1,*], **Kuolung Hu**[2], **Yang Zhang**[3], **Jean Pan**[4], **Joseph G. Ibrahim**[1]

[1]Department of Biostatistics, University of North Carolina, McGavran-Greenberg Hall, CB#7420, Chapel Hill, NC 27599

[2]Ionis Pharmaceuticals Inc., 2855 Gazelle Court, Carlsbad, CA 92010

[3]Atara Biotherapeutics Inc., 2380 Conejo Spectrum Street, Suite 200, Thousand Oaks, CA 91320

[4]Amgen Inc., One Amgen Center Drive, 24-1-B, Thousand Oaks, CA 91320

Summary:

In this paper, we propose a Bayesian design framework for a biosimilars clinical program that entails conducting concurrent trials in multiple therapeutic indications to establish equivalent efficacy for a proposed biologic compared to a reference biologic in each indication to support approval of the proposed biologic as a biosimilar. Our method facilitates information borrowing across indications through use of a multivariate normal correlated parameter prior (CPP) which is constructed from easily interpretable hyperparameters that represent direct statements about the equivalence hypotheses to be tested. The CPP accommodates different endpoints and data types across indications (e.g., binary and continuous) and can therefore be used in a wide context of models without having to modify the data (e.g., rescaling) to provide reasonable information borrowing properties. We illustrate how one can evaluate the design using Bayesian versions of the type I error rate and power with the objective of determining the sample size required for each indication such that the design has high power to demonstrate equivalent efficacy in each indication, reasonably high power to demonstrate equivalent efficacy simultaneously in all indications (i.e., globally), and reasonable type I error control from a Bayesian perspective. We illustrate the method with several examples, including designing biosimilars trials for follicular lymphoma and rheumatoid arthritis using binary and continuous endpoints, respectively.

Non-medical switch

- Non-medical switch (NMS) is referred to the switch from the reference product (more expensive) to an approved biosimilar product (less expensive) based on factors unrelated to clinical/medical considerations.
- It is a concern that this non-medical switch may present unreasonable risk (e.g., reduced efficacy or increase of the incidence rate of adverse events) to patient population with the diseases under study.

Possible solution

- It is suggested that non-medical switch should be evaluated following recent FDA guidance on interchangeability
  - Based on historical data or real-world data in conjunction with Bayesian approach
- Statistical considerations
  - Control arm and endpoint selection
  - Possible confounding/interaction
  - Heterogeneity of variability
  - Sample size requirement under the study design employed

DukeMedicine

Enliven Archive
Intensifying Thoughts

Research Article | www.enlivenarchive.org | Enliven: Biosimilars and Bioavailability

Non-Medical Switch in Biosimilar Product Development

Shein-Chung Chow[1*]

[1]Duke University School of Medicine, Durham, North Carolina, USA

*Corresponding author: Shein-Chung Chow, Professor, Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA, E-mail: sheinchung.chow@duke.edu

Received Date: 21st March 2018
Accepted Date: 23rd March 2018
Published Date: 30th March 2018

Citation: Chow SC (2018) Non-Medical Switch in Biosimilar Product Development Enliven: Biosimilars Bioavailab 2(1): e001.

Abstract

For an approved biosimilar product, it is a common practice that the provider (pharmacist or insurance company) may switch from the innovator product to the approved biosimilar product based on factors unrelated to clinical/medical consideration. In practice, it is a concern that this non-medical switch may present unreasonable risk (e.g., reduced efficacy or increase of the incidence rate of adverse events) to patients with the diseases under study. In recent years, several observational studies and a national clinical study (NOR-SWITCH) were conducted to evaluate the risk of non-medical switch from a reference product to an approved biosimilar product. The conclusions from these studies, however, may be biased and hence misleading due to lack of some scientific and/or statistical deficiencies in design and analysis of the data collected. In this article, valid study designs and appropriate statistical methods are recommended for a more accurate and reliable assessment of potential risk of medical/non-medical switch between a proposed biosimilar product and a reference product. The results can be easily extended for evaluation of the potential risk of medical/non-medical switch among multiple biosimilar products and a reference product.

Keywords: Drug interchangeability; Switching; Alternation; NOR-SWITCH; Switching design

Concluding remarks

- The use of historical data or real-world data in conjunction with Bayesian approach can help in evaluation of biosimilar drug products
  - Provide scientific justification (rationale)
  - Improve efficiency
- Statistical methods under valid study design should be developed
  - Complete n-of-1 trial design
  - Seamless adaptive design
  - Master protocol
  - Bayesian adaptive design

DukeMedicine

Johanna Mielke
Acknowledgment: Byron Jones, Heinz Schmidli

Incorporating historical information in biosimilar trials

Acknowledgment and disclaimer

I am presenting my personal opinion today - this should not be understood as the opinion of Novartis, Sandoz or Bayer.

This project was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 999754557. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Swiss Government.

Historical data and biosimilars

In biosimilar development:

- the originator product has already been on the market for several years when the biosimilar development begins
- the originator was already studied very often, both prior to market authorization and in post-marketing studies

Idea: incorporate this historical information into the Phase III studies that are used for the approval of the biosimilar with a Bayesian approach

- Summarize historical data in a prior distribution
- Combine historical data with data in new study to obtain posterior distribution (here: meta-analytic-predictive approach)
- Note: historical data is only used for the reference product!

Challenge: Type I error rate inflation is expected

Example: binary endpoint

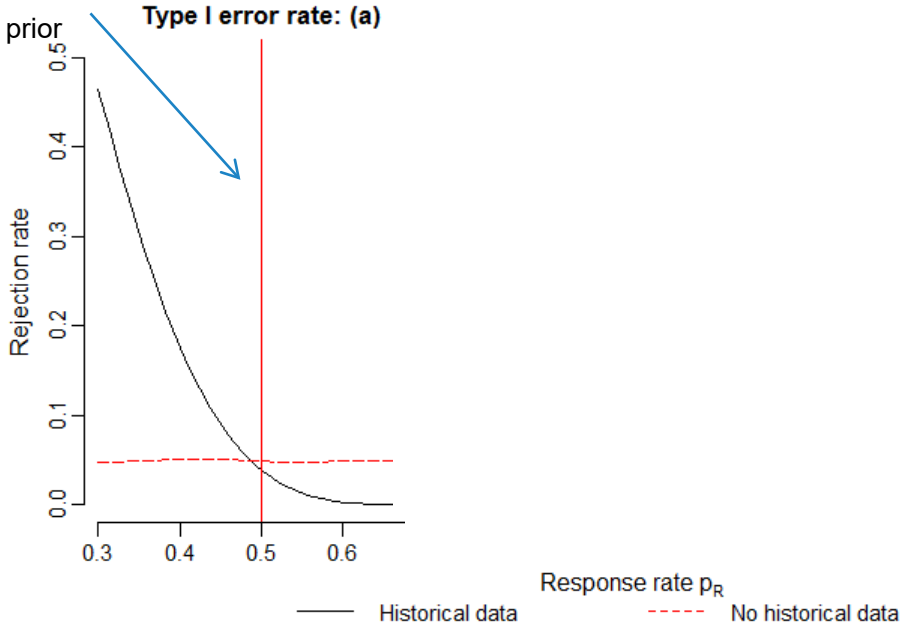
- Compare responder rate in both groups
- Goal is to confirm equivalence in response rates of the test (T) and reference (R) product:

$$H_0 : |p_R - p_T| \geq \Delta \text{ vs. } H_1 : |p_R - p_T| < \Delta$$

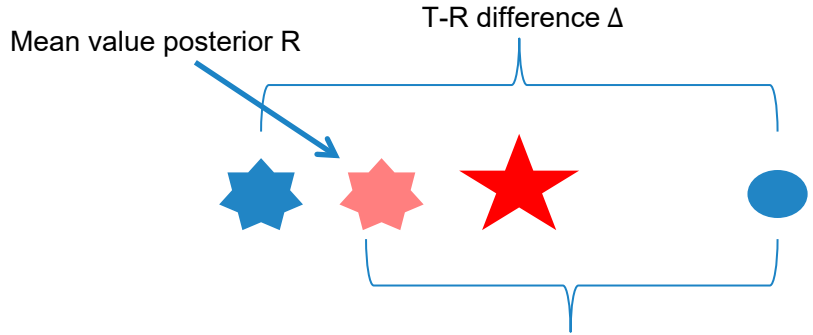
- Evaluate operating characteristics typically in 3 situations:
  - $p_T = p_R + \Delta$ (Type I error rate: Situation (a))
  - $p_T = p_R - \Delta$ (Type I error rate: Situation (b))
  - $p_R = p_T$ (power)

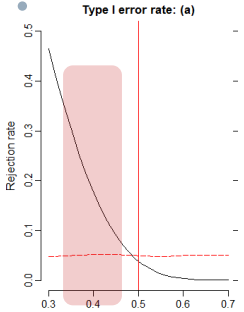
Operating characteristics

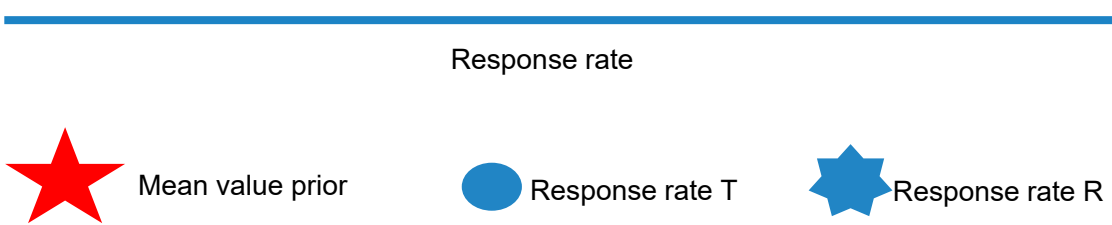
Mean value of the prior



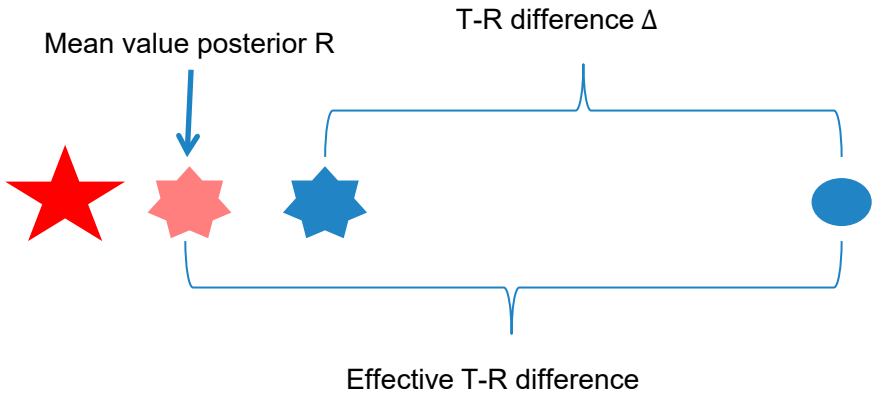
Why do we observe this profile?

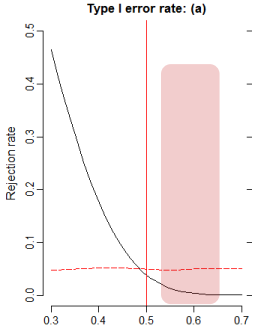


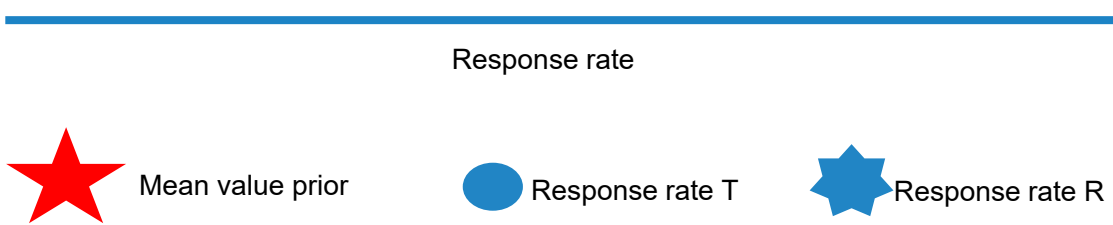




Why do we observe this profile?

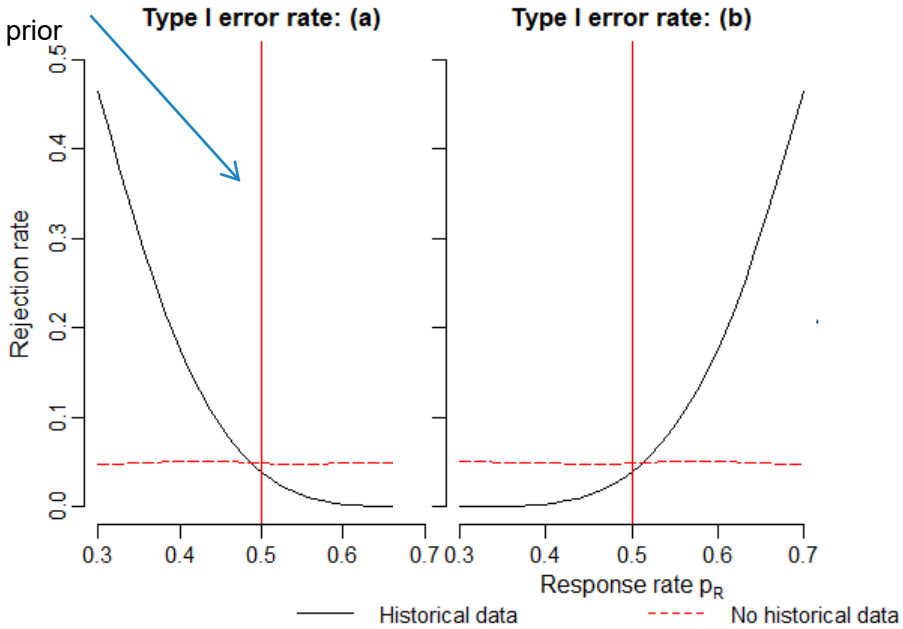






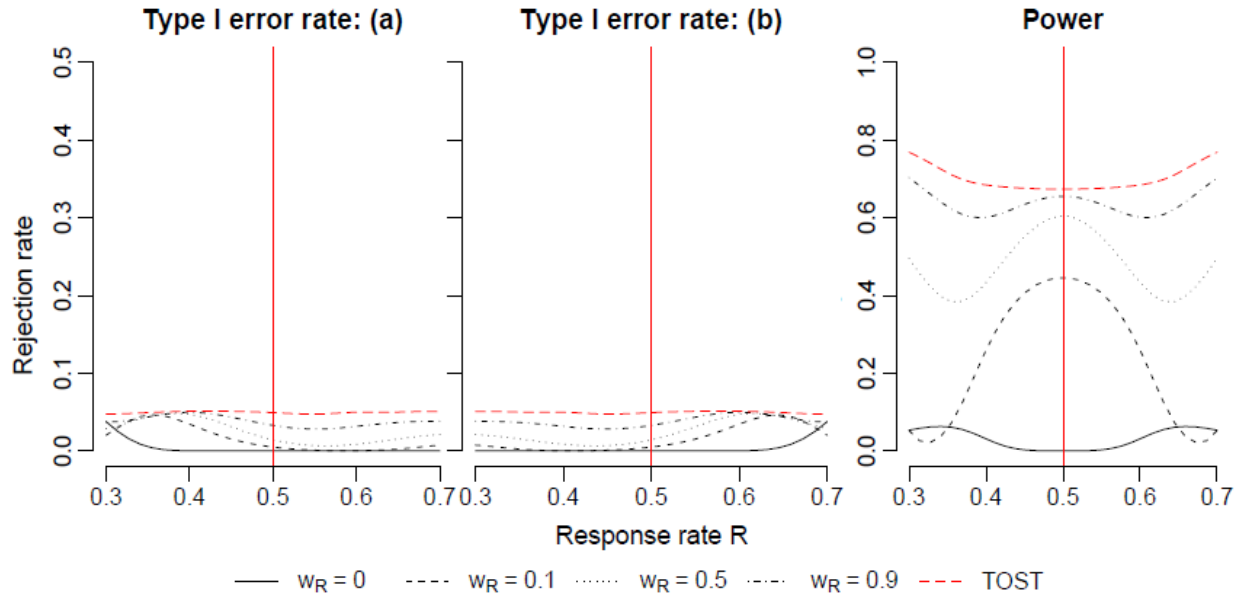
Operating characteristics

Mean value of the prior



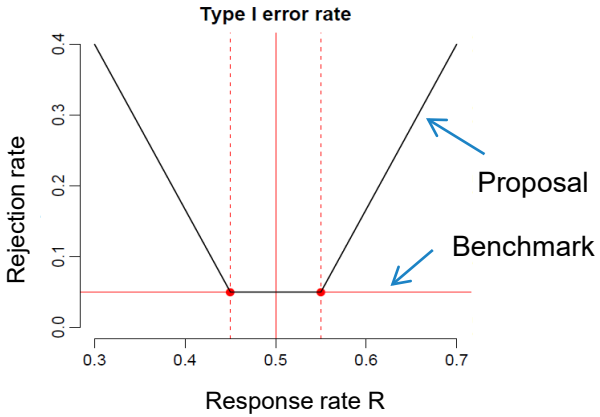
There is no gain in power if we require full control of Type I error rate!

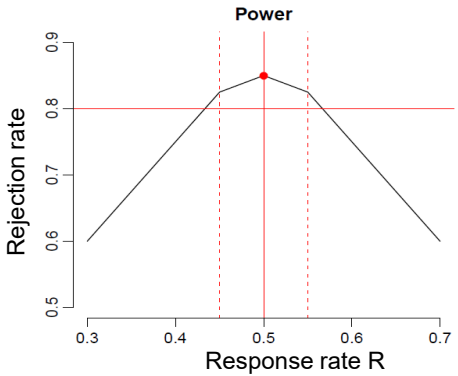
Robust MAP priors (or similar) do not allow for gain in power if T1E is controlled



Partial Type I error rate control

- We accept that strict Type I error rate control is incompatible with a gain in power
- For biosimilars, we expect that it is possible to conduct a "similar" study
- We define an interval C in which we aim to control the Type I error rate
- Note: standard approaches do not give a relevant gain in power even if only partial Type I error rate control is required





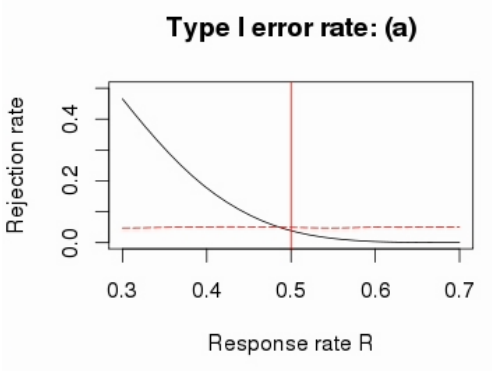
Overview of proposed method

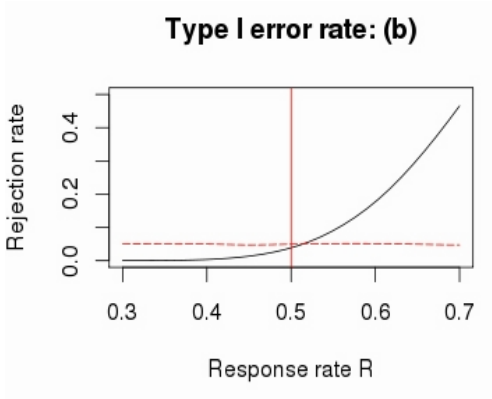
- Main goal: gain in power while controlling the Type I error rate in interval C
- Main concepts:
  - Switching rule I: if response rate of R in the new study and in the historical data are *very** different, do not use historical data
  - Switching rule II: if the response rates for T and R are *very** similar, use *lower** critical value
  - *Response rate-dependent critical values**

**: tuning parameters, can be chosen either automatically or be specified by the user*

Response rate-dependent critical values

- The Type I error rate highly depends on the response rate in the new study
  ➔ Set the critical value high in regions in which the test is too liberal, and low in regions in which the test is too conservative
- The location of these regions depends on the ordering of the response rate of T and R in the new study (Situation (a) vs. Situation (b))
  ➔ Use different critical values for Situations (a) and (b)
- True response rate is not known
  ➔ Use estimated response rate

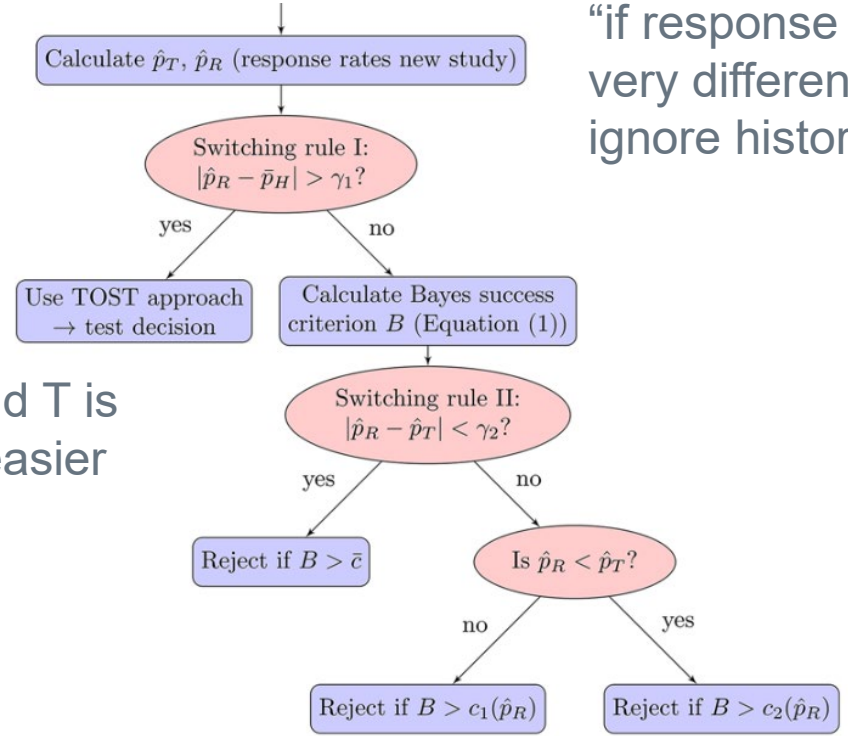




Response rate-dependent critical values

- Response rate-dependent critical values are chosen such that the Type I error rate is controlled in the interval C while the power is maximised under equality of response rates of T, R and the historical data

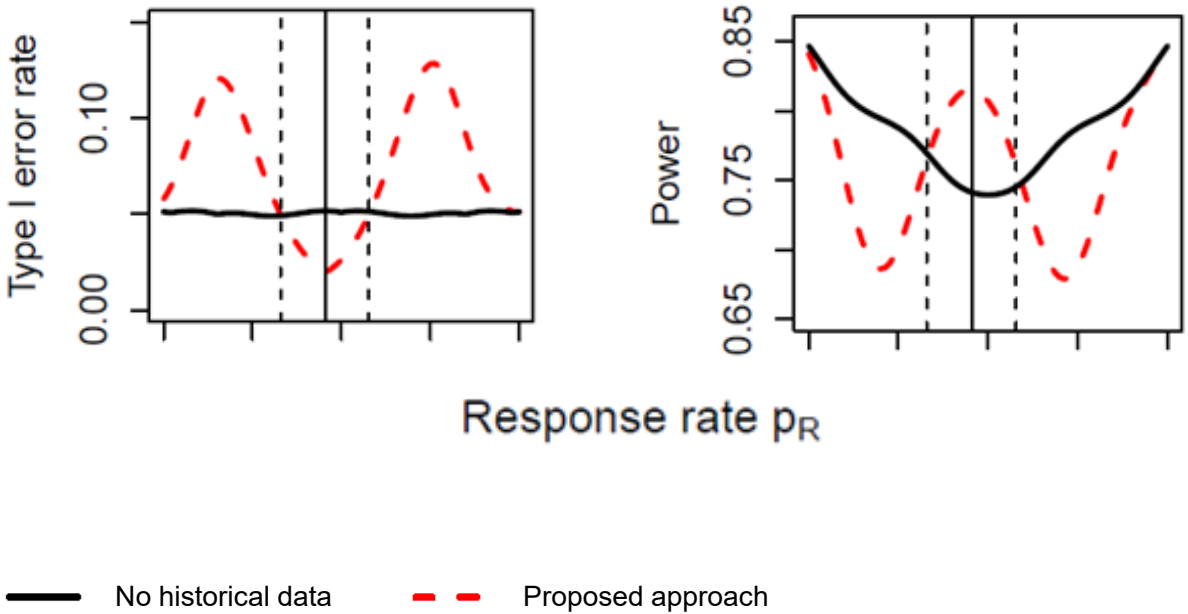
Application of algorithm is simple – just follow a stepwise approach



"if response rate R in new study is very different from mean value prior, ignore historical data"

"if response rate R and T is very similar, make it easier to reject"

Example operating characteristics



Case study: a Phase III study for a biosmilar for adalimumab (Humira)

- (Hypothetical) Phase III study for a proposed biosimilar with the active substance adalimumab (Humira)
- Indication: Psoriasis
- Endpoint: PASI 90 (Psoriasis Area and Severity Index) responder rate at week 16
- Equivalence margin: $\Delta = 0.15$
- Sample size new study: $n = 175$

Case study: a Phase III study for a biosmilar for adalimumab (Humira)

- Historical data is available for five studies with similar treatment regimen and study population

Study	Publication	Indication	Responders/Sample size (%)
1	Menter et al. (2008)	Moderate to severe psoriasis	366/814 (45.0)*
2	Saurat et al. (2008)	Moderate to severe plaque psoriasis	55/108 (51.3)*
3	Thaçi et al. (2010)	Moderate to severe psoriasis	183/364 (50.0)
4	Blauvelt et al. (2017)	Moderate to severe psoriasis	166/334 (49.7)
5	Reich et al. (2017)	Moderate to severe psoriasis	116/248 (46.8)
Total		886/1868 (47.4)	

The shown response rates correspond to PASI 90 at week 16.

