On Assessing Bioequivalence and Interchangeability between Generics Based on Indirect Comparisons

Jiayin Zheng[1], Shein-Chung Chow[1] and Mengdie Yuan[2]
[1]Department of Biostatistics & Bioinformatics, Duke University
[2]Center for Drug Evaluation and Research Organization, Food and Drug Administration

Outline

- Introduction
- Existing methods
- Fiducial probability and restricted confidence interval
- Some extensions
- Similarity assumptions
- Simulation study and real example analysis
- Conclusion

Introduction

- Before a generic is approved to the market place, FDA requires the sponsor to conduct a bioequivalence study to demonstrate the generic is bioequivalent to the brand name drug
- The rate and extent of drug absorption
- pharmacokinetic/pharmacodynamic (PK/PD) , C_{max}, AUC_{0-t}
- Average bioequivalence test: if the 90% confidence interval of the geometric mean ratio (GMR) between two drugs falls into a pre-specified interval, say 80.00% and 125.00%, we claim that two drugs are average bioequivalent.

Introduction (practical problem)

- Interchangeably use of approved generics without any mechanism of safety monitoring in practice
- The safety/efficacy concerns
- However, bioequivalence assessment for regulatory approval among generics is not required
- A lack of head-to-head comparative trials between all available generics: indirect comparison to estimate the relative bioavailabilities between generics by using the summary results from available trials

Notations

- Assume two generics, denoted by G_A and G_B, have been shown to be bioequivalent to the same brand name drug (denoted by B_R)
- Two existing trials: G_A versus B_R, G_B versus B_R
- Denote (L_A, U_A) and (L_B, U_B) as the $1 - 2\alpha$ confidence intervals of $\mu_A - \mu_R$ and $\mu_B - \mu_R$, respectively, where μ_A, μ_B, and μ_R are the true logarithmic geometric means of G_A, G_B, and B_R
- Denote $(\delta_L, \delta_U) = (\log(0.8), \log(1.25))$ as the bioequivalence limits. The approval of both generics require that both (L_A, U_A) and (L_B, U_B) fall within (δ_L, δ_U).
- $t\alpha(df)$ and $t(df)$ are the α quantile and the distribution of the Student's t distribution with the degree of freedom df, respectively.

Existing methods

- One existing method considered as the simplest and most suitable one among those performing indirect comparisons is adjusted indirect comparison.
- Assume $\mu_A - \mu_R$ and $\mu_B - \mu_R$ were estimated by $\hat{v}_{AR}$ and $\hat{v}_{BR}$ in the trials of G_A *vs* B_R and G_B vs B_R, respectively.
- Under the similarity assumption, based on indirect comparison, we get a point estimator of $\mu_A - \mu_B$ as $\hat{v}_{AR} - \hat{v}_{BR}$, of which the variance was estimated by $\widehat{Var}_{AB}$. With the normality assumption, the $1-2a$ confidence interval of $\mu_A - \mu_B$ by adjusted indirect comparison can be expressed as

$$\hat{v}_{AR} - \hat{v}_{BR} \pm t_{\alpha}(df)\sqrt{\widehat{Var}_{AB}}.$$

- $t_a(df)$ is the a quantile of the Student's t distribution with the degree of freedom df.

Existing methods

- However, this method has some limitations regarding both clinical side and statistical side.
- For clinical practice, the bioequivalence between generics makes sense only if both generics are bioequivalent to the corresponding brand name drug.
- It derives the confidence interval of $\mu_A - \mu_B$ ignoring the fact that the confidence intervals for $\mu_A - \mu_R$ and $\mu_B - \mu_R$ are already contained within (δ_L, δ_U), potentially resulting in a narrower confidence interval and overestimating clinical meaningful bioequivalence between generics.

Existing methods

- To get a better performance, Garcia-Arieta *et al.* extended the acceptance limit from ±20% to ±25% range, which seems arbitrary. From our simulation studies, it causes the type I error inflated and significantly greater than 0.05.