*: Response rate was given in the paper and number of responders was calculated using the information in the publication.

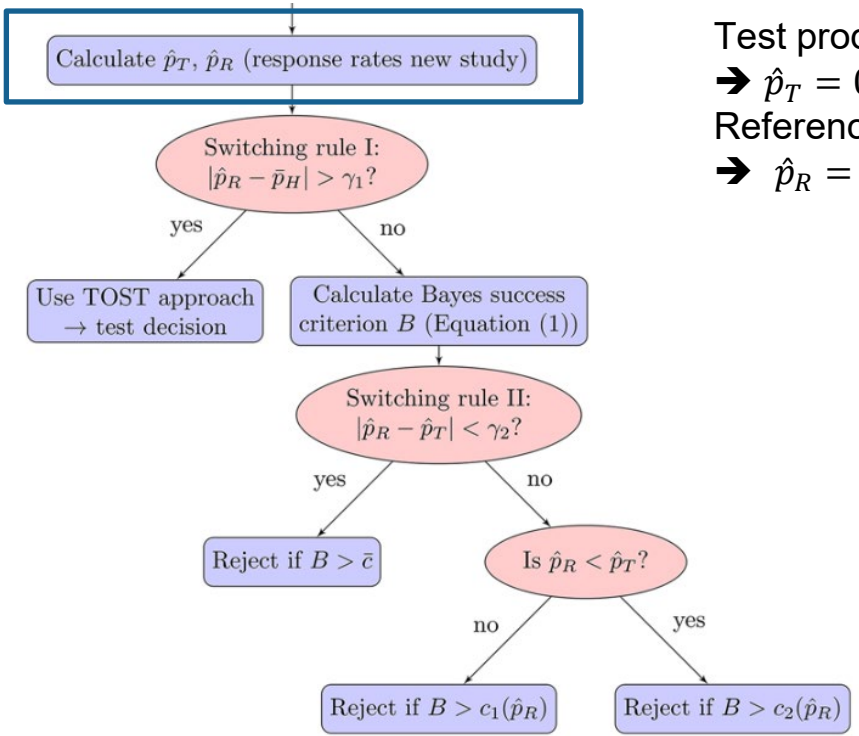
As a first step, we need to choose the study characteristics

- Construct prior for reference product with MAP-approach leads to ESS of 144 and $\bar{p}_H = 0.4813$.
- Control of Type I error in interval:

$$C = [\bar{p}_H - 0.05, \bar{p}_H + 0.05] = [0.43, 0.53]$$

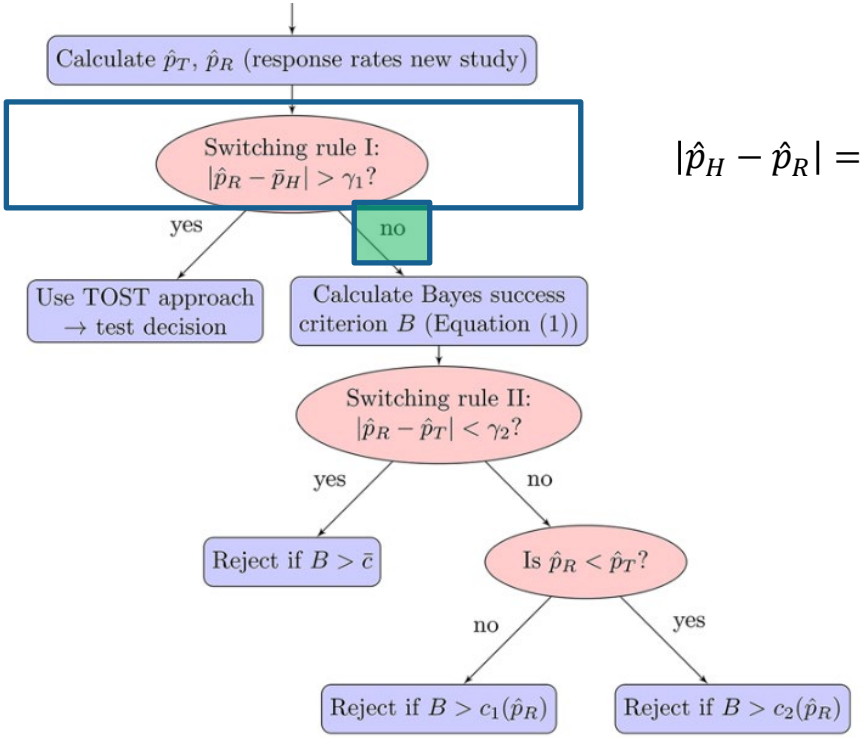
- Tuning parameters are chosen with the proposed algorithm

Application of algorithm to EPAR data from Amgevita (Amgen)



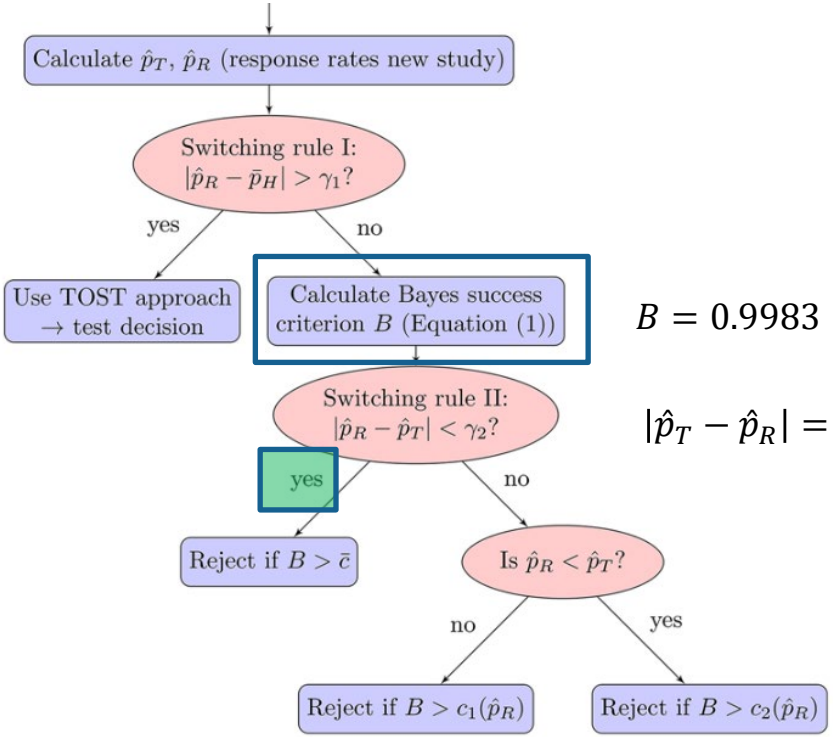
Test product: 81 out of 172 subjects responded
➔ $\hat{p}_T = 0.471$
Reference product: 82 out of 173 subjects responded
➔ $\hat{p}_R = 0.474$

Application of algorithm to EPAR data from Amgevita (Amgen)



$$|\hat{p}_H - \hat{p}_R| = |0.4813 - 0.474| = 0.0073 < 0.0944 = \gamma_1$$

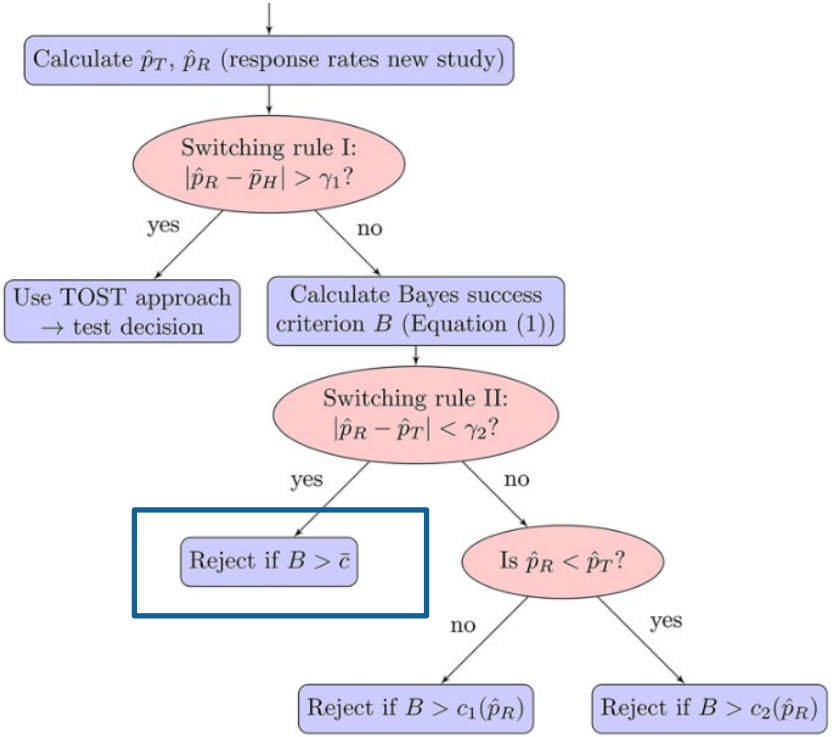
Application of algorithm to EPAR data from Amgevita (Amgen)



$B = 0.9983$

$|\hat{p}_T - \hat{p}_R| = |0.471 - 0.474| = 0.003 < 0.0567 = \gamma_2$

Application of algorithm to EPAR data from Amgevita (Amgen)



$B = 0.9983 > 0.9 = \bar{c}$

We claim equivalence between test and reference!

Discussion

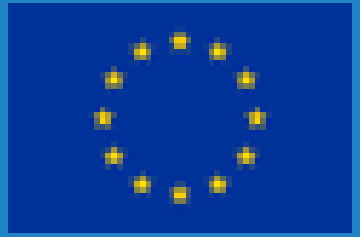
- Proposed approach provides a gain in power in comparison to not using historical data while controlling the Type I error rate in the interval C
- Choice of the interval C dependent on knowledge and confidence in conducting a new study which is similar to the historical studies
- The choice of the response rate-dependent critical values and tuning parameters for the switching rules is computationally very expensive, but not difficult for the user to perform

Mielke, J., Schmidli, H. and Jones, B. (2018): Incorporating historical information in biosimilar trials: challenges and a hybrid Bayesian-frequentist approach. *Biometrical Journal*, 60 (3), 564-582.

Thanks!

Any questions?

Please get in touch:
johanna.mielke@googlemail.com



Bayesian design of biosimilars clinical programs involving multiple therapeutic indications

Matthew A. Psioda
GSK Statistics and Data Science Innovation Hub

September 19, 2022

Background

- Purpose: For a proposed and reference biologic, a biosimilar clinical program is conducted to demonstrate
  - equivalent pharmacokinetics and clinical efficacy,
  - similar safety, and
  - similar immunogenicity.
- The overall objective is to demonstrate that there are no clinically meaningful differences between the two biologics [1].

Background

- One or more trials may be required to eliminate residual uncertainty regarding similarity of clinical efficacy and safety between the proposed and reference biologic.
- These trials are typically designed to provide statistical evidence that the proposed biologic is neither inferior nor superior (in most cases) to the reference biologic based on pre-specified equivalence margins.
- Due to clinically relevant differences in the MoA for a proposed biologic in different disease conditions (hereafter *indications*), a biosimilar program may include clinical trials for multiple indications.

Background

- The biologic rituximab (brand name MabThera/Rituxan) has indications for treatment of Non-Hodgkin's Lymphoma (NHL), chronic lymphocytic leukemia, rheumatoid arthritis (RA), and granulomatosis with polyangiitis and microscopic polyangiitis.

- One biosimilar program [2] included separate clinical trials in RA [3] and follicular lymphoma (FL, a type of NHL) [4] in addition to analytical characterization and non-clinical assessments of the proposed biologic.

- On the basis of these collective data, the product was approved as a biosimilar by the European Medicines Agency (EMA) in 2017 for use in all indications.

Motivating Example I

Two indications to be studied...

- Follicular Lymphoma (FL)
  - difference in proportions endpoint
  - equivalence margin $\delta_1 = 0.10$
  - reference response probability 0.81.
- Rheumatoid Arthritis (RA)
  - mean change from baseline endpoint
  - equivalence margin $\delta_2 = 0.60$
  - reference mean CFB = -2.0 ($\sigma_{\text{CFB}} = 1.4$)

Question: How should we approach biosimilars program design when multiple diseases will be studied?

Motivating Example II

Question: What are the available strategies for design?

1. Design equivalence trials that each independently produce substantial evidence of equivalent efficacy for the respective indication (90% power, $\alpha = 0.05$)

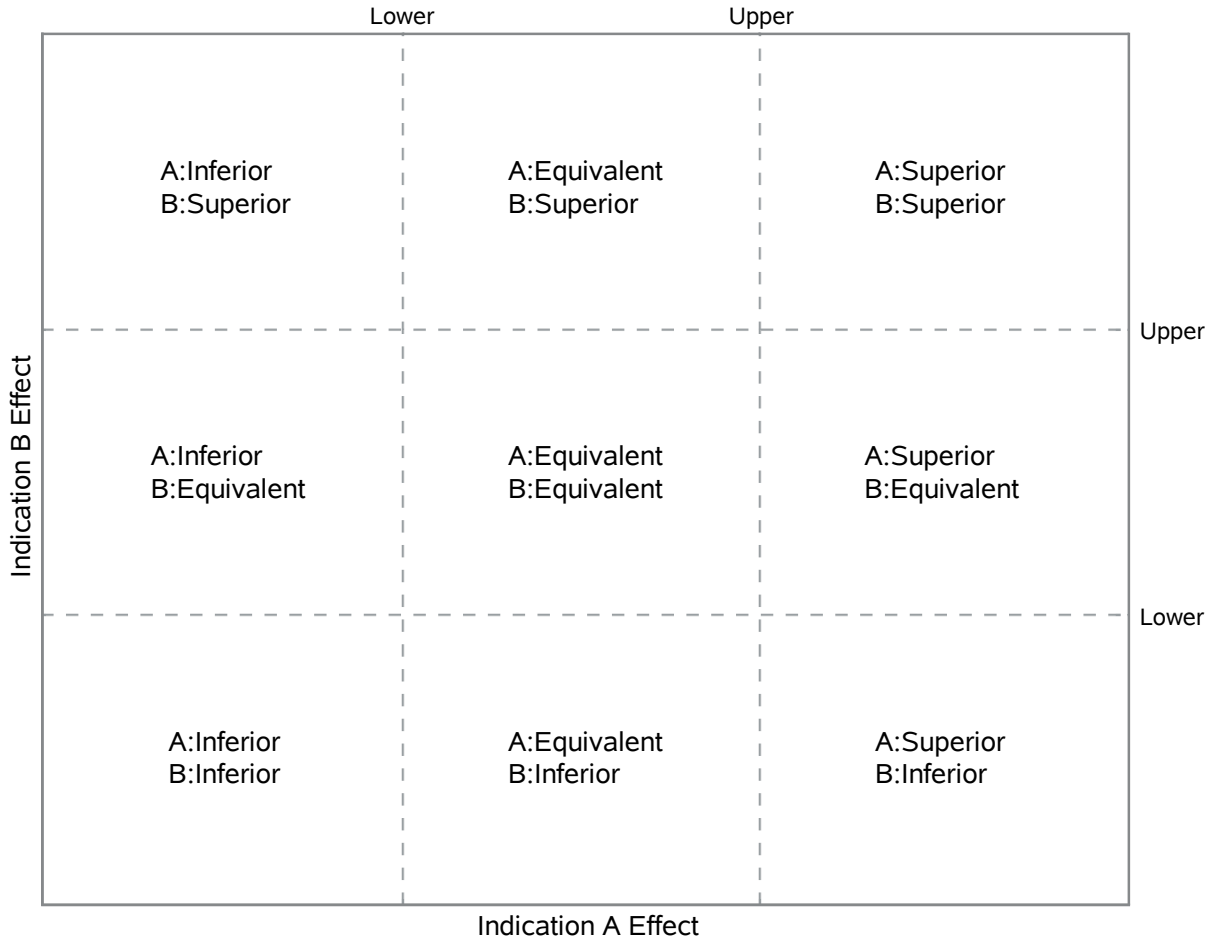
2. Design equivalence trials that incorporate supplemental information in some way in order to meet traditional evidentiary requirements (90% power, $\alpha > 0.05$)

    - Option A: hybrid control arm based on reference program trials or RWE

    - Option B: borrow information across indications within current program

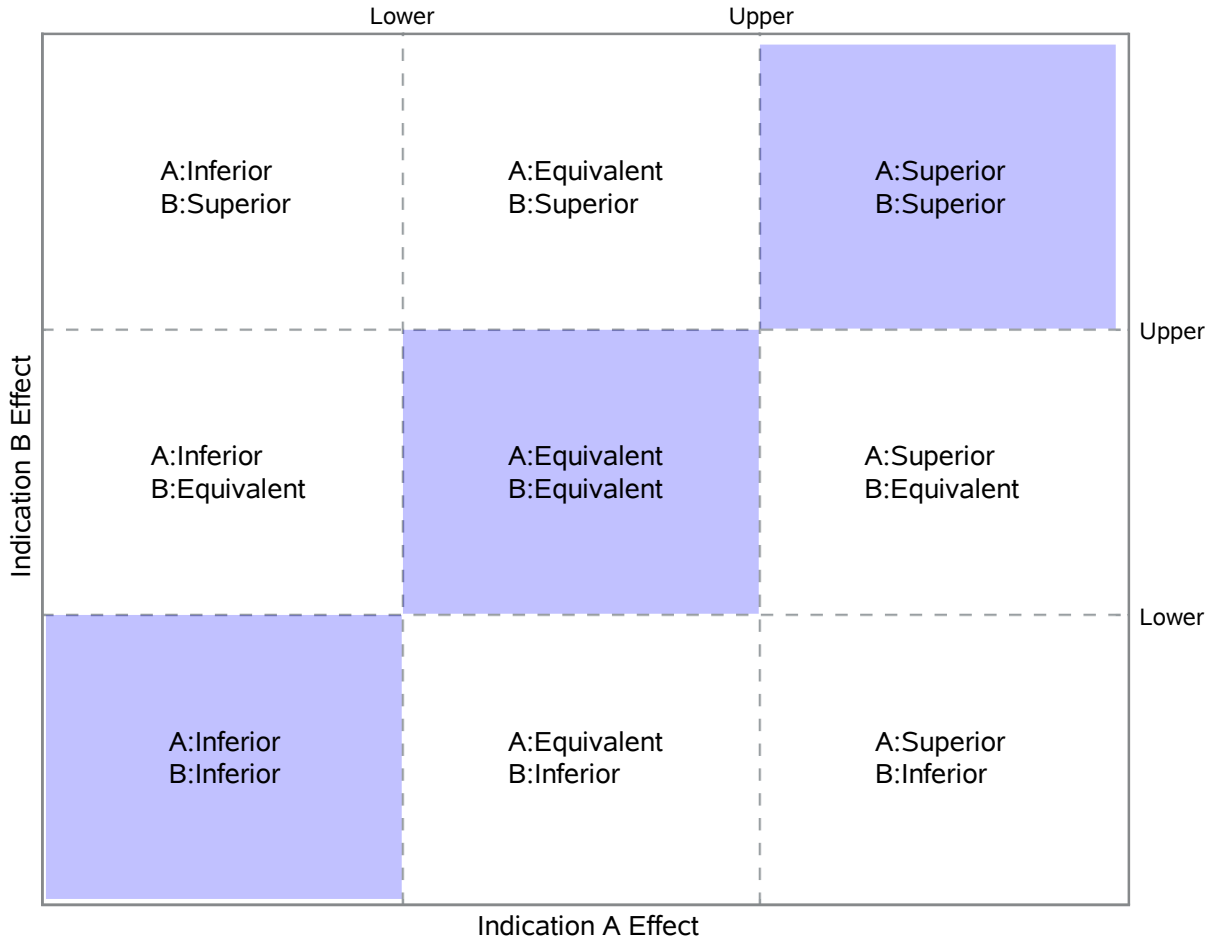
Intuition: Borrowing Information Across Indications

Question: How can we borrow information on indication-specific effects?



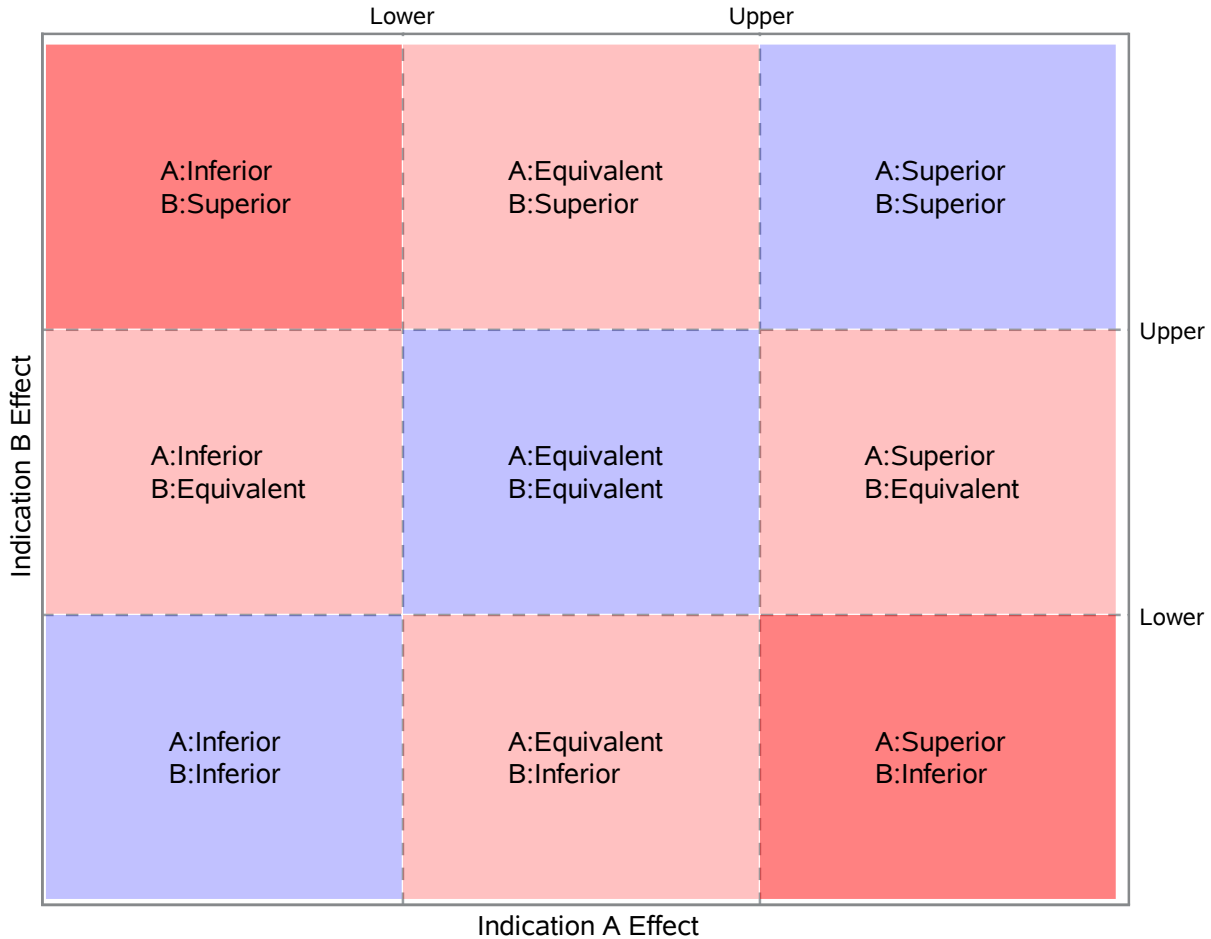
Intuition: Borrowing Information Across Indications

Question: How can we borrow information on indication-specific effects?



Intuition: Borrowing Information Across Indications

Question: How can we borrow information on indication-specific effects?



Methodology – Psioda et al [5]

- Develops an approach for design of a biosimilars program that concurrently investigates several treatment indications with the goal of establishing efficacy equivalence in all of them.
- Proposes a simple, informative prior that supports information borrowing regarding efficacy equivalence across indications.
- Uses Bayesian thinking to argue multiple well-powered equivalence trials may not be sensible in these settings.

Methodology: Indication-Specific Equivalence

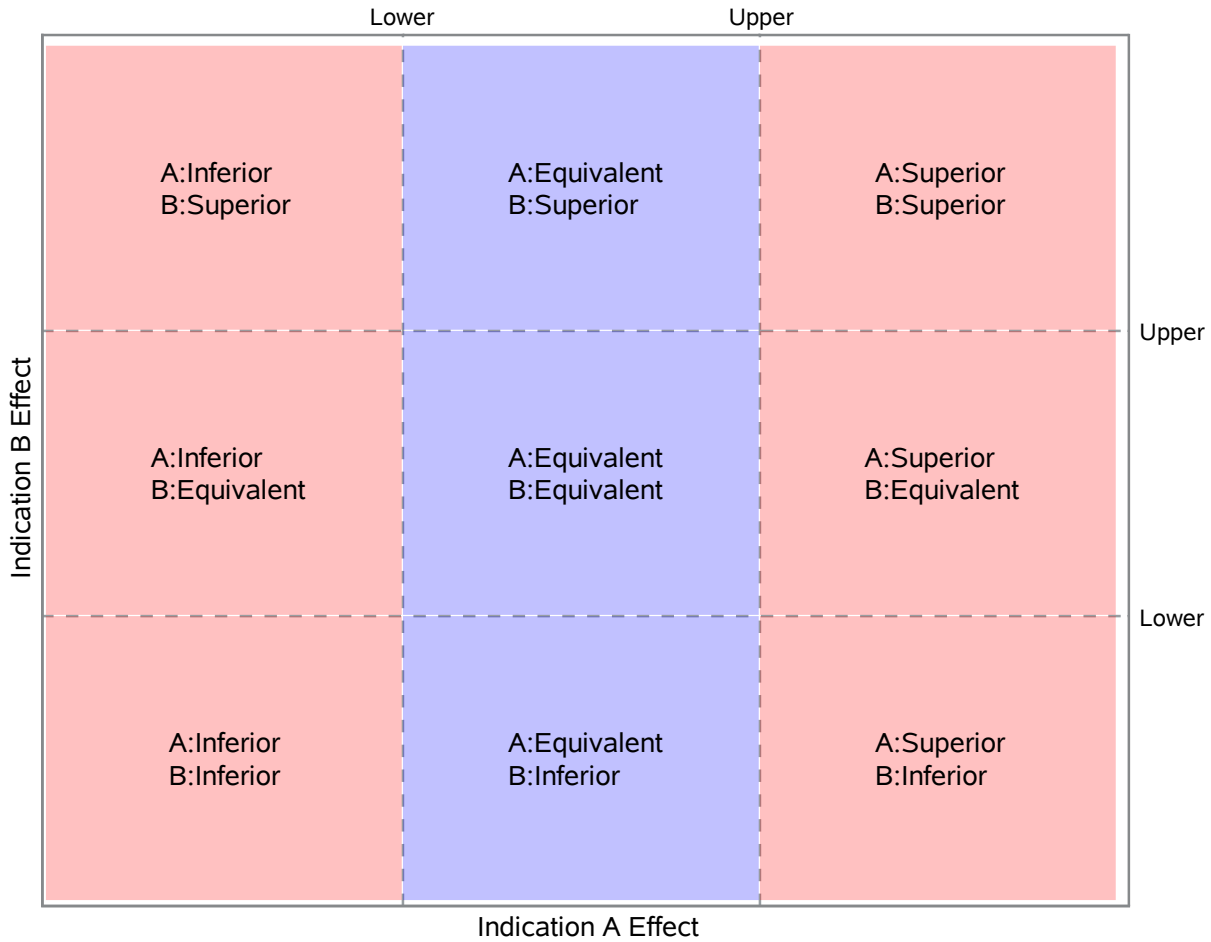
- Hypotheses can be formulated as

$$H_{0j}: |\gamma_j| \geq \delta_j \text{ versus } H_{1j}: |\gamma_j| < \delta_j, \tag{1}$$

where $\delta_j \geq 0$ i s the largest absolute value of γ_j that i s not clinically meaningful (i.e., the equivalence margin).

- One rejects the null hypothesis for indication j when $P(|\gamma_j| < \delta_j | \mathbf{D}) \geq p_{0j}$ where p_{0j} is a prespecified posterior probability critical value.

Methodology: Indication-Specific Equivalence

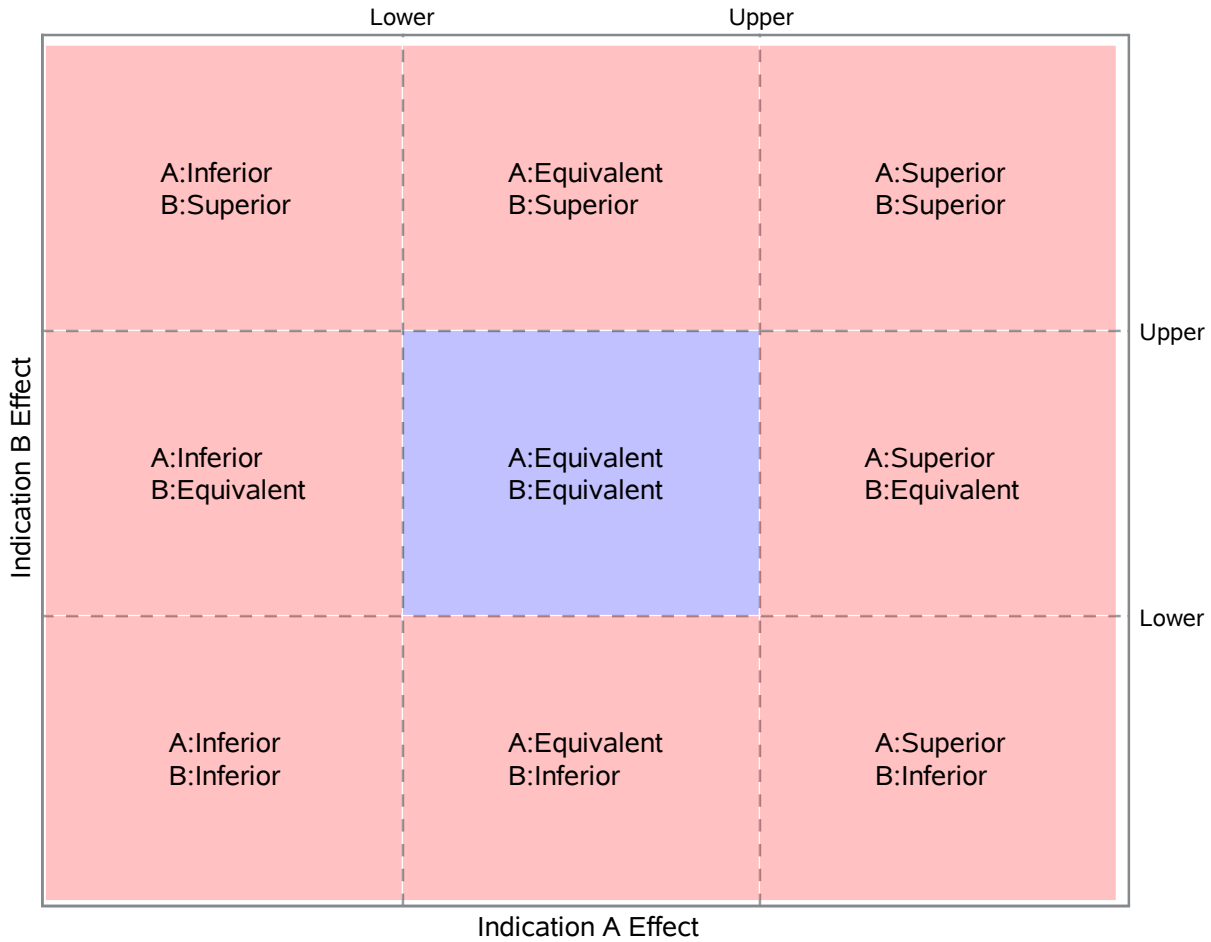


Methodology: Global Equivalence

- Let Θ be the parameter space for $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_J)$ and define the global alternative space as $\Theta_1 = \{\boldsymbol{\gamma} : |\gamma_j| < \delta_j, j = 1, ..., J\}$ with complement Θ_0.

- The global equivalence hypotheses may then be defined generally as $H_0 : \boldsymbol{\gamma} \in \Theta_0$ versus $H_1 : \boldsymbol{\gamma} \in \Theta_1$ with the decision to reject the global null hypothesis occurring when $P(\gamma \in \Theta_1 \mid \mathbf{D}) \geq p_0$.

Methodology: Global Equivalence



Methodology: The Correlated Parameter Prior

- CPP is a multivariate normal prior for γ and can be written as $\gamma \mid \pi_0, \pi_1 \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma})$, with covariance matrix $\mathbf{\Sigma}$ determined indirectly by scalar hyperparameters π_0 and π_1.

- $\pi_0 =$ marginal probability of treatment equivalence for an indication

  - $\pi_0 = P(|\gamma_j| < \delta_j)$ for all j

- $\pi_1 =$ conditional probability of treatment equivalence for an indication given equivalence in another

  - $\pi_1 = P\left(|\gamma_j| < \delta_j \middle| |\gamma_k| < \delta_k\right)$ for all j and k with $j \neq k$

Methodology: On Elicitation of π_0 and π_1

- Possible viewpoint – $\pi_0 = 0.\bar{3}$
  - equivalence is as likely as either inferiority or superiority
- More realistic viewpoint – $\pi_0 = 0.5$
  - equivalence is as likely as non-equivalence
- $\pi_0 \in [0.5, 0.8]$ would be consistent with having pertinent knowledge from earlier stages in the development program and the presence of a non-negligible degree of residual uncertainty.

Methodology: On Elicitation of π_0 and π_1

- To elicit π_1, one must ask the question, "If a trial were conducted in one indication and equivalence proved, how would that modify π_0 for the indications yet to be studied?"

- Percent reduction in residual uncertainty:

$$\Delta_{\text{RU}} = \left(\frac{\pi_1 - \pi_0}{1 - \pi_0} \right) \times 100$$

- Having determined π_0, eliciting Δ_{RU} is equivalent to eliciting π_1.

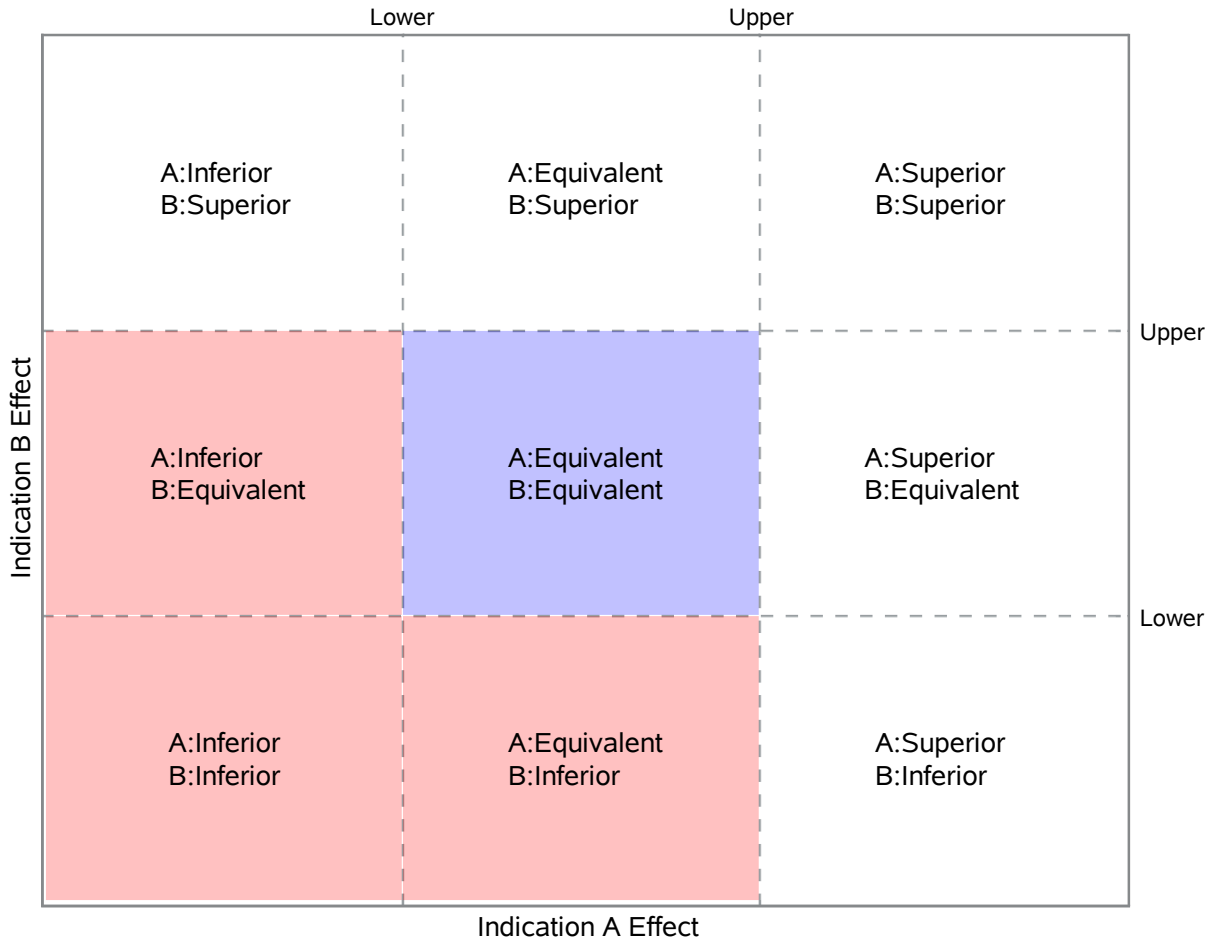
Methodology: Indication-Specific and Global Hypotheses

- Consider a program evaluating the equivalence of a proposed and reference biologic for treatment of FL ($j = 1$) and RA ($j = 2$).

Indication-Specific and Global Hypotheses

Scenario Label	True Indication Hypothesis $j = 1$	$j = 2$	True Global Hypothesis
AA	H_{11}	H_{12}	H_1
AN	H_{11}	H_{02}	H_0
NA	H_{01}	H_{12}	H_0
NN	H_{01}	H_{02}	H_0

Methodology: Indication-Specific Equivalence



Methodology: Sampling Priors over Hypotheses

- In order to evaluate operating characteristics based on each of the 2^J scenarios, one must specify sampling prior distributions for the treatment effects that are consistent with each of them.

- For example, a valid sampling prior distribution for the AN scenario would give non-zero mass to effects that satisfy both $|\gamma_1| < \delta_1$ and $|\gamma_2| \geq \delta_2$.

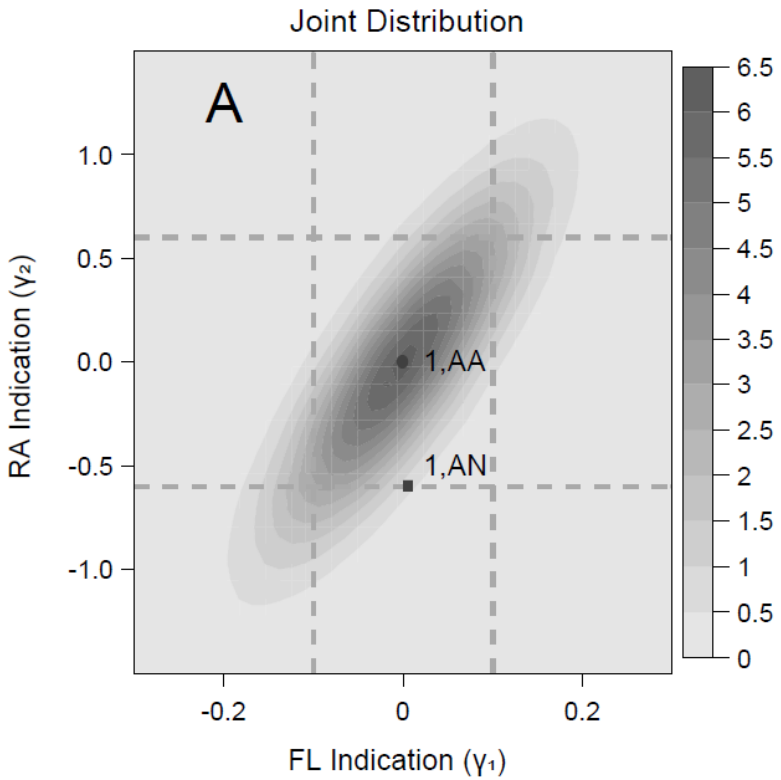
- What is a *sensible* sampling prior?

  - An extreme AN sampling prior: $\pi_{AN}^{(s)}(\boldsymbol{\gamma}) \propto 1(\gamma_1 = 0, \gamma_2 = -\delta_2)$.

  - In what sense is this extreme?

Methodology: Sampling Priors over Hypotheses

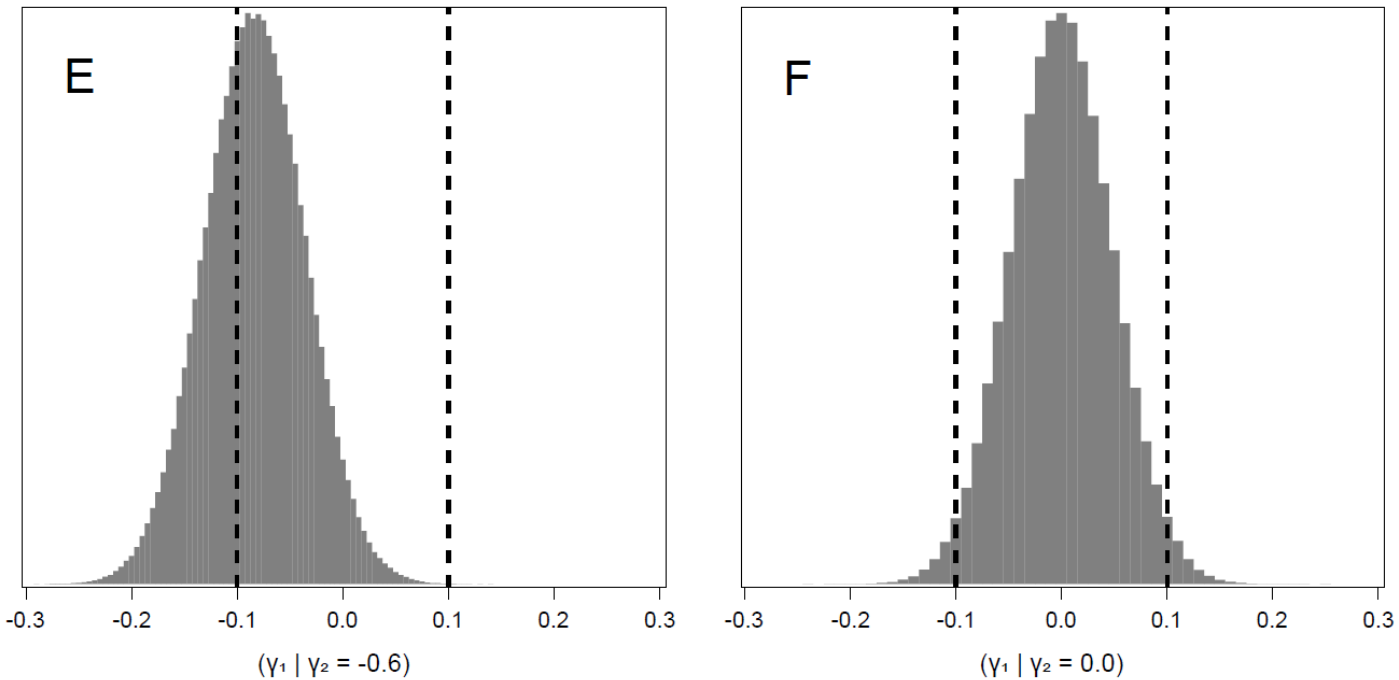
Consider a CPP with a prior probability of treatment efficacy equivalence equal to $\pi_0 = 0.75$, and $\pi_1 = 0.875$ (i.e., $\Delta_{RU} = 50$).



If we condition on $\gamma_2 = -0.60$ (inferiority), what does this imply about γ_1?

Methodology: Sampling Priors over Hypotheses

Panels E and F present, respectively, the prior distributions for the FL difference in proportions conditional on inferiority (panel E) and equivalence (panel F) for the RA indication.



The prior in panel E, further restricted to $\gamma_1 \in (-\delta_1, \delta_1)$ and denoted by $\pi_{AN}^{(s)}(\boldsymbol{\gamma})$, may be useful for considering error rates.

Application: Two Independent Trials

- FL
  - SS requirement – $N_1 = 756$ (based on score test)
  - Power $= 0.90$ ($\alpha = 0.05$)
- RA
  - SS requirement – $N_2 = 238$ (based on exact test)
  - Power $= 0.90$ ($\alpha = 0.05$)
- Thus, if two independent trials were conducted using these samples sizes (when the assumptions are met), the power to demonstrate global equivalence would be approximately 0.81 (i.e., $0.9^2 = 0.81$).

Application: Design Operating Characteristics

Sampling Prior	π_0	$\Delta_{RU} = 0$ π_1	%SS	$r_{FL}^{(s)}$	$r_{RA}^{(s)}$	$r^{(s)}$	$\Delta_{RU} = 50$ π_1	%SS	$r_{FL}^{(s)}$	$r_{RA}^{(s)}$	$r^{(s)}$
$\pi_{1,AA}^{(s)}$	0.50	0.50	1.00	0.92	0.92	0.82	0.75	0.80	0.91	0.91	0.81
	0.67	0.67	0.90	0.91	0.91	0.80	0.84	0.70	0.92	0.92	0.82
	0.75	0.75	0.85	0.91	0.92	0.80	0.88	0.60	0.91	0.91	0.81
$\pi_{1,AN}^{(s)}$	0.50	0.50	1.00	0.92	0.06	0.05	0.75	0.80	0.87	0.11	0.09
	0.67	0.67	0.90	0.91	0.08	0.06	0.84	0.70	0.84	0.18	0.14
	0.75	0.75	0.85	0.91	0.09	0.07	0.88	0.60	0.80	0.24	0.20
$\pi_{2,AN}^{(s)}$	0.50	0.50	1.00	0.44	0.06	0.02	0.75	0.80	0.35	0.07	0.03
	0.67	0.67	0.90	0.45	0.08	0.02	0.84	0.70	0.34	0.10	0.05
	0.75	0.75	0.85	0.47	0.09	0.03	0.88	0.60	0.33	0.12	0.07
$\pi_{1,NN}^{(s)}$	0.50	0.50	1.00	0.06	0.06	0.00	0.75	0.80	0.05	0.05	0.00
	0.67	0.67	0.90	0.08	0.08	0.00	0.84	0.70	0.05	0.05	0.01
	0.75	0.75	0.85	0.10	0.09	0.01	0.88	0.60	0.06	0.07	0.01

%SS=Fraction of standard sample size.

Concluding Remarks I

- The ideas in the associated paper offer a perspective on why multiple (fully powered) equivalence trials may not be ideal for many biosimilars programs.
- Like any information borrowing method, the approach requires choosing quantities that have no single correct value, but perhaps arguments can be made that there are sensible values of π_0 and π_1 that are practical in most cases (at least by the time programs progress to the efficacy trial stage).
- The method does not require use of data from other sources (which we did not have during the project) and which could be challenging to obtain beyond summary statistics.
- If data were available from other sources, it could certainly be incorporated into a more complex design.

Concluding Remarks II

- The strategy is amenable to embedding the set of indications studied in a master protocol which could provide a different type of efficiency, especially with the same IP and reference product being evaluated for each indication.

- Futility stopping could easily be folded into a design that uses a CPP if one indication were expected to enroll much more quickly.

- There is no theoretical reason why sample size reduction would need to be proportional across indications. Reducing sample size in more difficult to enroll indications would seem sensible.

Selected References I

[1] U.S. Food and Drug Administration. *Scientific Considerations in Demonstrating Biosimilarity to a Reference Product: Guidance for Industry*. U.S. Food and Drug Administration, 2015.

[2] Emma D Deeks. CT-P10 (Truxima): A Rituximab Biosimilar. *BioDrugs*, 31(3):275–278, 2017.

[3] Dae Hyun Yoo, Chang-Hee Suh, Seung Cheol Shim, Slawomir Jeka, Francisco Fidencio Cons-Molina, Pawel Hrycaj, Piotr Wiland, Eun Young Lee, Francisco G Medina-Rodriguez, Pavel Shesternya, Sebastiao Radominski, Marina Stanislav, Volodymyr Kovalenko, Dong Hyuk Sheen, Leysan Myasoutova, Mie Jin Lim, Jung-Yoon Choe, Sang Joon Lee, Sung Young Lee, Taek Sang Kwon, and Won Park. A multicentre randomised controlled trial to compare the pharmacokinetics, efficacy and safety of CT-P10 and innovator rituximab in patients with rheumatoid arthritis. *Annals of the Rheumatic Diseases*, 76(3):566 – 570, 2017.

Selected References II

[4] Won Seog Kim, Wojciech Jurczak, Juan-Manuel Sancho, Edvard Javrid, Jin Seok Kim, Jose Angel Hernandez Rivas, Aliaksandr Prokharau, Mariana Vasilica, Rajnish Nagarkar, Dzhelil Osmanov, Christian Buske, Larry Kwak, Michinori Ogura, Sang Joon Lee, Sung Young Lee, Yunju Bae, and Bertrand Coiffier. Double-blind, randomized phase 3 study to compare efficacy and safety of the biosimilar CT-P10 to rituximab combined with CVP therapy in patients with previously untreated advanced-stage follicular lymphoma. *Journal of Clinical Oncology*, 35(15_suppl):7532–7532, 2017.

[5] Matthew A. Psioda, Kuolung Hu, Yang Zhang, Jean Pan, and Joseph G. Ibrahim. Bayesian design of biosimilars clinical programs involving multiple therapeutic indications. *Biometrics*, 76(2):630–642, 2020.

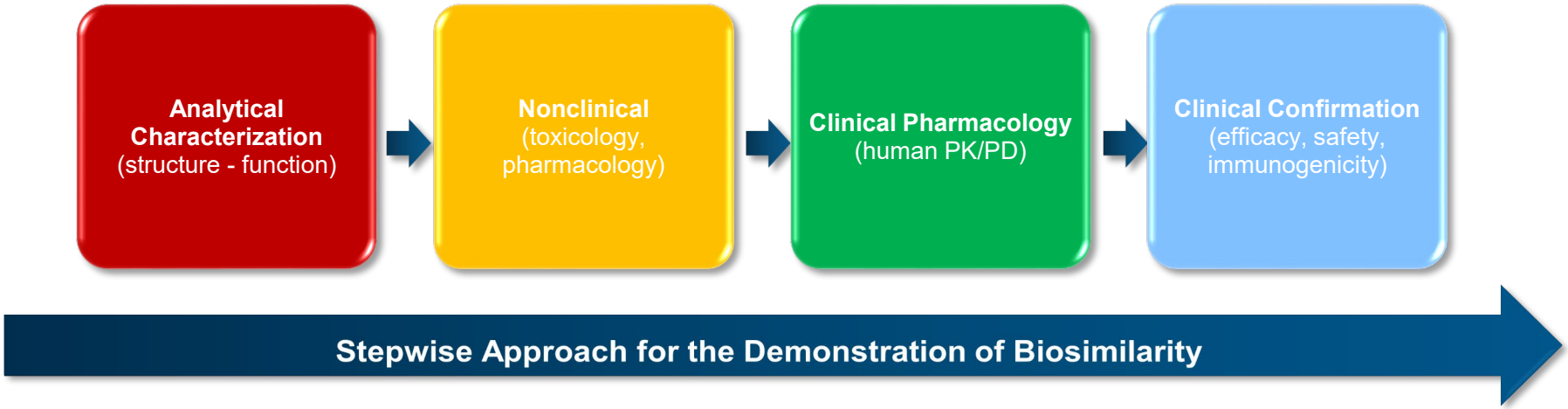
Improving the Power to Establish Clinical Similarity in a Phase 3 Efficacy Trial by Incorporating Prior Evidence of Analytical and Pharmacokinetic Similarity

Danyu Lin, Ph.D.
Dennis Gillings Distinguished Professor of Biostatistics
The University of North Carolina at Chapel Hill

Joint Work with Donglin Zeng, Jean Pan, Kuolung Hu & Eric Chi

BACKGROUND

BIOSIMILAR DEVELOPMENT AND APPROVAL IS BASED ON THE TOTALITY OF EVIDENCE



CBER. *Scientific Considerations in Demonstrating Biosimilarity to a Reference Product. Guidance for Industry*. Silver Spring, MD: FDA. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM291128.pdf. Published April 2015. Accessed September 27, 2015.

STATISTICAL METHODOLOGY TO ASSESS BIOSIMILARITY BASED ON TOTALITY OF THE EVIDENCE

OBJECTIVE

- To incorporate prior similarity evidence (e.g., function and PK) into the comparative clinical study to
  - reduce sample size at the design stage, and
  - enhance study power and estimation precision at the analysis stage

METHODS

NOTATION

	Prior Study	Clinical Study
Data	$Y_{1T} \sim \mathrm{N}(\theta_T, \sigma_T^2)$, $Y_{1R} \sim \mathrm{N}(\theta_R, \sigma_R^2)$	$Y_{3T} \sim B(1, p_T)$, $Y_{3R} \sim B(1, p_R)$
Parameter	$\delta_1 = \theta_T - \theta_R$	$\delta_3 = log(p_T) - log(p_R)$
Pre-specified margin	$L' \leq 0 \leq U'$	$L \leq 0 \leq U$
Null Hypothesis	$H_0^{(1)}: \delta_1 \leq L'$ *or* $\delta_1 \geq U'$	$H_0^{(3)}: \delta_3 \leq L$ or $\delta_3 \geq U$
Alternative Hypothesis	$H_a^{(1)}: L' < \delta_1 < U'$	$H_a^{(3)}: L < \delta_3 < U$
Sample Size/Arm	N_1	N_3
Parameter Estimator	$\hat{\delta}_1 = \hat{\theta}_T - \hat{\theta}_R$	$\hat{\delta}_3 = log(\hat{p}_T) - log(\hat{p}_R)$
Variance Estimator	$\hat{v}_1^2 = \mathrm{var}(\hat{\delta}_1) = \hat{\sigma}_T^2 + \hat{\sigma}_R^2$	$\hat{v}_3^2 = \mathrm{var}(\hat{\delta}_3)$

LEVERAGE SIMILARITY EVIDENCE FROM A PRIOR STUDY WHEN DEMONSTRATING EFFICACY SIMILARITY

Challenges	Potential Solutions
1. θ and p have different scales.	1. Rescale by defining a relative similarity measurement (RSM): $RSM_1 = \frac{\lvert\theta_T - \theta_R\rvert}{\lvert U' - L'\rvert}$, $RSM_3 = \frac{\lvert log(p_T) - log(p_R)\rvert}{\lvert U - L\rvert}$
2. Assumptions about relationship between θ and p are required.	2. Impose a structural assumption: If $RSM_1 < c_1$, then $RSM_3 < \frac{max\{\lvert L\rvert, \lvert U\rvert\}}{\lvert U - L\rvert}$
3. Evidence from θ is empirical, hence random.	3. Allocate the overall type I error rate α: • α_1 $(0 \le \alpha_1 \le \alpha)$ for rejecting $H_0^{(1)*}$: $RSM_1 \ge c_1$ using prior study data • $\alpha_3 = \alpha - \alpha_1$ for rejecting $H_0^{(3)}$ using efficacy study data

REJECTION REGIONS FOR $H_0^{(3)}$

- **Rejection region to control the type I error rate for rejecting $H_0^{(1)*}: RSM_1 \geq c_1$ at α_1, based on the prior data:**

$$\hat{Z}_{\alpha_1} \leq c_1 |U' - L'|$$

where $\hat{Z}_{\alpha_1} = max\left(\left|\hat{\delta}_1 - z_{1-\alpha_1/2}\hat{v}_1\right|,\ \left|\hat{\delta}_1 + z_{1-\alpha_1/2}\hat{v}_1\right|\right)$

- **Rejection region to control the type I error rate for rejecting $H_0^{(3)}$ at $\alpha_3 = \alpha - \alpha_1$, based on the efficacy data:**

$$[\hat{L}_{\alpha_3}, \hat{U}_{\alpha_3}] \subset [L, U]$$

where $[\hat{L}_{\alpha_3}, \hat{U}_{\alpha_3}]$ **is the** $100(1-2\alpha_3)\%$ **CI for** δ_3**.**

- **Overall rejection region for $H_0^{(3)}$:**

$$\mathcal{R} = \boldsymbol{I}(\hat{Z}_{\alpha_1} \leq c_1 |U' - L'| \ \boldsymbol{or}\ [\hat{L}_{\boldsymbol{\alpha_3}}, \hat{U}_{\boldsymbol{\alpha_3}}] \subset [L, U])$$

TYPE I ERROR

- **Overall probability of rejection**

$$\boldsymbol{P}\left(\mathcal{R} = \mathbf{1}|H_0^{(3)}\right) \leq \boldsymbol{P}\left(\hat{Z}_{\alpha_1} \leq c_1|U' - L'|\right) + \boldsymbol{P}([\hat{L}_{\alpha_3}, \hat{U}_{\alpha_3}] \subset [L, U])$$

$$\leq \alpha_3 + \alpha_1$$

$$= \alpha$$

Type I error rate is preserved.

OPTIMAL α_1 TO MAXIMIZE POWER AT DESIGN STAGE

- **Under $H_a^{(3)}$, power is**

$$1 - \mathrm{G}(\alpha_1) = 1 - \mathrm{P}\left(\hat{\mathrm{Z}}_{\alpha_1} > c_1 | \mathrm{U}' - \mathrm{L}' | \mathrm{H}_{\mathrm{a}}^{(3)}\right) \mathrm{P}\left(\mathrm{L} > \hat{\mathrm{L}}_{\alpha_3} \text{ or } \hat{\mathrm{U}}_{\alpha_3} > \mathrm{U}\right)$$

  - $1 - \mathrm{G}(0) =$ **power without using prior similarity evidence**
- **For a given *N*$_3$, steps to obtain the optimal α_1(α_1^{opt}) which maximizes power are:**
  - **Step 1: Obtain $\hat{\delta}_1$ and $\hat{\nu}_1$.**
  - **Step 2: Replace $\hat{p}_T$ and $\hat{p}_R$ in $\mathrm{G}(\alpha_1)$ with p_T and p_R under $H_a^{(3)}$.**
  - **Step 3: Search over a grid of $\alpha_1 \in [0, \alpha]$ to find α_1^{opt} which maximizes $1 - \mathrm{G}(\alpha_1)$.**
- **The above steps can be used to determine *N*$_3$ that will give the desired power, e.g., $1 - \mathrm{G}(\alpha_1^{opt}) \geq 80\%$.**

OPTIMAL α_1 TO REFINE CONFIDENCE INTERVAL AT ANALYSIS STAGE

- **For a pre-specified c_1, steps to obtain a refined CI for δ_3 with the shortest length are:**
  - **Step 1: Obtain $\hat{\delta}_1$ and $\hat{v}_1$.**
  - **Step 2: Obtain $\hat{\delta}_3$ and $\hat{v}_3^2$.**
  - **Step 3: Search over a grid of $\alpha_1 \in [0, \alpha]$ for a CI $[\hat{L}_f, \hat{U}_f]$ for δ_3 with the shortest length. For each α_1,**
    - **Step 3.1: Calculate $\hat{Z}_{\alpha_1}$.**
    - **Step 3.2: Obtain CI $[\hat{L}_f, \hat{U}_f]$:**

$$[\hat{L}_f, \hat{U}_f] = \begin{cases} [\hat{L}_{\alpha-\alpha_1}, \hat{U}_{\alpha-\alpha_1}] & \text{if } \hat{Z}_{\alpha_1} > c_1|U' - L'| \\ [\hat{L}_{\alpha-\alpha_1} \vee L, \hat{U}_{\alpha-\alpha_1} \wedge U] & \text{otherwise} \end{cases}$$

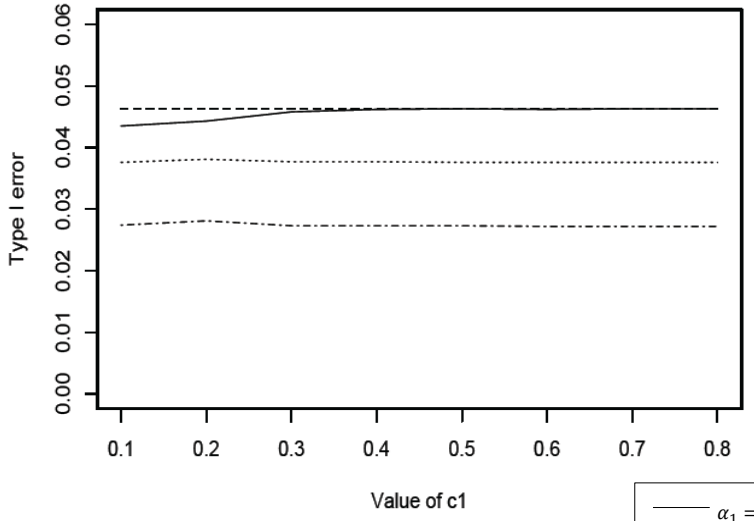
where $a \vee b = max(a, b)$, $a \wedge b = min(a, b)$

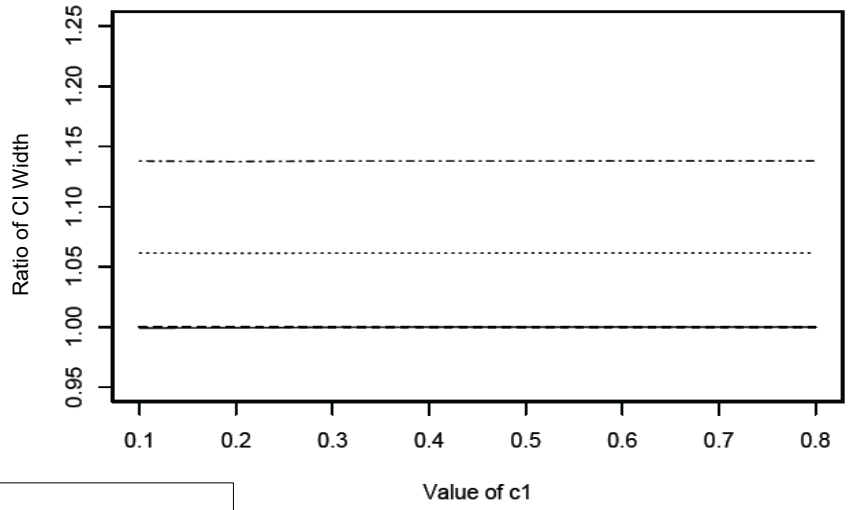
SIMULATION STUDIES

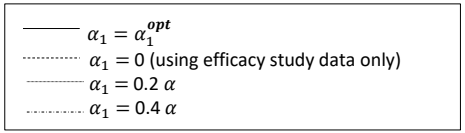
S1: TYPE I ERROR

- **Simulate data under $H_0^{(3)}$**
  - **Prior study:**
    - $\mathrm{N}_1 = 50$
    - $\mathrm{L}' = -\mathrm{U}' = \log(0.8)$
    - $\theta_\mathrm{R} = 5$
    - $\theta_\mathrm{T} = 1.25c_1(\mathrm{U}' - \mathrm{L}') + \theta_\mathrm{R}$ as θ_T must satisfy $\mathrm{H}_0^{(1)*}: \mathrm{RSM}_1 \geq \mathrm{c}_1$.
    - $\sigma_\mathrm{R} = 1.1,\ \sigma_\mathrm{T} = 1$
  - **Efficacy study:**
    - $\mathrm{N}_3 = 300$
    - $\mathrm{L} = -\mathrm{U} = \log(0.75)$
    - $\mathrm{p}_\mathrm{R} = 0.4,\ \mathrm{p}_\mathrm{T} = 0.75\mathrm{p}_\mathrm{R}$
    - Note: $H_a^{(3)}: \mathrm{p}_\mathrm{R} = \mathrm{p}_\mathrm{T} = 0.4$ for obtaining α_1^{opt}

S1: TYPE I ERROR (CONT'D)





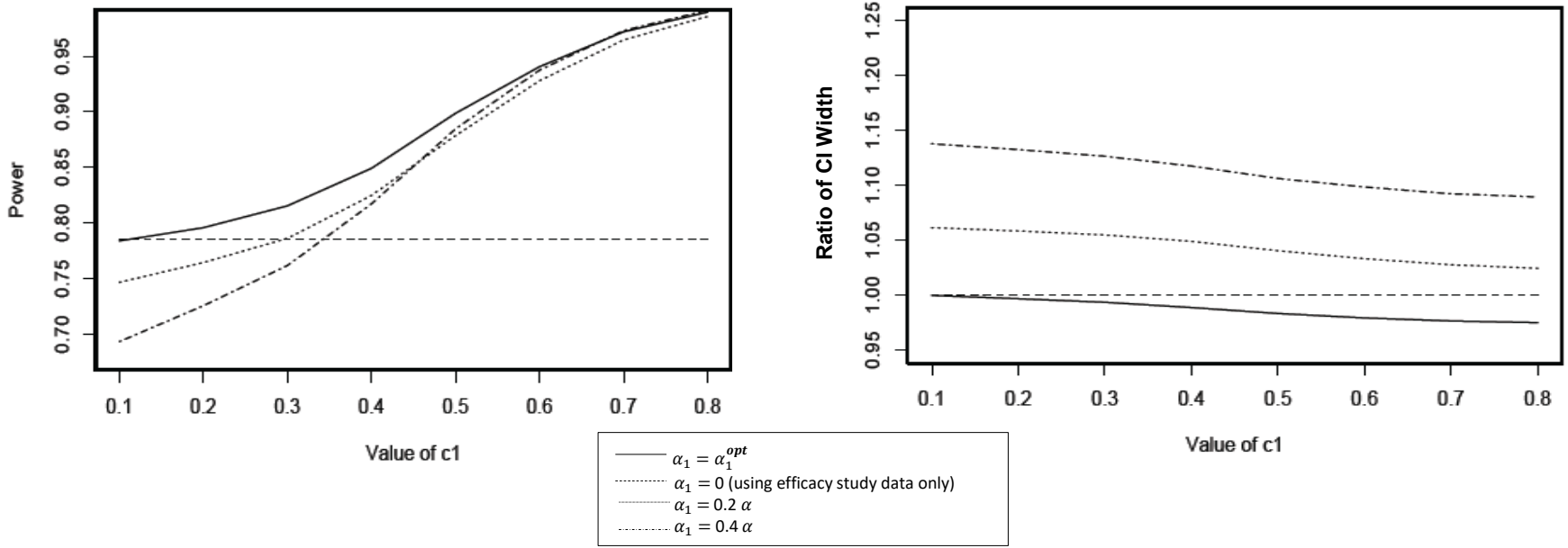


- Overall type I error rate for α_1^{opt} is well controlled at α=0.05 for all c_1.
- Confidence interval width for α_1^{opt} is the shortest among all α_1.

S2: POWER

- **Simulate data under $H_a^{(3)}$**
  - **Similarly to that in S1 except the following:**
    - **Prior study:**
      - $\theta_T = 0.75c_1(U' - L') + \theta_R$ as θ_T must satisfy $RSM_1 < c_1$.
    - **Efficacy study:**
      - $p_R = p_T = 0.4$

S2: POWER (CONT'D)



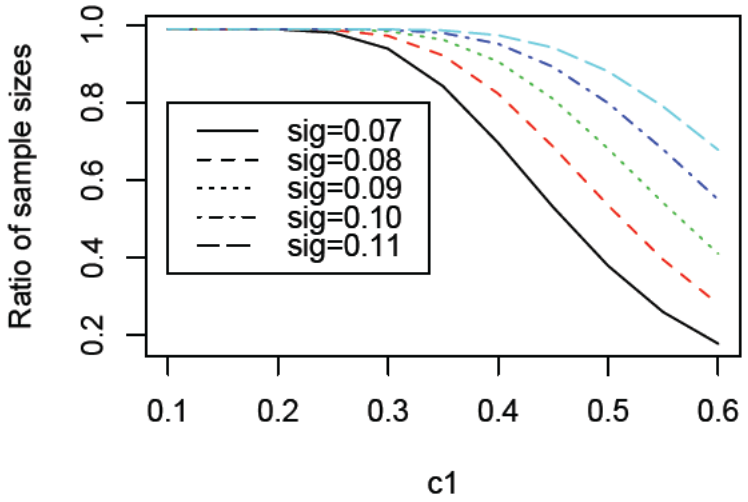
- Power for α_1^{opt} is highest among all α_1, and increases as c_1 increases.
- Confidence interval width for α_1^{opt} is the shortest among all α_1, and decreases as c_1 increases.

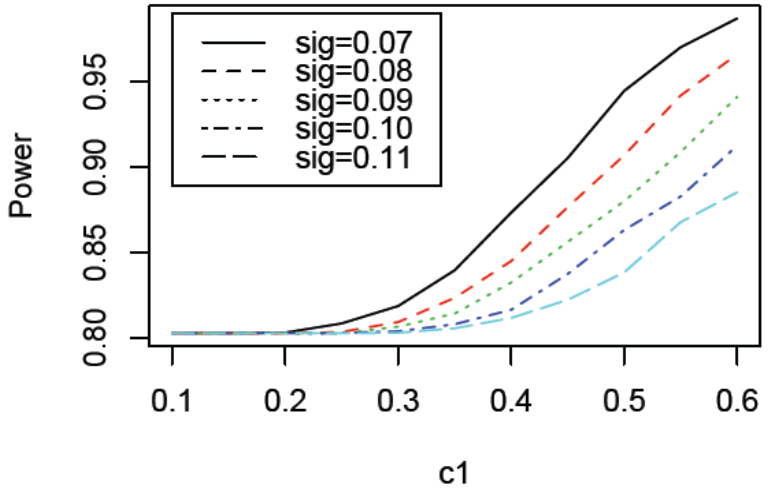
S3: IMPACT OF STRENGTH OF PRIOR SIMILARITY EVIDENCE

- **Simulate data**
  - **Functional activity (TNFα binding affinity):**
    - $N_1 = 10$
    - $\theta_{R1} = 1.08 + \delta,\ \theta_{T1} = 1.08$
    - $\sigma_R = \sigma_T = \sigma$
    - $L_1' = -U_1' = -0.1497$
  - **Efficacy study:**
    - $\delta_3 = p_T - p_R$
    - $L = -U = -0.15$
    - $p_R = p_T = 0.8$
    - Desired power = 80%
- **Smaller variability (σ) or difference (δ) ➔ stronger prior similarity evidence**

S3: IMPACT OF VARIABILITY OF MEASUREMENT

- **Fix δ = 0.04**

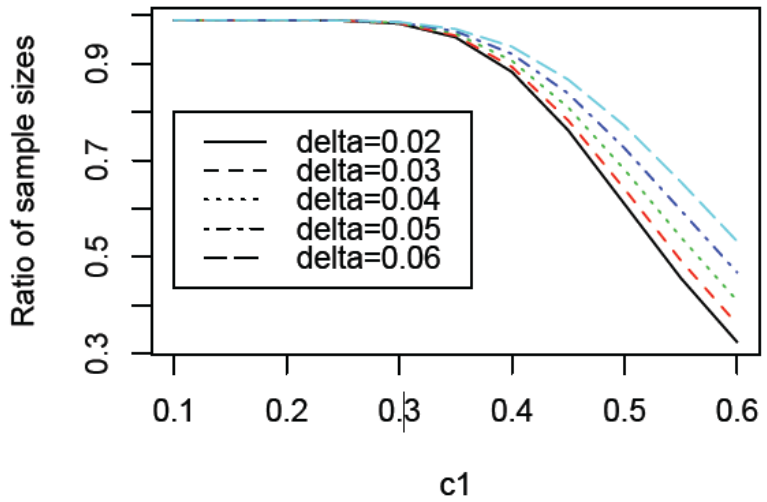


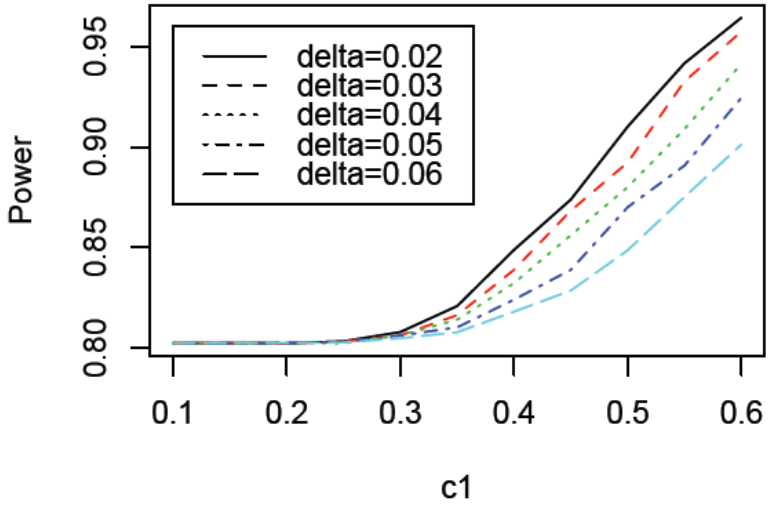


Sample size decreases and power increases as variability decreases.

S3: IMPACT OF DIFFERENCE OF MEASUREMENT

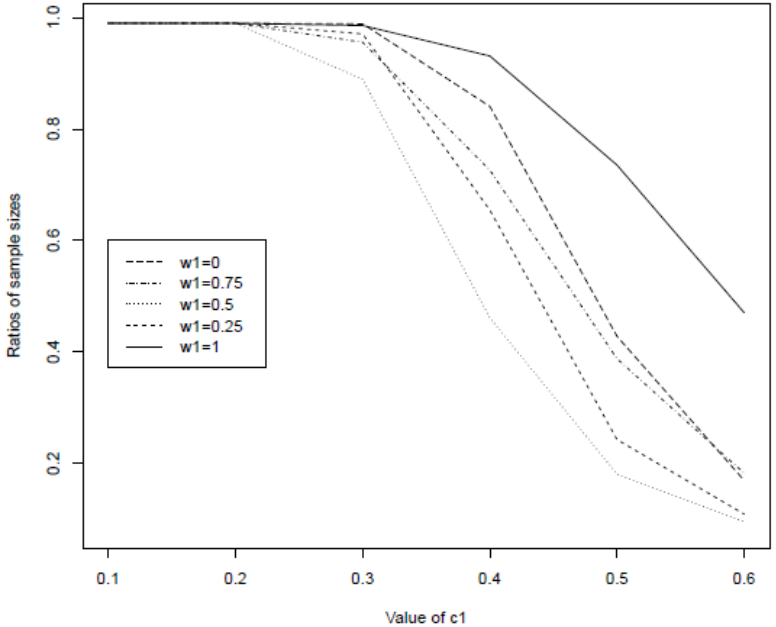
- **Fix** $\sigma = 0.09$

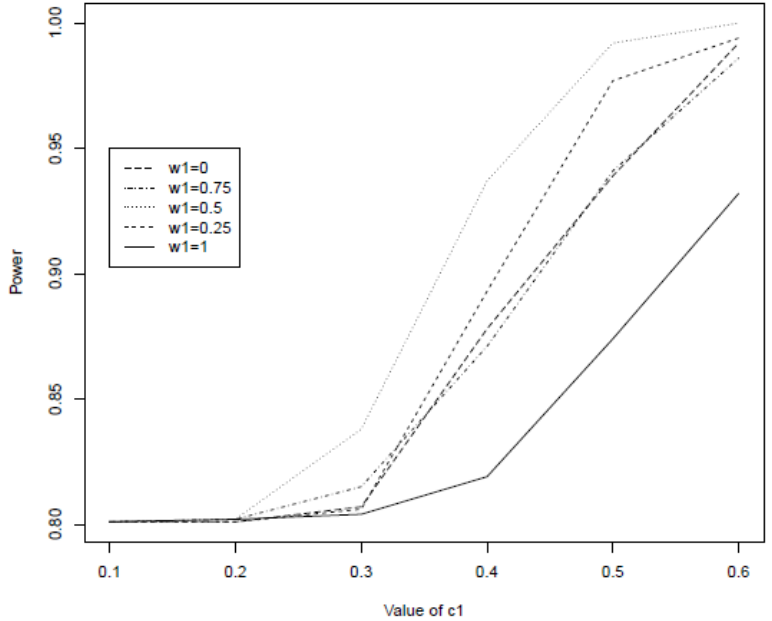




Sample size decreases and power increases as difference decreases.

S4: TWO SOURCES OF PRIOR SIMILARITY EVIDENCE (CONT'D)





Additional efficiency is gained when adding prior PK similarity as an additional source of evidence to prior functional similarity.

DISCUSSION

STRUCTURAL ASSUMPTION

- A key and minimal assumption for the proposed methods
  - Key: Strong prior similarity evidence implies similar efficacy.
  - Minimal: qualitative instead of an actual functional relationship between parameters from previous studies and efficacy study
- A prior qualitative distribution for the two similarity measures from a Bayesian point of view

CHOICE OF c_1

- c_1 governs the amount of information to borrow from prior sources.
- Recommend to determine c_1 using historical evidence from the competing drugs or similar drugs.
- Conduct sensitivity analysis to examine a range of c_1.

REFERENCES

Zeng D, Pan J, Hu K, Chi E, Lin DY: Improving the power to establish clinical similarity in a phase 3 efficacy trial by incorporating prior evidence of analytical and pharmacokinetic similarity. Journal of Biopharmaceutical Statistics, 28: 320-332, 2018.

FDA WORKSHOP: INCREASING THE EFFICIENCY OF BIOSIMILAR DEVELOPMENT PROGRAMS

LUNCH BREAK

We will return to continue Session #2 at 1:00 pm Eastern Time

Discussion

"Bayesian Integration of Data Sources to Inform the Stepwise Approach and Comparative Clinical Study"

Peter F. Thall, PhD
Department of Biostatistics
The University of Texas
MD Anderson Cancer Center

FDA Workshop
Increasing the Efficiency of Biosimilar Development Programs
September 19, 2022

Properties of Bayesian Statistical Models

- The main objects in a statistical model are observable variables $\boldsymbol{Y}$ and unknown parameters $\boldsymbol{\theta}$.

- Elements of $\boldsymbol{Y}$ may be a treatment response indicator, survival time, or the area under a plasma concentration curve (AUC)

- Entries of $\boldsymbol{\theta}$ may be Pr(response), median survival, or mean AUC

- Bayesians consider $\boldsymbol{\theta}$ to be random. A Bayesian model includes a likelihood function, $p(\boldsymbol{Y} \mid \boldsymbol{\theta})$, and a prior, $p(\boldsymbol{\theta})$

- Bayes' Law allows one to compute the posterior $p(\boldsymbol{\theta} \mid \mathit{data})$ which summarizes knowledge about $\boldsymbol{\theta}$ after observing data on $\boldsymbol{Y}$. The posterior is used to make inferences about $\boldsymbol{\theta}$

- During a multi-stage experiment with interim decisions, Bayes' Law may be applied repeatedly, with the posterior at each stage used as the prior for the next stage ("Bayesian Learning ")

Some Advantages of Bayesian Models

- They can account for multiple sources of variability
- They can be used to combine different data sources, while accounting for between-source variability: (1) Multiple clinical trials of the same treatments, for a meta-analysis, (2) historical control data and current control data in a clinical trial (3) observational (non-experimental) data and experimental data
- $p(\theta \mid data)$ can be used to compute the posterior probability of a hypothesis: For response probabilities θ_T for T = new treatment and θ_R for R = reference treatment, $\Pr\{H_1(\theta_T, \theta_R) \mid data\} = \Pr(-.10 < \theta_T - \theta_R < .10 \mid data)$ = posterior probability that T and R are "equivalent"
- They facilitate prediction by computing P(Future observation | past observations). Examples :

Pr(a future patient will survive $\geq$ 5 years | past patient data)
Pr(The Nationals will win the World Series | current MLB data)

1. Incorporate historical data $\mathcal{H}$ on the reference treatment R.
2. Use the posterior degree of similarity between $\theta_{R,\mathcal{H}}$ and $\theta_{R,trial}$ to incorporate $\mathcal{H}$ into tests comparing T to R
3. Assume non-informative priors on new treatment T model parameters θ_T
4. Accommodate different types of endpoints: Binary, ordinal, time-to-event, real valued.
5. For PK variables $Y_{PK,T}$ and $Y_{PK,R}$ with means $\theta_{PK,T}$ and $\theta_{PK,R}$ it makes sense to do the usual interval test of

   $|\theta_{PK,T} - \theta_{PK,R}| < \Delta_{PK}$ versus $|\theta_{PK,T} - \theta_{PK,R}| \geq \Delta_{PK}$

6. For clinical efficacy outcomes $Y_{T,Eff}$ and $Y_{R,Eff}$ with means $\theta_{T,Eff}$ and $\theta_{R,Eff}$, and toxicity outcomes $Y_{T,Tox}$ and $Y_{R,Tox}$ with means $\theta_{T,Tox}$ and $\theta_{R,Tox}$,
it makes sense to do 2 correlated 1-sided tests:

   Efficacy ϵ-Equivalence: $H_{0,Eff}: \theta_{Eff,T} \leq \theta_{Eff,R} - \epsilon$ versus $H_{1,Eff}: \theta_{T,Eff} > \theta_{R,Eff} - \epsilon$ for small fixed slippage $\epsilon > 0$.

   Safety: $H_{0,Tox}: \theta_{T,Tox} > \theta_{R,Tox}$ vs $H_{1,Tox}: \theta_{T,tox} \leq \theta_{R,Tox}$

7. Give Specific Guidelines for
   a. determining numerical prior hyperparameters
   b. specifying fixed Δ (PK equivalence) or ϵ (Eff slippage)
   c. setting limits on Pr(Type I error) and power
   d. determining n based on (a), (b) and (c)

8. A testing scheme should allow equivalence in one prespecified subgroup but not others (subgroup = indication, disease subtype, biomarker +).
9. If a new Bayesian design is both reliable and practical, it may motivate modifying FDA guidelines
10. *Caveat*: "One size fits all" designs usually must be modified, and FDA guidelines should accommodate this.

Comments on the presentation by J. Mielke

Review of The Proposed Methodology

1. For binary outcomes with response probabilities p_R and p_T, a Bayesian alternative to doing "two one-sided tests" (TOST) is proposed.
2. Historical data $\mathcal{H}$ from studies of R are used to construct a Bayesian meta-analytic-predictive (MAP) hierarchical reference prior $f_{\mathcal{H},robust}(p_R)$ given fixed weight w_R.
3. A non-informative prior $f(p_T)$ is assumed.
4. Test $H_0 : |p_R - p_T| \geq \Delta$ vs $H_1 : |p_R - p_T| < \Delta$
5. Evaluate operating characteristics (OCs) for three cases:

$$p_T = p_R + \Delta \quad \text{(Case a, Type I error)}$$
$$p_T = p_R - \Delta \quad \text{(Case b, Type I error)}$$
$$p_T = p_R \quad \text{(Power)}$$

A Major Difficulty

If Pr(Type I error) is controlled at $p_T = p_R + \Delta$ and $p_T = p_R - \Delta$ then using historical data $\mathcal{H}$ on R does not improve the power at $p_T = p_R$

A Proposed Solution

1. Control Type I error rate in an interval $[p_R - .05,\ p_R + .05]$.
2. Switching Rule I: If $|\hat{p}_{R,trial} - \hat{p}_{R,\mathcal{H}}| > \gamma_1$ then switch to the TOST method
3. Switching Rule II: If $|\hat{p}_{R,trial} - \hat{p}_T| < \gamma_2$ then apply the Bayesian test; if not, make different decisions depending on whether $\hat{p}_R < \hat{p}_T$ or $\hat{p}_R > \hat{p}_T$

Problems With the Bayesian MAP Test

1. A total of 7 test cut-offs and tuning parameters must be chosen. It is not clear how this should be done.

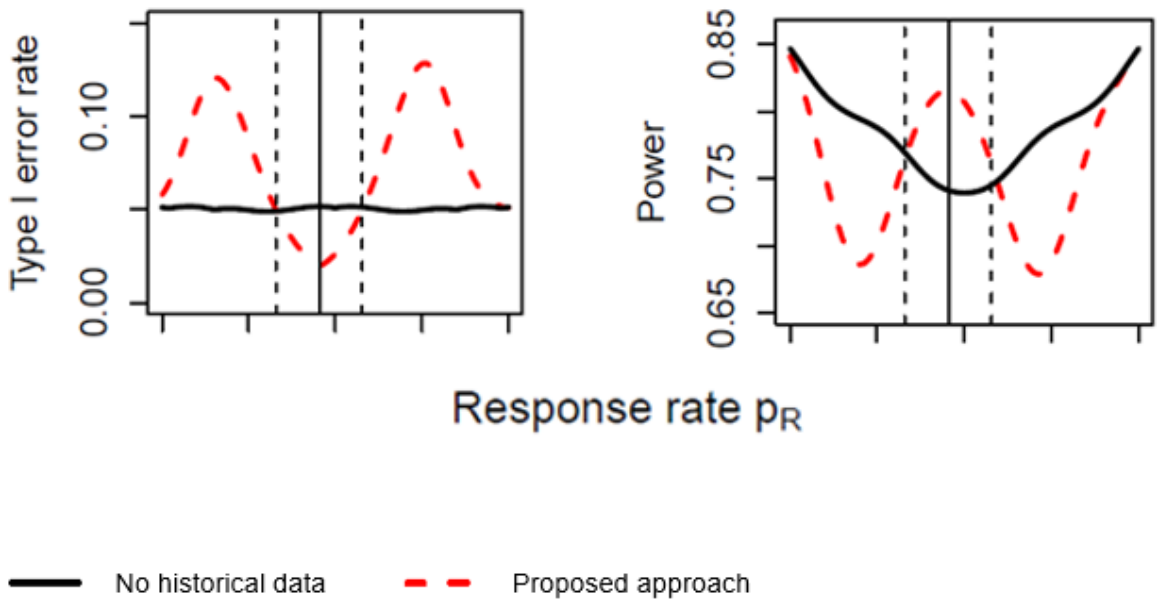
2. For Switching Rule I, instead of switching to TOST, why not downweight $\mathcal{H}$ proportionally to the dissimilarity between $\hat{p}_{R,trial}$ and $\hat{p}_{R,\mathcal{H}}$?

3. Psoriasis Study Application: It is unclear how the 5-study historical data were used. Was a hierarchical model accounting for between-study variability assumed?

4. Type I error of the test depends heavily on the estimate $\hat{p}_T$

5. Attempted Fix: Adjust the test cut-offs in regions where the test is either too liberal or too conservative, depending on whether $\hat{p}_R < \hat{p}_T$ or $\hat{p}_R > \hat{p}_T$. A complex multi-step algorithm for adjusting the test cut-offs is proposed.

6. The Bayesian MAP test has a very irregular power curve that is extremely non-monotone in p_R. Consequently, the test has poor properties and should not be used.



Comments on the presentation by M. Psioda

Review of Proposed Methodology

1. Multiple indications are tested simultaneously. For $J = 2$ indications A and B, 9 rectangular sets as functions of (γ_A, γ_B) = (Indication A effect, Indication B effect) are defined.
2. A multivariate normal correlated parameter prior (CPP) is constructed from hyperparameters representing statements about interval equivalence hypotheses. Denoting

   $\pi_0 = \Pr(|\gamma_j| < \delta_j)$ for all indications j and

   $\pi_1 = \Pr(|\gamma_j| < \delta_j \mid |\gamma_k| < \delta_k)$ for all $j \neq k$,

   the CPP is given by $\boldsymbol{\gamma} \mid \pi_0, \pi_1 \sim N(0, \Sigma)$ with Σ determined by π_0 and π_1, which are elicited.
3. Different endpoints are accommodated by using a Generalized Linear Model with linear term = [Indication j effect] + [treatment effect in indication j] + [baseline covariate effects].

4. n is determined to control Bayesian Type I error and power
5. Different indications may have qualitatively different endpoints. The CCP borrows information on treatment efficacy equivalence across indications.
6. The hypotheses $H_{0,j}: \ |\gamma_j| \geq \delta_j$ (non-equivalence) vs $H_{1,j}: \ |\gamma_j| < \delta_j$ (equivalence) for each indication $j = 1, \cdots, J$ are used to define the global alternative $\Theta_1 = \cap_j H_{1,j} =$ [Equivalence in all J indications] and global null $\Theta_0 = \Theta_1^c$
7. Reject Θ_0 if the posterior probability of global equivalence is large, e.g. $\Pr(\Theta_1 \mid data) \geq .95$.
8. Simulate the trial over an exhaustive set of 2^J scenarios for J indications.

Comments

1. Borrowing strength between indications by assuming the CPP is a very useful idea.
2. Testing Θ_1 = [equivalence in all indications] is very demanding, and obtaining large power seems unrealistic.
3. The global equivalence test of Θ_1 does not allow equivalence in some indications but not others.

   Example: For A=Rheumatoid Arthritis and B=Follicular Lymphoma, if
   Pr(equivalence for A | data) = .10 and
   Pr(equivalence for B | data) = .90
   then, under independence for simplicity,
   Pr(Θ_1 | $data$) = .10 $\times$ .90 = .09
   In this case, the global test for Θ_1 is a very bad idea.

4. 2^J simulation scenarios are OK for $J = 2$ or even 3 indications, but the simulations become impractical for $J \geq 4$
5. For nuisance parameters Ψ, a **sampling prior** is a probability distribution on $\boldsymbol{\theta} = (\gamma, \Psi)$. **Bayesian Type I error for indication** j is defined as the null hypothesis rejection rate

$$r_j(\boldsymbol{\theta}) = E\left\{ I[\textit{Reject } H_{0j}] \mid \boldsymbol{\theta} \right\},$$

with $r_j(\boldsymbol{\theta})$ averaged over a sampling prior on $\boldsymbol{\theta}$ to obtain $r_j^{(s)}$. Bayesian power seems to be computed at a targeted value $\boldsymbol{\theta}^{target}$.

6. The application with 2 indications: For the Bayesian design, size and power are functions of (CCP, sampling prior), n = smallest sample size to get power $\geq .90$ in each indication, power $\geq .80$ globally for the 2 indications, and type I error prob $\leq .10$.
7. As comparators, two conventional .05-.90 tests require $n = 756$ for A and $n = 238$ for B. From the simulations, the fractions of these values provided by the Bayesian test in various cases are computed, 60% to 100% .
8. Specific guidelines for establishing sampling priors and decision cut-offs are needed.
9. Suggestion: Borrow strength between indications by assuming a Bayesian model with correlated $\gamma_1, \cdots, \gamma_J$ but do not do a global test. A realistic approach is separate tests for $J = 2$ or 3 indications, allowing $2^J = 4$ or 8 possible conclusions. This should give higher power and smaller n, compared to assuming independence.

Comments on the presentation by D. Lin

Review of The Proposed Methodology

1. For $\Delta = \theta_T - \theta_R$ test $H_1 : \Delta_L < \Delta < \Delta_U$ vs $H_0 : \Delta \leq \Delta_L$ or $\Delta \geq \Delta_U$
2. **Strategy**: Incorporate prior data to (a) reduce n at the design stage and (b) improve test power and estimation precision at the analysis stage
3. **The setting being addressed**: A prior study has real-valued normally distributed outcomes Y_{1T}, Y_{1R} and $\delta_1 = \theta_T - \theta_R$ and a later clinical trial has binary outcomes Y_{3T}, Y_{3R} and $\delta_3 = logit(p_T) - logit(p_R)$
4. Rescale the prior study effect δ_1 and new study effect δ_3 to obtain **relative similarity measurements** $RSM_1 = \frac{\delta_1}{U' - L'}$ and $RSM_3 = \frac{\delta_3}{U - L}$ on the same scale for comparability
5. Allocate α = overall type I error prob under $H_0^{(3)}$ with α_1 for the prior study test and $\alpha - \alpha_1$ for the study data test

6. Design stage: Use prior data to obtain parameter estimates, and do a grid search to determine sample size that gives a desired power. Messy but practical.
7. Analysis stage: Use prior parameter estimates to obtain a refined CI for δ_3 with the shortest width, $P(L < \delta_3 < U) \geq 1 - 2\alpha$. This looks like a Bayesian posterior credible interval. If it is not, then it should be.
8. Extend this to incorporate multiple sources of prior data. Obtain a weighted relative similarity measurement:

$$RSM_1 = \sum_{k=1}^{K} w_k \left(\frac{\theta_{Tk} - \theta_{Rk}}{U'_k - L'_k} \right)$$

9. Illustration/application in a setting with prior similarity data on TNF-α binding similarity (N_1=10) and PK similarity (N_{R2}=45, N_{T2}=96)

Comments

1. The simulations under $H_0^{(3)}$ show good control of Type I error rate and good power for $N_1 = 50$ and $N_3 = 300$.
2. When multiple sources of prior data are available, the parameters $\{(\theta_{kR}, \theta_{kT}), k = 1, \cdots, K\}$ should have a hierarchical prior that induces correlation among them, if they are exchangeable.
3. The methodology is very detailed with lots of parameters $\Longrightarrow$ OK, but step-by-step guidelines are needed.
4. Use Bayesian posterior credible intervals $[L,\ U]$ defined by $\Pr(L < \theta < U \mid data) = .95$, rather than confidence intervals.
5. Since c_1 governs the amount of prior evidence to borrow, it should be determined adaptively during the trial using $p(\theta_{R,\mathcal{H}}, \theta_{R,trial} \mid data)$ with greater similarity between $\theta_{R,\mathcal{H}}$ and $\theta_{R,trial}$ determining a larger weight for $\theta_{R,\mathcal{H}}$ automatically.

INNOVATIVE STATISTICAL METHODS FOR INTEGRATION OF DATA SOURCES INFORMING BIOSIMILAR COMPARATIVE CLINICAL STUDIES

PANEL DISCUSSION

FDA WORKSHOP: INCREASING THE EFFICIENCY OF BIOSIMILAR DEVELOPMENT PROGRAMS

BREAK

We will return and start Session #3 at 2:00 pm Eastern Time

FDA Workshop: Increasing the Efficiency of Biosimilar Development Programs

BIOSIMILAR COMPARATIVE CLINICAL ENDPOINT STUDY DESIGN: CHOICES TO OPTIMIZE EFFICIENCY

Context Matters: The Purpose of Comparative Clinical Endpoint Studies (CCES) in Biosimilar Development Programs

Carol Kim, Pharm.D.
Scientific Review Staff,
Office of Therapeutic Biologics and Biosimilars (OTBB)

9/19/2022

Disclaimer

- The opinions and information in this presentation are those of this presenter, and do not represent the views and/or policies of the U.S. Food and Drug Administration.

Three Questions

- What is the purpose of comparative clinical endpoint studies (CCES) in biosimilar development programs?
- How are CCES different than comparative efficacy / active control trials intended to support efficacy conclusions?
- In light of their supportive rather than pivotal role, how can CCES be designed more efficiently?

Biosimilar Definition

- Biosimilar or Biosimilarity means that
  - the (proposed) biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components;

  AND there are

  -no clinically meaningful differences between the (proposed) biological product and the reference product in terms of safety, purity, and potency of the product (~safety and efficacy)

  ➢ **Key point:** Clinical studies may be helpful to provide context regarding the "clinical meaningfulness" of differences in other results in a biosimilar development program but also may not be necessary if the other results do not create questions or uncertainty

Different Goals for "Stand-alone" vs Biosimilar Development

"Stand-alone": 351(a) BLA
Goal: To establish *de novo* safety and efficacy of a new product

"Abbreviated": 351(k) BLA
Goal: To demonstrate biosimilarity (or interchangeability) to a reference product based on comparative assessments

Comparative Analytical Assessment is the Foundation

- Compare multiple physicochemical and biological attributes of each product
  - Analytical studies are generally more **sensitive** than clinical studies in detecting differences between products, should differences exist
  - A biosimilar product with **highly similar structure and function** to the reference product should behave like the reference product
- Analyze multiple lots of the reference product and proposed biosimilar for product quality attributes, including:
  - Primary amino acid sequence
  - Higher order structure (protein folding)
  - Post-translational modifications (glycosylation, etc.)
  - Heterogeneity (charge, size, aggregates, etc.)
  - Biological activity - evaluation of attributes that affect the known MoAs

Role of Clinical Studies

- As a scientific matter, FDA expects an adequate clinical PK, and PD if relevant, comparison between the proposed biosimilar product and reference product and a clinical immunogenicity assessment
- Additional clinical studies are not considered “pivotal” in the way Phase 3 clinical trials are for standalone development
- Add to the totality-of-the-evidence that supports a demonstration of biosimilarity

“Totality of Evidence”: Scenarios

Conclusion	Product A	Product B	Product C	Product D	Product E
CAA	**Not highly similar**	Highly similar; minimal or no residual uncertainty	Highly similar, minimal or no residual uncertainty	Highly similar with residual uncertainty noted	Not clear if highly similar: residual uncertainty noted
PK bioeq	Pass or fail	Pass	Pass	Pass	Pass or fail
CCES	“Pass” i.e., met equivalence margin & descriptively similar in safety and efficacy	“Pass” i.e., met equivalence margin & descriptively similar in safety and efficacy	“Fail” i.e., didn’t meet equivalence margin but was descriptively similar in safety and efficacy	“Pass” i.e., met equivalence margin & descriptively similar in safety and efficacy	“Fail” i.e., didn’t meet equivalence margin or difference in descriptive safety profile noted
Totality of evidence	**Product A is not biosimilar**	Product B is biosimilar	Product C is biosimilar?	Product D is biosimilar	Data is insufficient to demonstrate that Product E is biosimilar
Implication	PK and CCES results are not sufficient to overcome issues with analytical similarity	CCES supportive but may have been an unnecessary cost if minimal analytical differences	Should potential flaws with CCES take precedence over totality of data in a biosimilar program?	CCES supports the other data in the development program to facilitate “no clinically meaningful differences” determ.	CCES aligns with the other data in the development program suggesting clinically meaningful differences cannot be ruled out

Not for "pivotal" Role but Burdensome

New Drug/Biologics

- A study for new efficacy safety determination
- Substantial evidence for effectiveness demonstrated
- Two different products are compared (e.g., comparator arm, placebo arm)
- PK data+

Why use same statistical approach?

Biosimilars

- A study for detecting the difference between products to conclude no clinical concern with respect to CAA
- ✓ Equivalent margin depends on Treatment effect size of the reference product
- ✓ Two highly similar products are being compared
- ✓ PK similarity
- Large SS~ >2-3 X pivotal

Generics

- A study for determining Bioequivalence
- Difference between products falls within the standardized margin that is considered acceptable
- Two same active ingredients are being compared
- PK data not available in general
- SS~ <2 X pivotal (placebo arm included)

Estimation of treatment effect size

Power	ORR	Sample size	Sample size after accounting for 10% dropout
Equivalence margin (0.73, 1.38)			
0.80	0.38	574	638
0.85	0.38	636	707
0.90	0.38	724	804
0.80	0.40	528	587
0.85	0.40	586	651
0.90	0.40	666	740
Equivalence margin (0.75, 1.34)			
0.80	0.38	688	764
0.85	0.38	764	849
0.90	0.38	870	967
0.80	0.40	634	704
0.85	0.40	704	782
0.90	0.40	800	889

Table1. sample size calculations under different scenarios

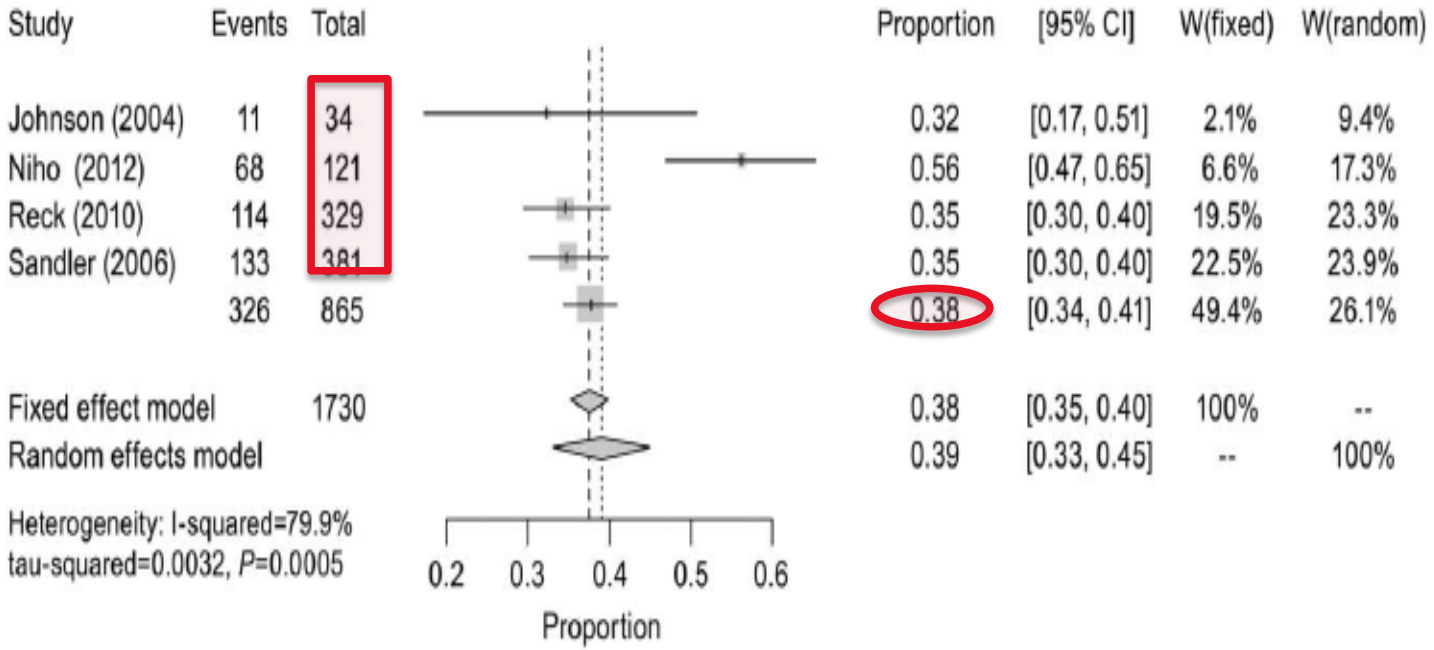


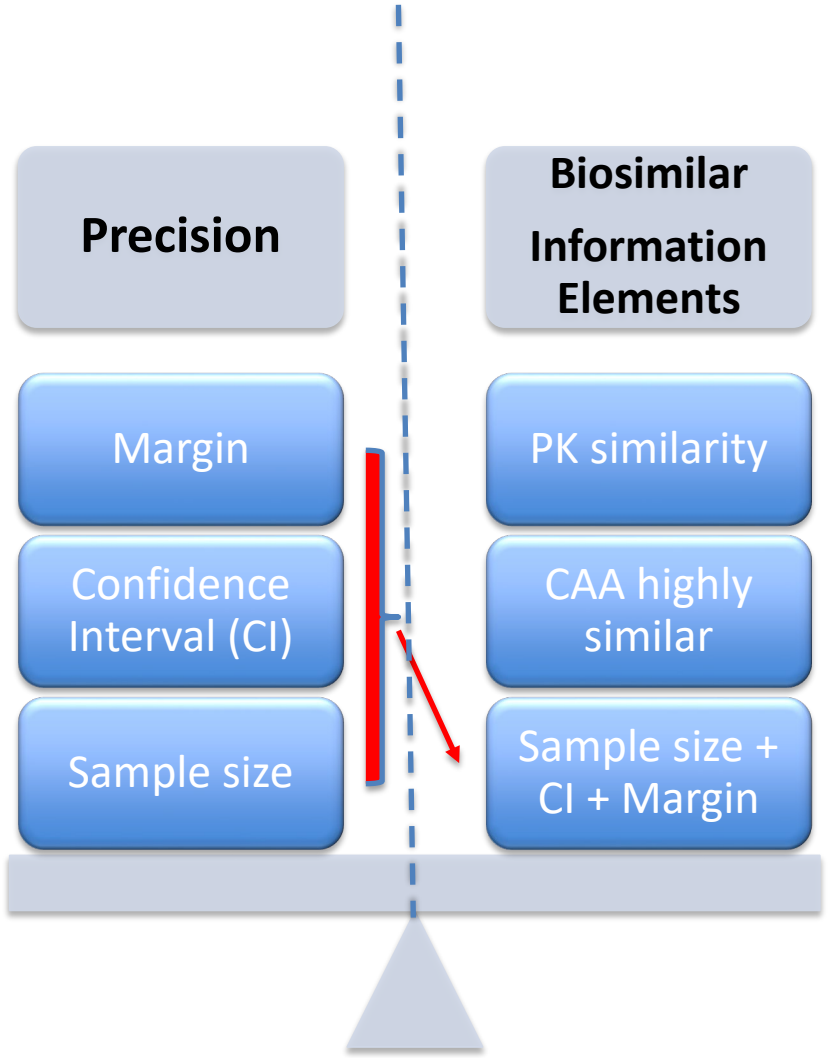
Figure 1. Treatment effect calculation for ORR (with 95% CI) for bevacizumab plus chemotherapy using historic data

Isakov L., Jin B, et al. "Statistical Primer on Biosimilar Clinical Development". Am J Ther, 2016, 23(6)

Statistical Consideration

New Drug

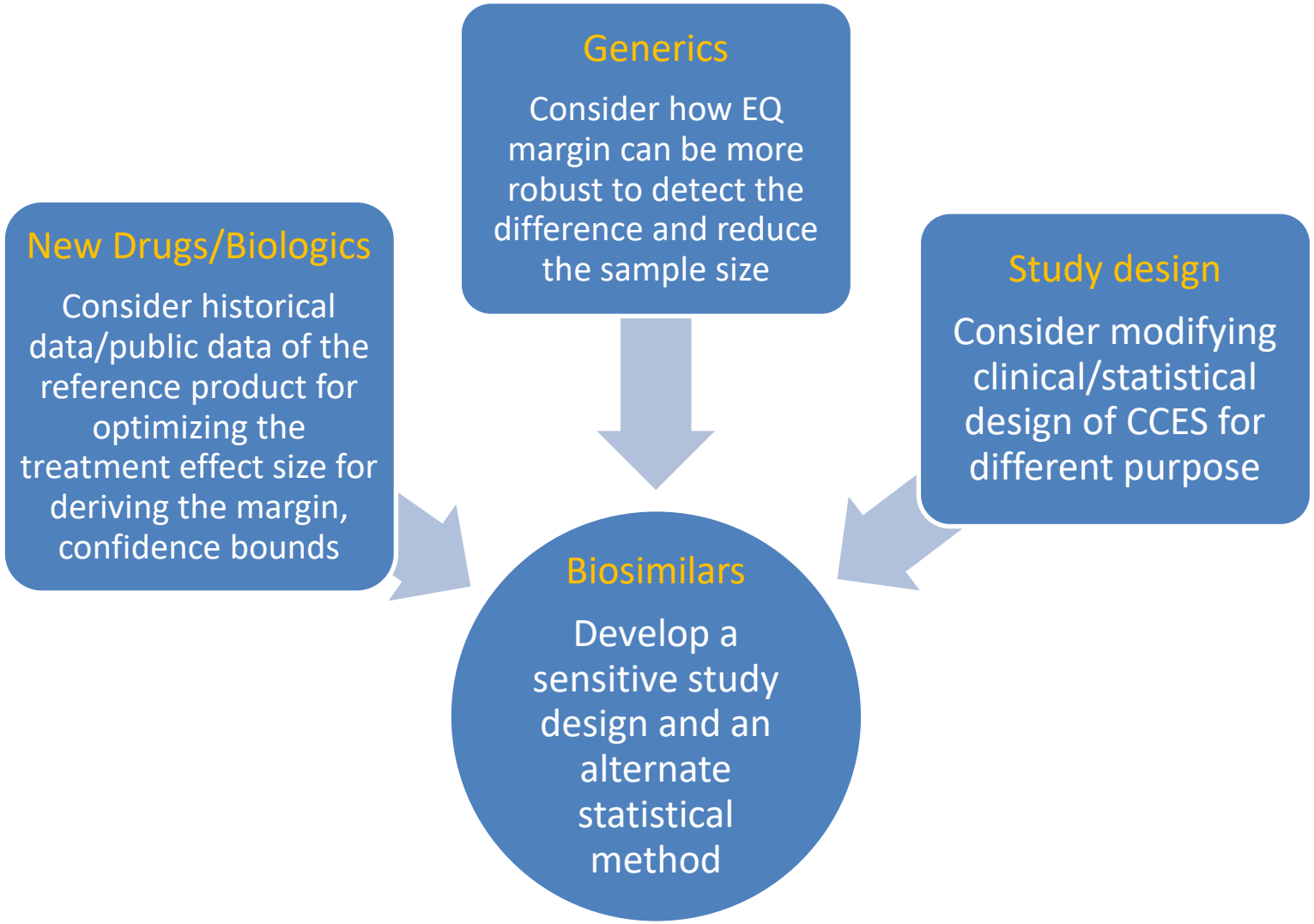
Biosimilar



Substantial Evidence of Effectiveness

No Clinically meaningful Difference Between Reference and Biosimilar

What else can be considered?



Summary

- ✓ CCES is not a confirmatory study or re-establishing safety or efficacy between products.
- ✓ The CCES should be used to support resolving the extent of the residual uncertainty noted in a comparative analytical assessment (CAA) to ensure that the difference observed in CAA is not clinically meaningful.
- ✓ In light of the different purpose, a sensitive study design and other alternate statistical approaches should be considered to reduce the burden of a large sample size.


FDA
U.S. FOOD & DRUG
ADMINISTRATION

Selecting an Endpoint

FDA Workshop: Increasing the Efficiency of Biosimilar Development Programs

Session 3

Kathleen Fritsch, PhD

Office of Biostatistics/Division of Biometrics III

September 19, 2022

How to Choose an Endpoint for a Comparative Clinical Study

- Guidance recommendations – Endpoints should be:
  - Relevant to clinical outcomes
  - Sensitive to detect clinically meaningful differences
- Other considerations:
  - Public availability of estimates to select similarity margins and calculate sample sizes
  - Designs of other programs with same reference product or patient population may be informative

Differences in Rationale for Selecting Endpoints

Phase 3 Trial for a New Drug or Biologic	Comparative Clinical Study for a Biosimilar
Clinically meaningful endpoint(s)	Clinically relevant endpoint
Comprehensive endpoint(s) (capture most important disease elements)	Focused endpoint (may assess more limited disease elements)
Endpoint(s) that can be effectively communicated in labeling	Endpoint that is sensitive to differences (lower variability)

Points to Consider for Selecting a Sensitive Endpoint

- Continuous vs. binary endpoints (generally higher power)
- Scales/instruments with lower variability
- If product has more than one indication, which indication may lead to most efficient design

Reference

Guidance for Industry: *Scientific Considerations in Demonstrating Biosimilarity to a Reference Product (April 2015)*

https://www.fda.gov/regulatory-information/search-fda-guidance-documents/scientific-considerations-demonstrating-biosimilarity-reference-product


FDA
U.S. FOOD & DRUG
ADMINISTRATION

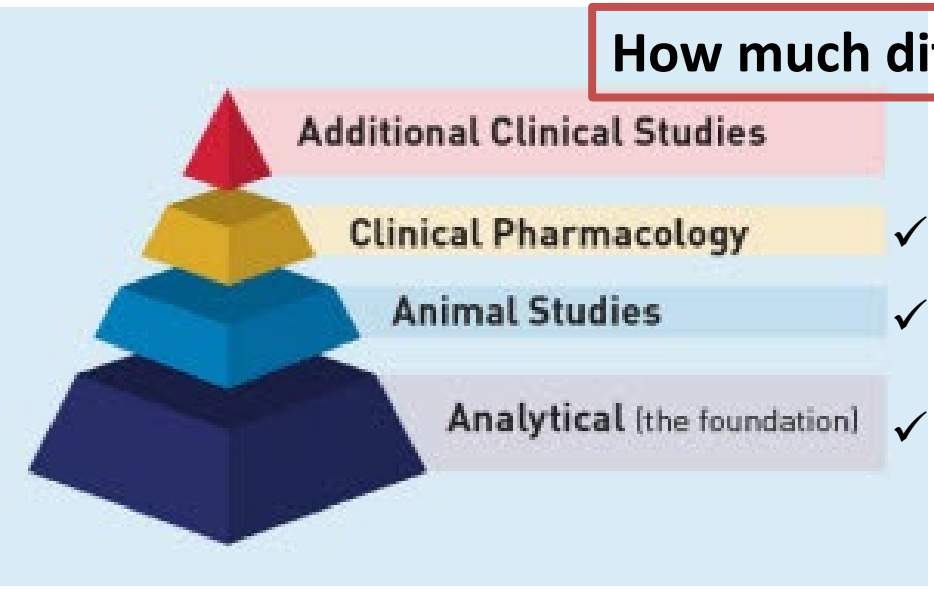
Similarity Margin in Biosimilar Studies

Jessica Kim, Ph.D.
Division of Biometrics VIII
Office of Biostatistics/OTS/CDER/FDA

FDA Public Workshop (9/19/2022)

Similarity Margin
No Clinically Meaningful Difference

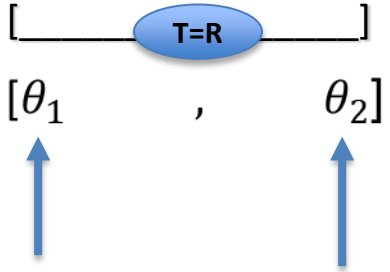
- **No clinically meaningful difference** between highly similar products (the biological product and the reference product) in "comparative clinical studies"



Source: "Abbreviated" Development Program, 351(k), FDA

Statistical Equivalence
Beyond Product Quality and Performance

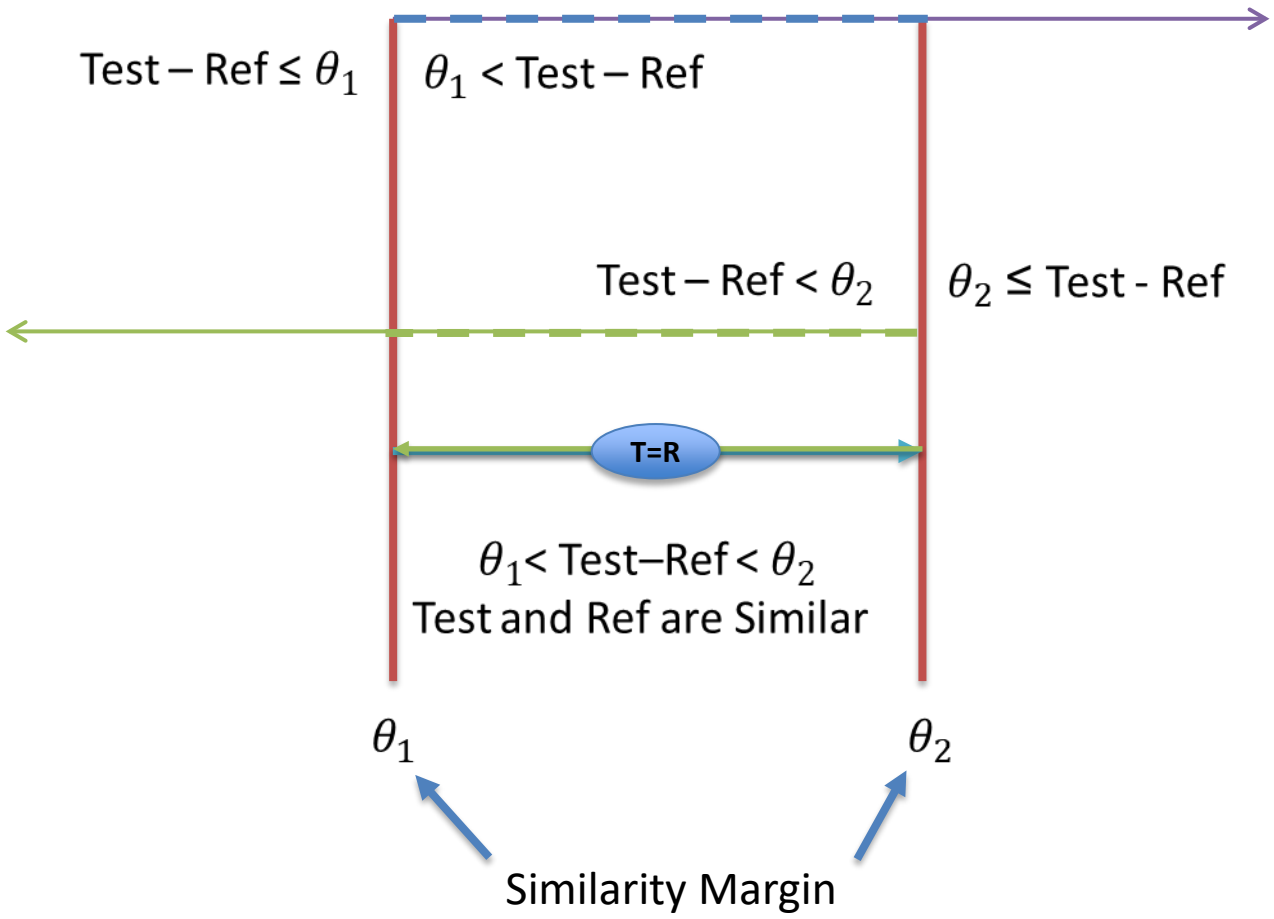
No significant difference between
the biologic product and the reference product



Equivalence (Similarity) Margin

- Statistical equivalence approach:
  - ✓ Two One-Sided Non-Inferiority (NI) Tests Approach
  - ✓ Equivalence (Similarity) Margin : NI margin framework

Statistical Equivalence: No Significant Difference



Big Question

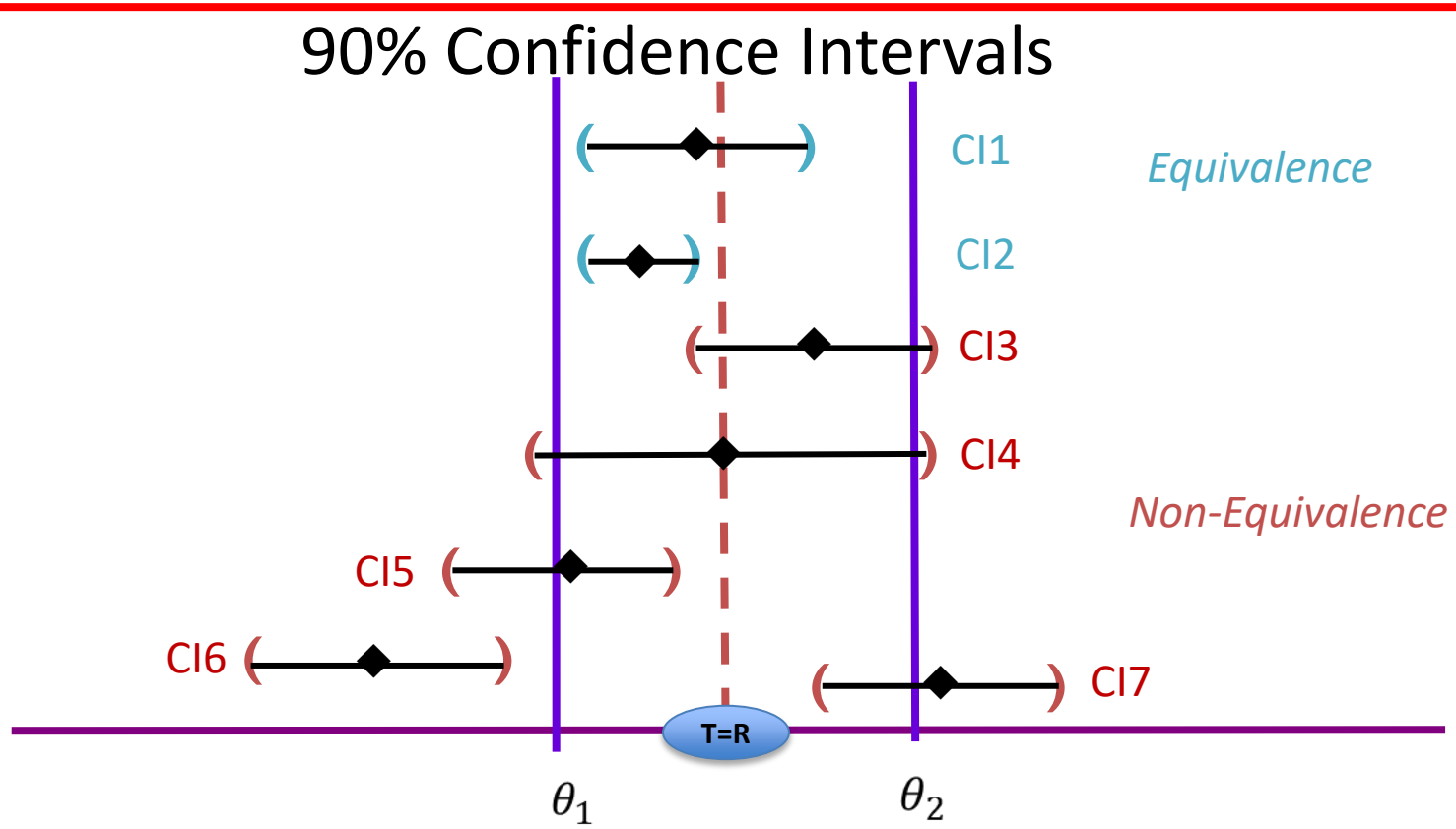
- Similarity margin: **NI margin framework**
  - How much difference to rule out?
  - How do we determine { θ_1, θ_2 }?

  ✓ Science-based

  ✓ Regulation

  ✓ Synthesizing evidence
    - ✓ Non-Inferiority Clinical Trials | FDA
    - ✓ Meta-Analyses of Randomized Controlled Clinical Trials to Evaluate the Safety of Human Drugs or Biological Products | FDA
    - ✓ Demonstrating Biosimilarity | FDA

Statistical Equivalence



- ✓ Typically, M1 is estimated by using 95% confidence level. How about 80% or 90%?
- ✓ CI3 and CI7 cases, asymmetric limits (or lower limit only) relaxing the positive direction?
- ✓ CI4 case, apply an alternative approach for endpoints with large variation?


FDA
U.S. FOOD & DRUG
ADMINISTRATION

More Statistical Considerations for Biosimilar Studies

Wanjie Sun

CDER/OTS/OB/DBVIII

FDA

Disclaimer

- The opinions and information in this presentation are those of the authors, and do not represent the views and/or policies of the U.S. Food and Drug Administration.

Outline of Three Topics

- ITT vs. PP for biosimilar studies
- Difference vs. ratio in measure for biosimilar studies
- Quantitative vs. qualitative treatment-by-subgroup interactions for biosimilar studies

Topic 1:
ITT vs. PP for Biosimilar Studies

Difference between ITT and PP in Efficacy vs. Biosimilar/Bioequivalence(BE)/Non-Inferiority (NI) Trials

- "The **full analysis** set and the **per protocol** set play **different** roles
- in **superiority** trials

  (which seek to show the investigational product to be **superior**) and
- in **equivalence** or **noninferiority** trials

  (which seek to show the investigational product to be **comparable**)"

(ICH E9 1998)

FDA

ITT Analysis in Efficacy vs. Biosimilar/BE/NI Trials

- **Intent-to-treat (ITT)**: **Full Analysis Set/Treatment Policy**:
  - "**Avoid over-optimistic** estimates of efficacy resulting from a **PP** analysis" (ICH E9 1998)
  - "**Non-compliers** included in the fully analysis set will generally **diminish** the estimated treatment effect." (ICH E9 1998)

  - Conservative for **superiority** test/efficacy trials
  - "However, in an **equivalence** or **noninferiority** trial, use of the full analysis set is generally **NOT conservative** and its role should be considered very **carefully**"(ICH E9 1998)
  - **Diluted treatment effect -> easier** to establish Equivalence/NI -> **Anti-conservative** for **Biosimilar/BE/NI** test

  (Sanchez & Chen 2006, Snapinn 2000, FDA NI Guidance 2016, ICH E9 1998, ICH E9 R1 Addendum 2019)

PP Analysis in Efficacy vs. Biosimilar/BE/NI Trials

- **Per – Protocol (PP)**: **completers** & **compliers**
  - For **Superiority/Efficacy:**

"**Over-optimistic** estimates of efficacy" -> less conservative than ITT analysis

- For **Biosimilars/BE/NI** trials:

PRO:

- "**most closely** reflects the **scientific model** underlying the **protocol**" (ICH E9 1998), "Pure treatment effect"
- **more conservative** than ITT in that it **excludes non-compliance** which **dilutes** treatment difference (Sanchez & Chen 2006, ICH E9 1998)

CON:

- Whether or not a subject is PP: **post-treatment** or **intermediate variable (intercurrent event)**
- A crude comparison within the observed PP population between treatment groups is subject to **selection bias**
- Bias can be in **either** direction – mechanism **complicated**

(FDA 2016 NI Guidance, Frangakis and Rubin 2009, Sanchez & Chen 2006, Snapinn 2000, Sanchez & Chen 2006, Lou et al 2018/2019)

Lou, Jones & Sun JBS 2019, SIM 2019

- **Quantified** the **bias** of the PP estimator
- Identified **three conditions** under which PP estimator can be unbiased.
- Proposed a **tipping point sensitivity** method based on causal estimand and principal stratification strategy to evaluate the **robustness** of the observed PP analysis when sensitivity parameters **deviate** from the three conditions but stay within a clinically meaningful range

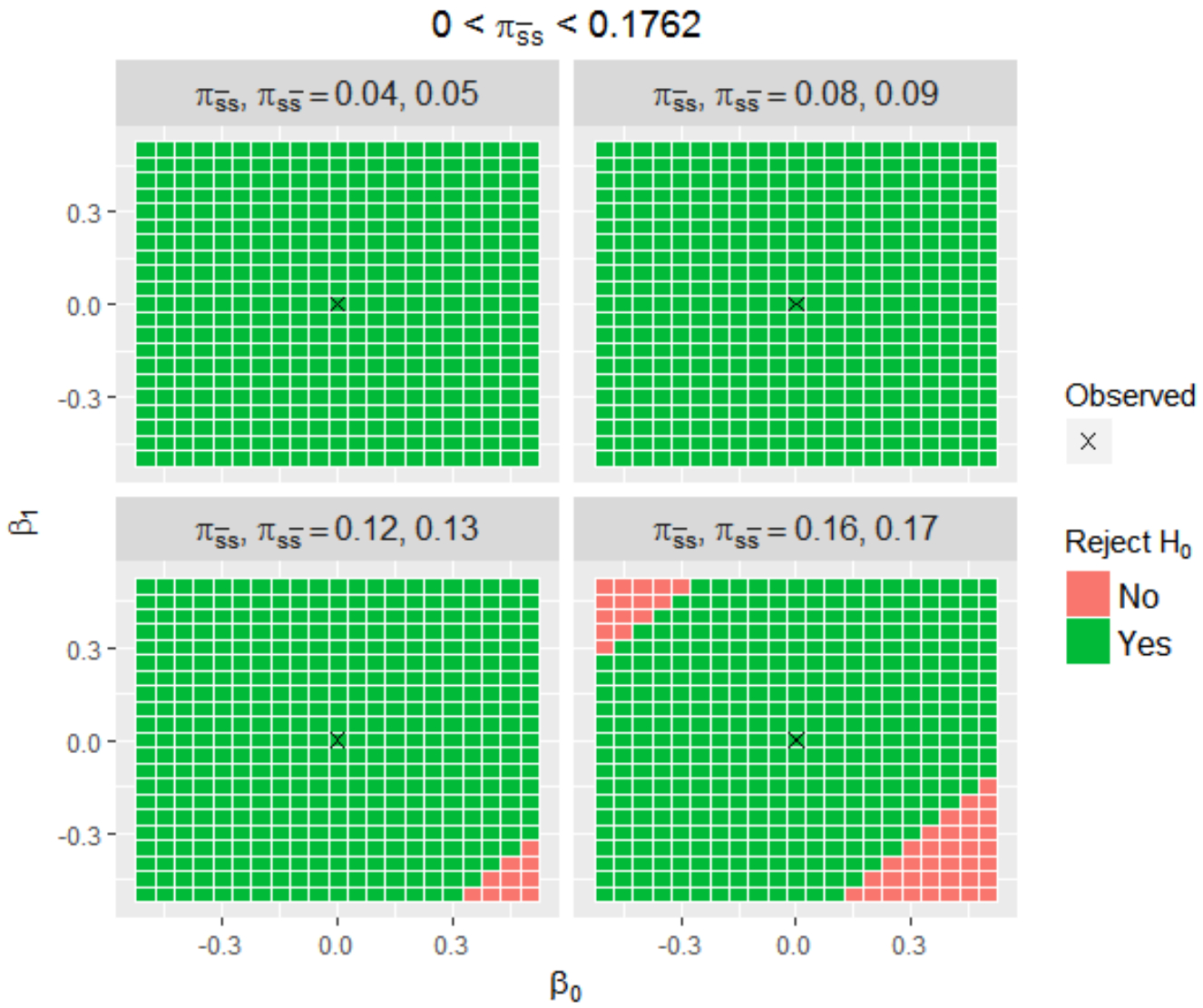
(**Quite a few** sensitivity analyses have been proposed for **ITT** analysis regarding how to deal with missing data.

However, **very few** sensitivity analyses have been proposed for **PP** analysis).

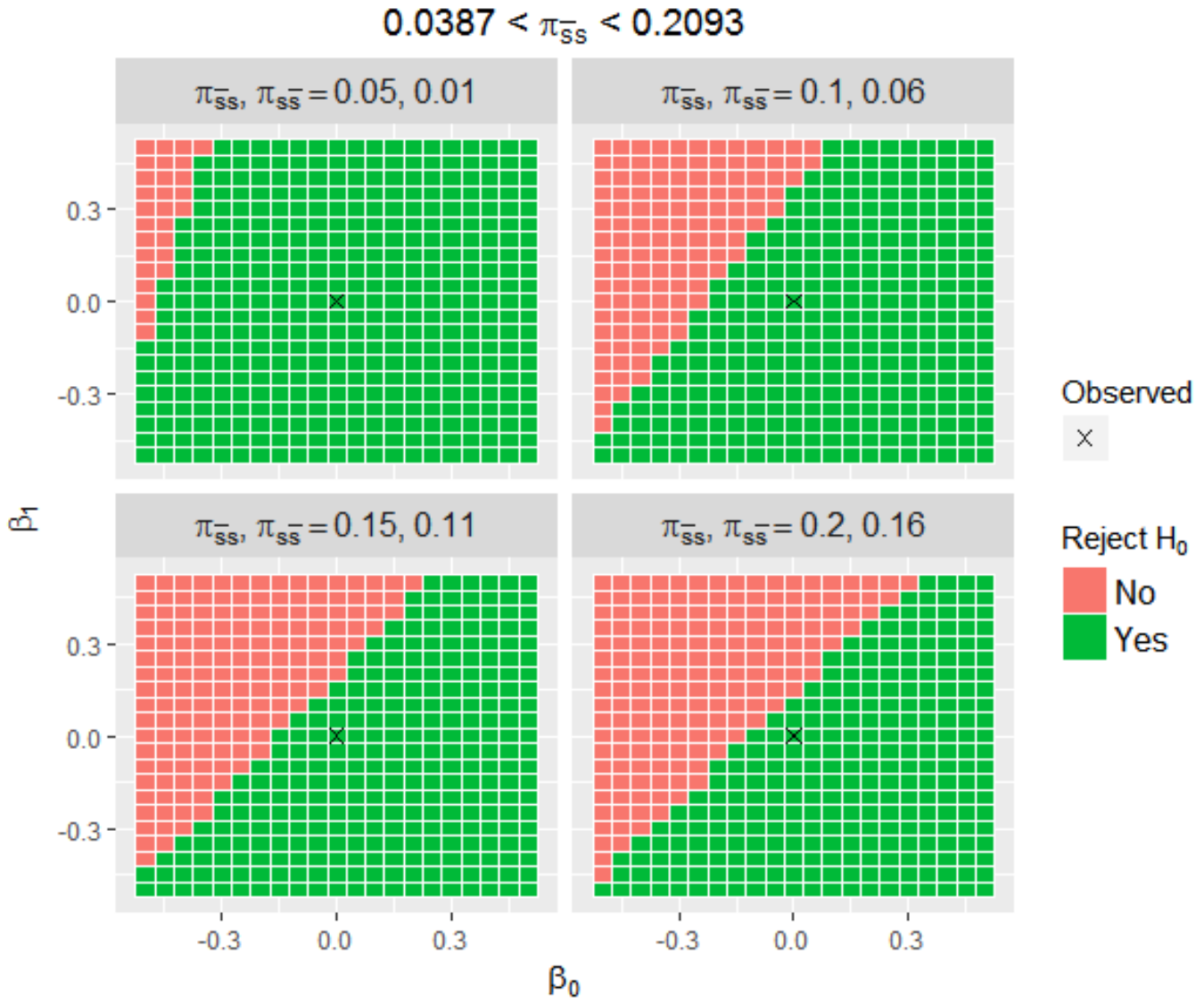
- This proposed tipping point sensitivity method can be applied to
  - **comparative clinical** studies for **biosimilar** products
  - **comparative clinical endpoint BE** studies for **generic** products

FDA

Tipping Point Sensitivity Analysis for PP Analysis (Lou, Jones & Sun JBS 2019, SIM 2019) Example 1



Tipping Point Sensitivity Analysis for PP Analysis (Lou, Jones & Sun 2018, 2019) Example 2



Topic 2:
Difference vs. Ratio in Measure for Biosimilar Studies

Sun, Grosser, Tsong, JBS 2017

Table 3. Summary of recommendations for choice between difference of means and ratio of means for the superiority, NI and ABE tests.

	Measure			
	Difference of Means (DOM)		Ratio of Means (ROM)	
Condition	Superiority	NI or ABE	Superiority	NI or ABE
(1) Biological effect of the treatment for difference control group values	Additive		Multiplicative	
(2) Units of the outcome of interested	Need to be identical units		Can be different units	
(3) Mean value of reference group	Does not matter		Cannot be very small (e.g., close to zero)	
(4) Sign of means of two treatment groups	Does not matter		Needs to be both positive or both negative Rothmann et al. (2012)	
(5) Direction of the scoring system for the outcome of interested	Does not matter		It doesn't matter because ROM and DOM are identical for superiority.	Power can change significantly as the direction of the scores changes.
(6) Shift in scoring systems with the same margins				
Location shift	Does not matter	Does not matter	Does not matter	Power can change
Scale shift with a positive scale factor	Does not matter	Power can change	Does not matter	Does not matter
Scale shift with a negative scale factor	Does not matter	Power can change	Does not matter	Power can change
Combined location & scale shift	Does not matter	Power can change	Does not matter	Power can change
Other shift (non-location or -scale)	Power can change	Power can change	Power can change	Power can change

JOURNAL OF BIOPHARMACEUTICAL STATISTICS

With a fixed effect size and margin, Power is an increasing function of Reference Mean for ROM and constant for Reference Mean for DOM in Biosimilar/BE studies

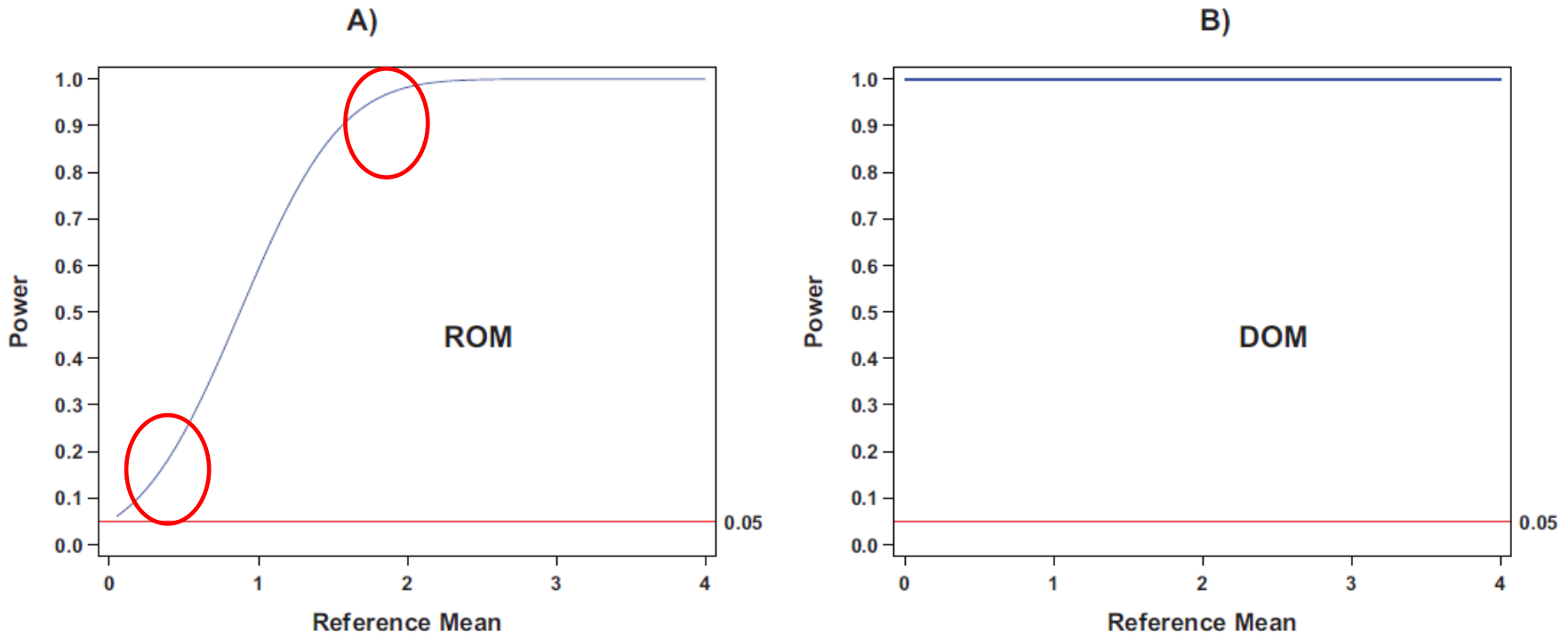
-> ROM in Biosimilar/BE studies: Direction of Y value impacts power vastly:

If Y is defined as the larger the better -> Power is high when Reference is good;

If Y is defined as the smaller the better -> Power is low when Reference is good;

-> Caution should also be used for Proportion Ratio, Odds Ratio, & Hazard Ratio in Biosimilar/BE studies

(Sun, Grosser, Tsong, JBS 2017; Sun, Grosser, Kim, & Raney JBS 2019)

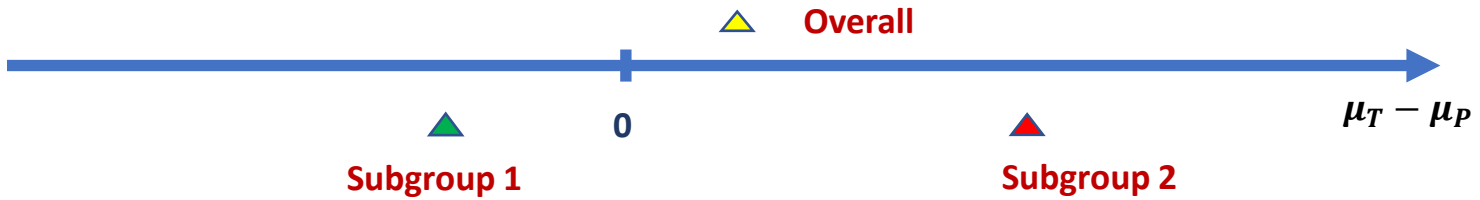


Topic 3:
Qualitative vs. Quantitative Treatment-by-Subgroup Interaction for Biosimilar Studies

For Efficacy Studies

Quantitative Interaction

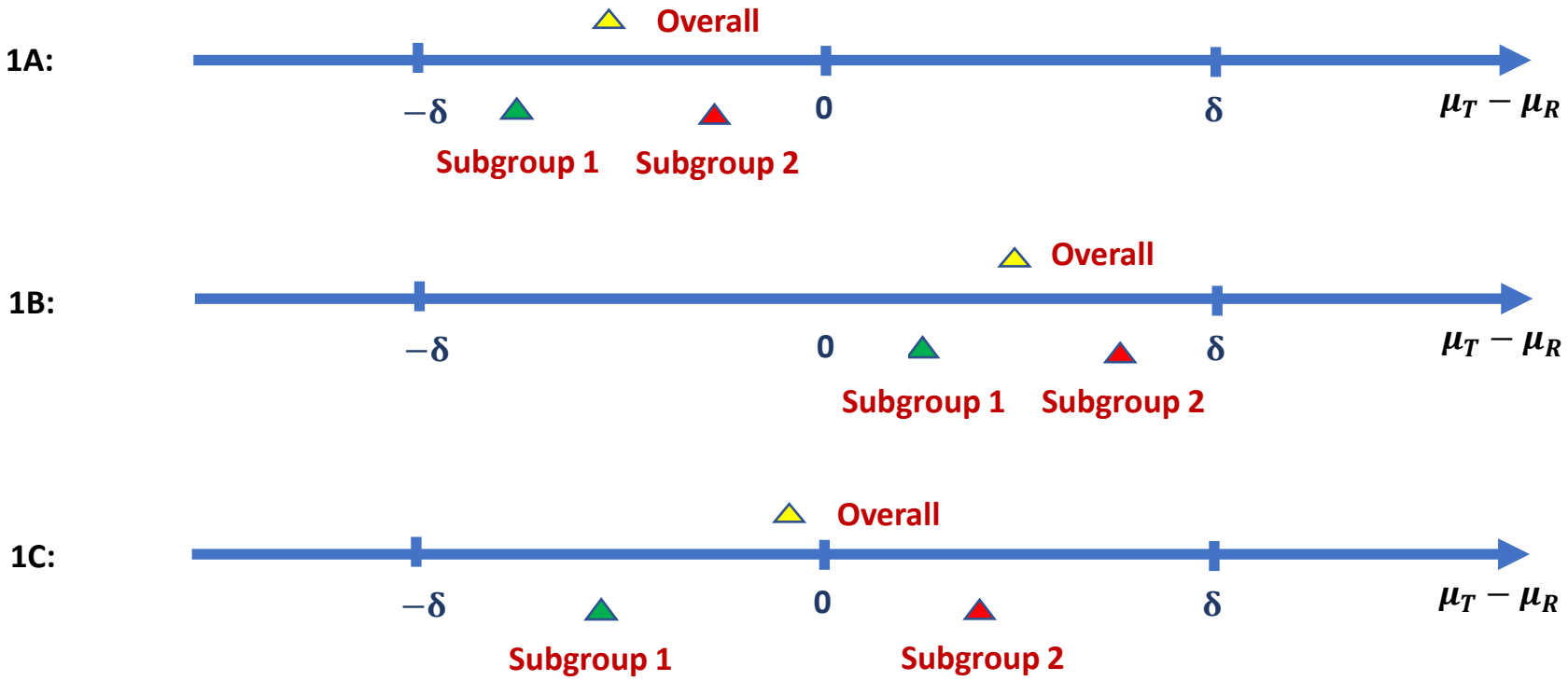
Qualitative Interaction



- Overall Population Mean Treatment Difference Δ
- Subgroup 1 Population Mean Treatment Difference Δ_1
- Subgroup 2 Population Mean Treatment Difference Δ_2

FDA

For Biosimilar/BE Studies
Quantitative Treatment-by-subgroup Interaction
(Sun W, Schuirmann D, Grosser S, SBR 2022)



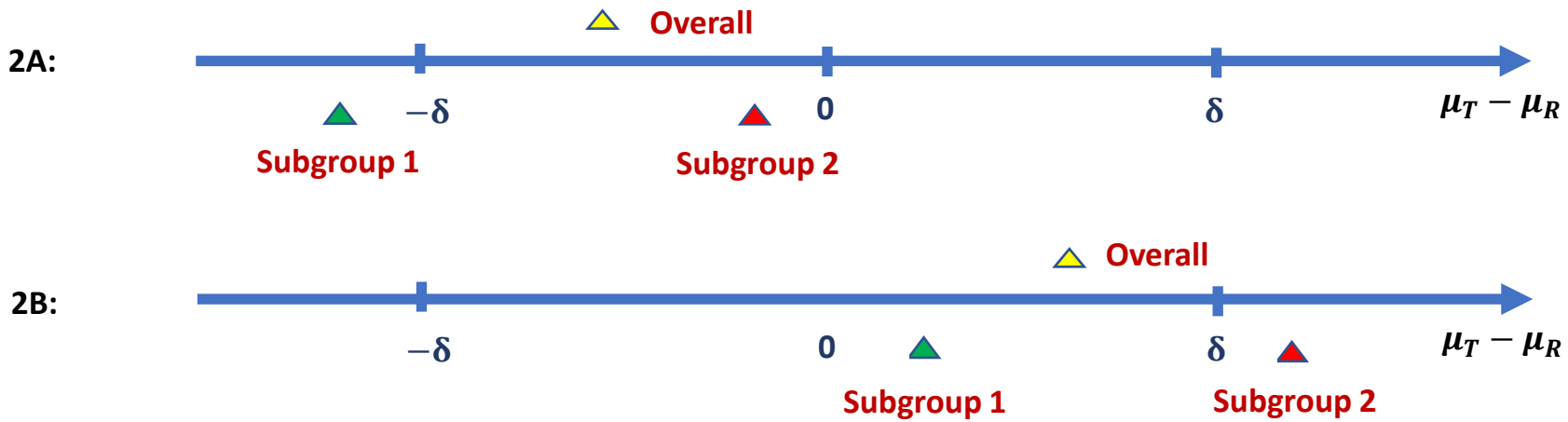
Overall Population Mean Treatment Difference Δ

Subgroup 1 Population Mean Treatment Difference Δ_1

Subgroup 2 Population Mean Treatment Difference Δ_2

FDA

For Biosimilar/BE Studies
Concordant Qualitative Treatment-by-subgroup Interaction
(Sun W, Schuirmann D, Grosser S, SBR 2022)



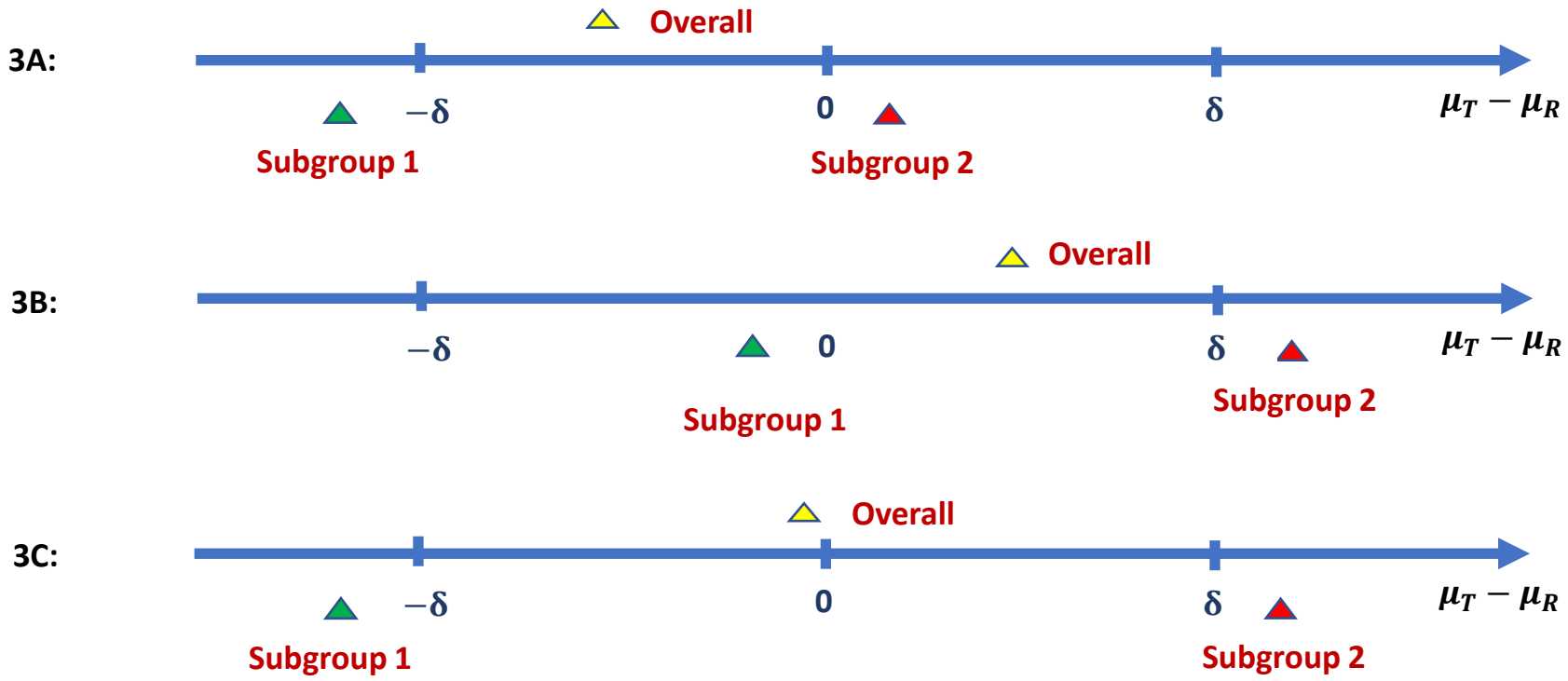
Overall Population Mean Treatment Difference Δ

Subgroup 1 Population Mean Treatment Difference Δ_1

Subgroup 2 Population Mean Treatment Difference Δ_2

FDA

For Biosimilar/BE Studies
Discordant Qualitative Treatment-by-subgroup Interaction
(Sun W, Schuirmann D, Grosser S, SBR 2022)



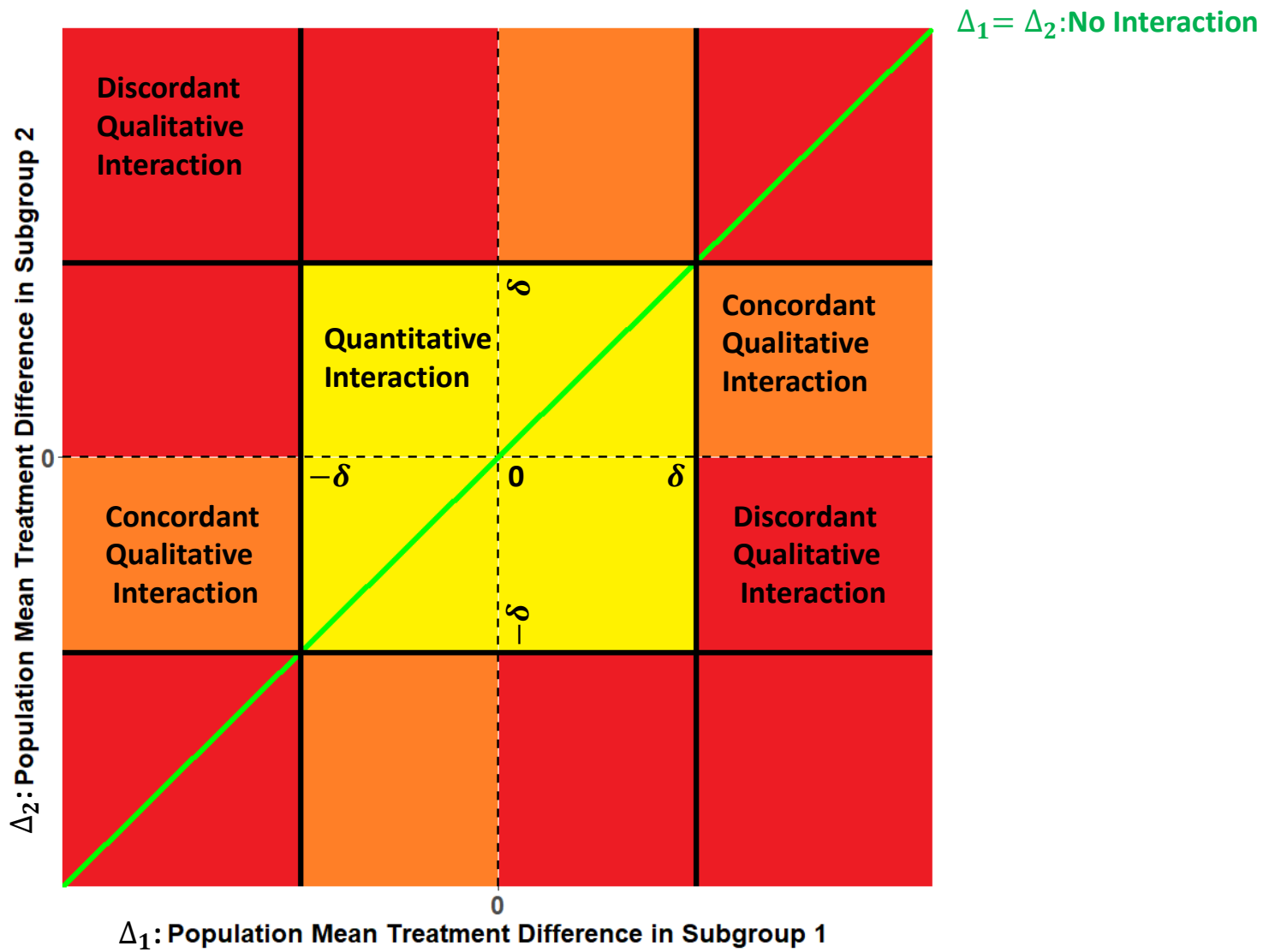
Overall Population Mean Treatment Difference Δ

Subgroup 1 Population Mean Treatment Difference Δ_1

Subgroup 2 Population Mean Treatment Difference Δ_2

FDA

For Biosimilar/BE Studies
No Interaction (Green Line), Quantitative (Yellow), Concordant (Orange) and Discordant (Red) Qualitative Treatment-by-subgroup Interaction (Sun W, Schuirmann D, Grosser S, SBR 2022)



Sun, Schuirmann & Grosser SBR 2022

- **Defined** quantitative and qualitive treatment-by-subgroup interactions for equivalence studies.
- Qualitative interactions are **not common** in placebo-controlled **efficacy** studies.
- However, qualitative interactions are **more likely** to arise in **biosimilar/BE** studies because the two active drugs can go **either** direction in relative treatment effect.
- The **impact** of discordant qualitative interaction on biosimilar/BE studies is **greater** than that of qualitative interaction in superiority.
- Because it is possible that there is equivalence in none of the subgroups but the **overall** treatment effect can be **equivalent** due to the **offset** of the **opposite** directions of **inequivalence** in subgroups.
- Be **cautious** about qualitative treatment-by-subgroup interactions in biosimilar studies!

FDA

References

- Lou, Yiyue, Jones, Michael P., Sun, Wanjie. Assessing the Ratio of Means as a Causal Estimand in Clinical End-point Bioequivalence Studies. *Statistics in Medicine* (2019); 38 (27): 5214-5235.
- Lou, Yiyue, Jones, Michael P., Sun, Wanjie. Estimation of causal effects in clinical endpoint bioequivalence studies in the presence of intercurrent events: noncompliance and missing data. *Journal of Biopharmaceutical Statistics (*2019); 29 (1), 151-173.
- Sun, Wanjie, Grosser, Stella, and Tsong, Yi. Ratio of means vs. difference of means as measures of superiority, noninferiority, and average bioequivalence. *Journal of biopharmaceutical statistics* (2017), 27(2), pp.338-355.
- Sun, Wanjie, Stella Grosser, Carol Kim, and Sam G. Raney. "Statistical considerations and impact of the FDA draft guidance for assessing adhesion with transdermal delivery systems and topical patches for ANDAs." *Journal of Biopharmaceutical Statistics* 29, no. 5 (2019): 952-970.
- Sun, Wanjie, Schuirmann, Don, Grosser, Stella. Qualitative vs. quantitative treatment-by-subgroup interaction in equivalence studies with multiple subgroups. *Statistics in Biopharmaceutical Research* (2022); in press.

FDA


FDA
U.S. FOOD & DRUG
ADMINISTRATION

FDA U.S. FOOD & DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION & RESEARCH
OFFICE OF CLINICAL PHARMACOLOGY

Clinical Pharmacology Experience in Streamlining Biosimilar Development

FDA Public Workshop (9/19/2022)

Yow-Ming Wang, PhD
Office of Clinical Pharmacology
FDA/CDER/OTS
Yowming.Wang@fda.hhs.gov

Disclaimer

- The presentation today should not be considered, in whole or in part as being statements of policy or recommendation by the United States Food and Drug Administration.

- Throughout the talk, representative examples of commercial products may be given to illustrate a methodology or approach to problem solving. No commercial endorsement is implied or intended.

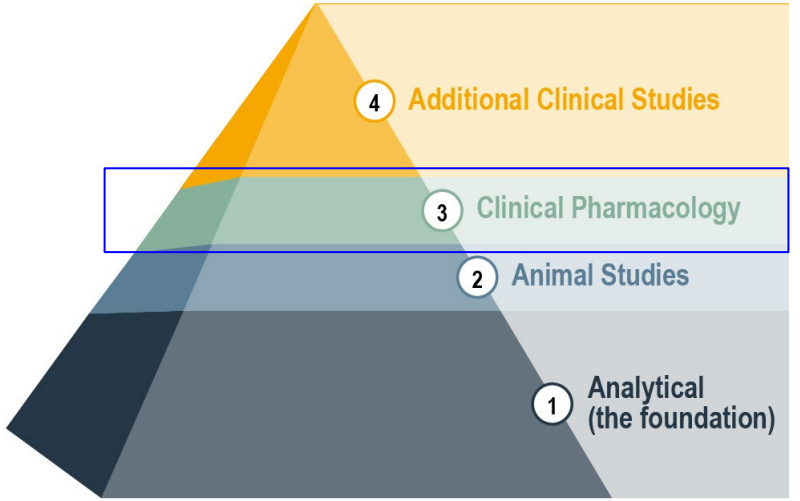
Role of clinical pharmacology data in biosimilar program

351(k) biosimilar BLA

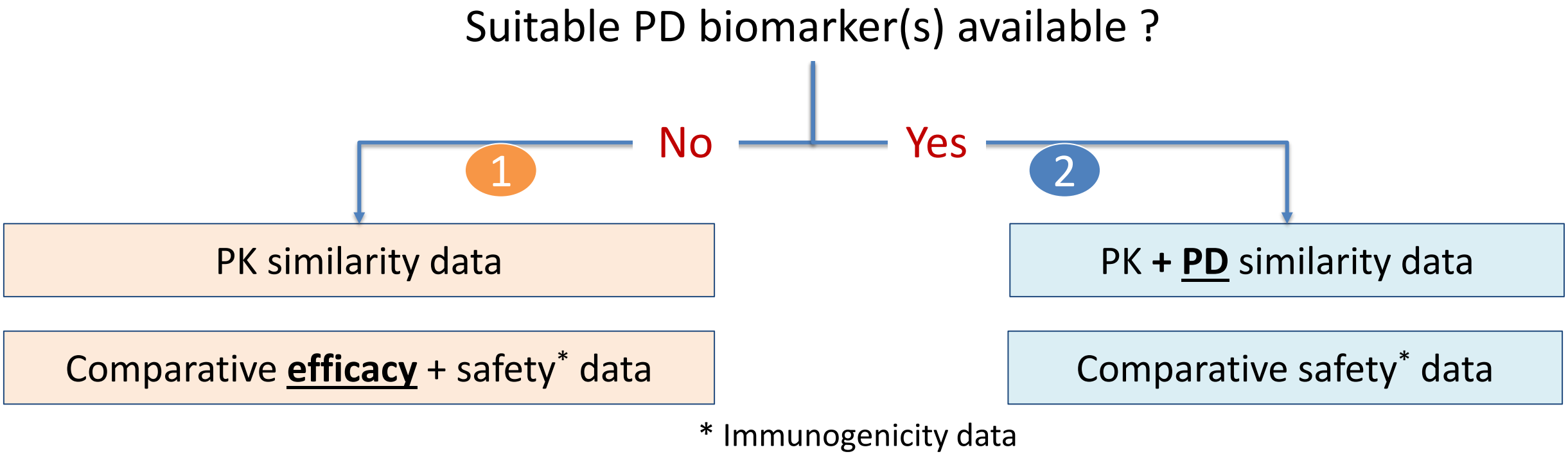
- **Goal:** To demonstrate biosimilarity (or interchangeability) to a reference product

- Clinical pharmacology studies compare pharmacokinetics (PK) and pharmacodynamics (PD) between products
  - Similar exposure (PK)
  - Similar response (PD), if applicable

- With the foundation of analytical similarity, similar PK and PD can support biosimilarity <u>without a comparative clinical study</u>



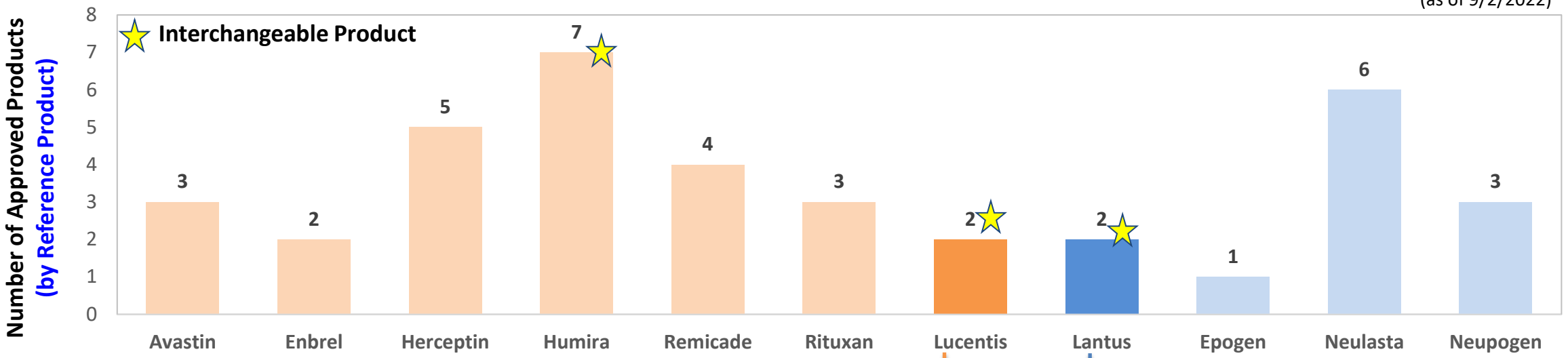
Two approaches supported biosimilar approvals
(when systemic PK is available)



i.e., PD similarity data ***in lieu of*** comparative efficacy data

The FDA has approved 38 biosimilar products, including 3 interchangeable products

(as of 9/2/2022)



N=26

1

PK similarity data

Comparative **efficacy** + safety data (immunogenicity)

X

✓

N= 12

✓

2

PK + **PD** similarity data

(in lieu of comparative efficacy data)

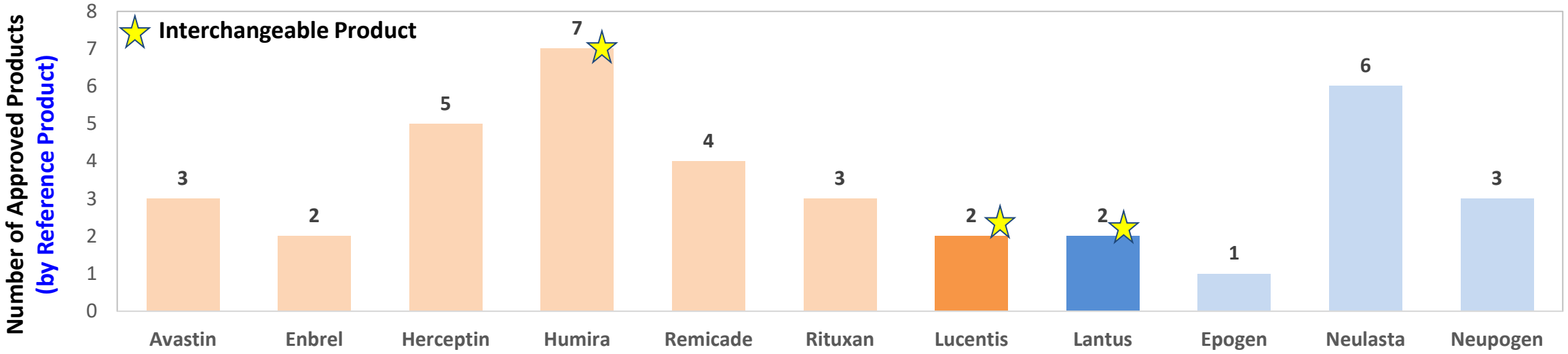
X

Comparative safety data (immunogenicity)

See the FDA's Purple Book for lists of licensed biological products, with reference product exclusivity and biosimilarity or interchangeability evaluations
https://go.usa.gov/xz6Ud

Biosimilar programs with PK and PD studies are more efficient

- Sample size — Comparative Clinical Study (CCS) > PK and PD similarity study
- Study duration — Longer for CCS than PK and PD study

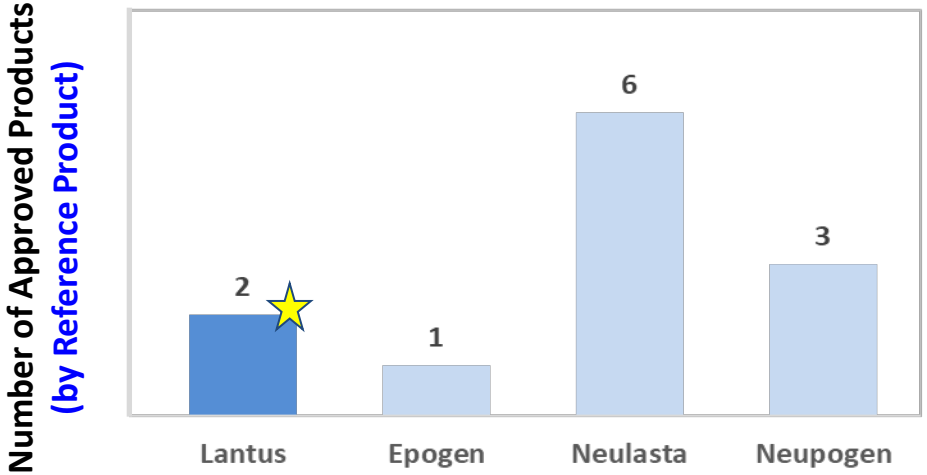


Study Sample Size (total number of subjects)										
Comparative Clinical Study (CCS)							PK + PD study			
~700	~500	~700	~600	~600	~300	~600	95	80-130	~270	25-60

Biosimilar programs with PK and PD studies are more efficient

- Sample size — Comparative Clinical Study (CCS) > PK and PD similarity study
- Study duration — Longer for CCS than PK and PD study

Study Sample Size (total number of subjects)				
	PK + PD approach			
PK + PD study	95	80-130	~270	25-60
Comparative Clinical Study (CCS, not required for BLA)	~600	~400	~300	~200



PD biomarkers used

Glucose infusion rate — Lantus

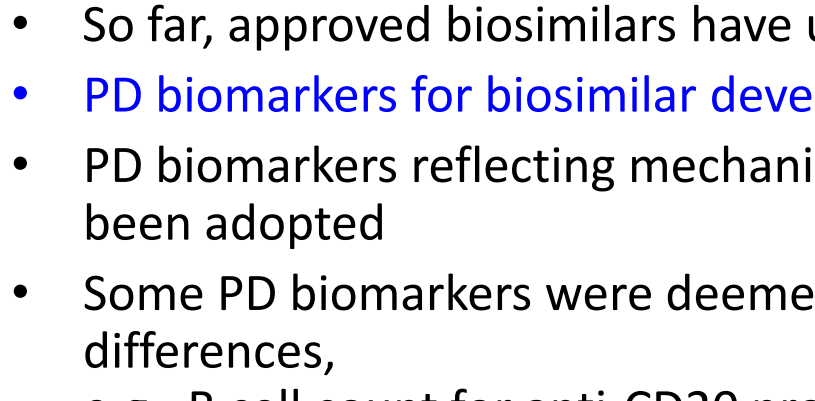
Reticulocyte count & Hemoglobin (Hb) — Epogen

Absolute neutrophil count (ANC) — Neulasta/Neupogen

CD34+ cell count — Neupogen

PD biomarker is an active topic in biosimilar IND discussions

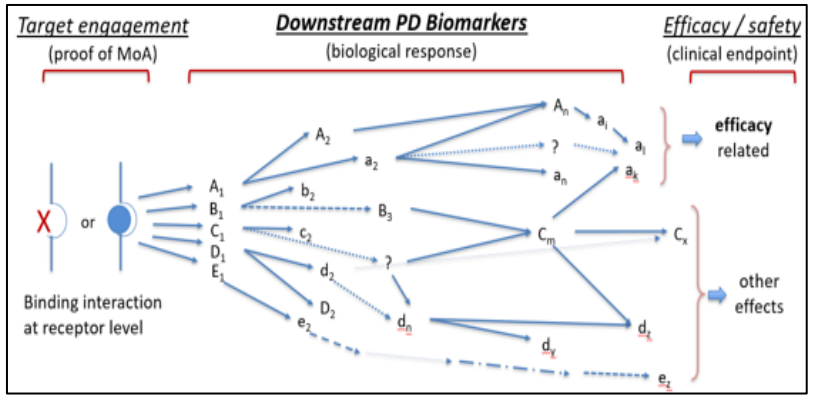
- FDA is engaged in advancing PD biomarkers for biosimilar development & approval
- So far, approved biosimilars have used PD biomarkers that are tied to clinical efficacy
- PD biomarkers for biosimilar development are **not** required to reflect clinical efficacy
- PD biomarkers reflecting mechanism of action while **not** tied to clinical efficacy have been adopted
- Some PD biomarkers were deemed not suitable for detecting clinically meaningful differences,
  e.g., B cell count for anti-CD20 products
- Opportunities exist for continued investigation of PD biomarkers for future biosimilars

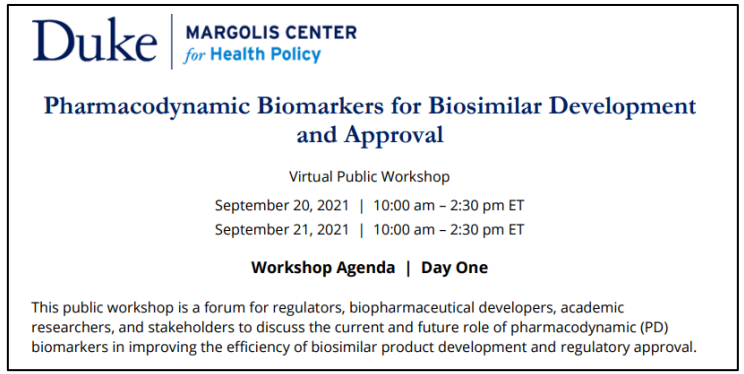

Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

December 2016
Biosimilars




Duke | MARGOLIS CENTER for Health Policy

Pharmacodynamic Biomarkers for Biosimilar Development and Approval

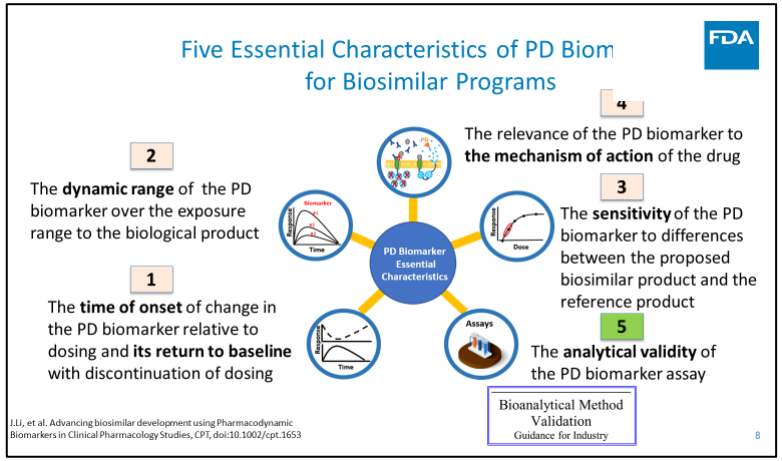
Virtual Public Workshop

September 20, 2021 | 10:00 am – 2:30 pm ET
September 21, 2021 | 10:00 am – 2:30 pm ET

Workshop Agenda | Day One

This public workshop is a forum for regulators, biopharmaceutical developers, academic researchers, and stakeholders to discuss the current and future role of pharmacodynamic (PD) biomarkers in improving the efficiency of biosimilar product development and regulatory approval.

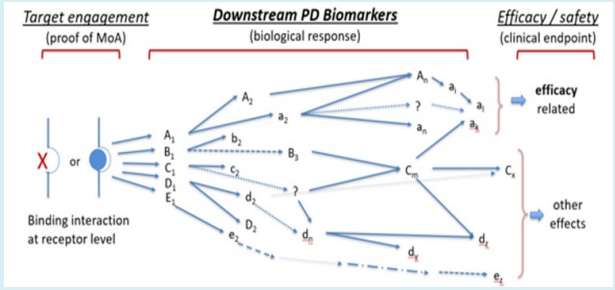
https://healthpolicy.duke.edu/events/biosimilar



Clinical pharmacology experience may be applicable to streamlining clinical studies for biosimilars

PD biomarker(s) for PD similarity study

PD biomarkers tied to efficacy



Consider mechanism of action

Sensitive PD biomarkers

Clinical endpoint(s) for CCS

Primary efficacy endpoint(s)

Early clinical studies → Dose-finding studies → Confirmatory trials

- Clinical study endpoints throughout development
- Primary, secondary, exploratory

Sensitive clinical endpoint(s)

Key criterion

Ability to detect clinically meaningful differences

Approaches to streamline biosimilar/interchangeable programs

Adopt the approach of PK + PD similarity studies

Benefits of PK and PD studies over comparative clinical studies:

- Smaller sample size (higher sensitivity with PD endpoints vs. clinical endpoints)
- Shorter study duration
- Ease of recruitment when feasible in healthy subjects

See the FDA guidance Clinical Pharmacology Data to Support a Demonstration of Biosimilarity to a Reference Product

Certain studies are not necessary when scientifically justified

For example, comparative immunogenicity data for insulin products may not be needed:

See the FDA guidance Clinical Immunogenicity Considerations for Biosimilar and Interchangeable Insulin Products

More scientific innovations are needed!

e.g., SMART clinical endpoint?


FDA
U.S. FOOD & DRUG
ADMINISTRATION

BIOSIMILAR COMPARATIVE CLINICAL ENDPOINT STUDY DESIGN: CHOICES TO OPTIMIZE EFFICIENCY

CLINICAL PERSPECTIVES

Steven Lemery, MD, MHS
Director, Division of Oncology 3

BIOSIMILAR COMPARATIVE CLINICAL ENDPOINT STUDY DESIGN: CHOICES TO OPTIMIZE EFFICIENCY

CLINICAL PERSPECTIVES

Nikolay Nikolov, MD
Director, Division of Rheumatology and Transplant Medicine

BIOSIMILAR COMPARATIVE CLINICAL ENDPOINT STUDY DESIGN: CHOICES TO OPTIMIZE EFFICIENCY

PANEL DISCUSSION

FDA WORKSHOP: INCREASING THE EFFICIENCY OF BIOSIMILAR DEVELOPMENT PROGRAMS

WORKSHOP SUMMARY AND CONCLUDING REMARKS

Sarah Yim, M.D.
Director, Office of Therapeutic Biologics and Biosimilars
Food and Drug Administration