Restricted confidence interval

- The trial of G_A versus B_R : $\mu_A - \mu_R$ was estimated as $\hat{V}_{AR}$ ($\widehat{Var}(\hat{V}_{AR})$). $(L_A, U_A) = \hat{V}_{AR} \pm t\alpha(df_A)\,\widehat{Var}(\hat{V}_{AR})$.
- The trial of G_B versus B_R : $\hat{V}_{BR}$, $\widehat{Var}(\hat{V}_{BR})$, and $(L_B, U_B) = \hat{V}_{BR} \pm t\alpha(df_B)\,\widehat{Var}(\hat{V}_{BR})$.
- Fiducial distribution of $\mu_I - \mu_R$: $\hat{V}_{IR} \pm t(df_I)\,\widehat{Var}(\hat{V}_{IR})$, where $I = A$ or B, $t(df_I)$ are the Student's t distribution random variable with the degree of freedom df_I.
- f_A and f_B: the *pdf* of the fiducial distribution of $\mu_A - \mu_R$ and $\mu_B - \mu_R$, assumed independent

Restricted confidence interval

- Based on f_A and f_B, for any q that $0 < q \leqslant pr\{\delta_L \leqslant \mu_l - \mu_R \leqslant \delta_U, l = A, B\}$, find the minimal Δ (denoted as Δ_q) satisfying

$$pr\{\delta_L \leqslant \mu_l - \mu_R \leqslant \delta_U, l = A, B \text{ and } -\Delta \leqslant \mu_A - \mu_B \leqslant \Delta\} \geqslant q,$$

then we get a restricted confidence interval of $\mu_A - \mu_B$ as $(-\Delta_q, \Delta_q)$, with the confidence level of q.

- Denote (δ_L', δ_U') $(\delta_L' = -\delta_U')$ as the target bioequivalence limits for $\mu_A - \mu_B$. Given the confidence level of $1-2\beta$, if $\Delta_{1-2\beta} \leqslant \delta_U'$, the clinical meaningful bioequivalence between G_A and G_B can be concluded.
- In other words, in this case, we have

$$pr\{\delta_L \leqslant \mu_l - \mu_R \leqslant \delta_U, l = A, B \text{ and } \delta_L' \leqslant \mu_A - \mu_B \leqslant \delta_U'\} \geqslant 1 - 2\beta.$$

Restricted confidence interval

- Conversely, simply calculate the fiducial probability

$pr\{\delta_L \leqslant \mu_l - \mu_R \leqslant \delta_U,\ l = A, B \text{ and } \delta'_L \leqslant \mu_A - \mu_B \leqslant \delta'_U\}$

(denoted as $q_{\delta'_U}$) and then compare $q_{\delta'_U}$ with the pre-specified confidence level. If $q_{\delta'_U}$ is greater, conclude G_A and G_B are average bioequivalent.

- The fiducial probability can be expressed as follows and numerical integration can be used to get an accurate estimate:

$$\int_{\delta L}^{\delta U} \int_{\delta L \vee (x+\delta'_L)}^{\delta U \wedge (x+\delta'_U)} f_A(x) f_B(y) \mathrm{dydx}$$

Some extensions

- The issue of multiple comparison may arises from a number of pairwise comparisons.
- The family-wise error rate (FWER) versus false discovery rate (FDR)
- Bonferroni's correction
- The Benjamini–Hochberg procedure: this method is still expected to perform well in this article's case, which is under the pairwise comparisons setting, a specific situation of dependence.

  (Benjamini and Hochberg 1995, Benjamini 2010)
- We recommend controlling a low FDR (say 0.1)
- The raw p-value of each comparison can be calculated by

$$1 - pr\{\delta_L \leqslant \mu_l - \mu_R \leqslant \delta_U,\ l = A, B \text{ and } \delta'_L \leqslant \mu_A - \mu_B \leqslant \delta'_U\}.$$

Some extensions

- The extension to accommodate the comparison of more than two generics simultaneously.
- In practice, the comparison of a basket of generics as a whole may arises. Denote G_A, G_B and G_C as three generics corresponding to the same brand-name drug B_R, with their log-geometric means denoted by μ_A, μ_B, μ_C and μ_R. Denote f_l as the probability density function of the fiducial distribution of $\mu_l - \mu_R$.
- The fiducial probability

  $pr\{\delta_L \leqslant \mu_l - \mu_R \leqslant \delta_U,\ l = A, B, C \text{ and } \delta'_L \leqslant \mu_A - \mu_B,\ \mu_A - \mu_C,\ \mu_B - \mu_C \leqslant \delta'_U\}$

  can be obtained.

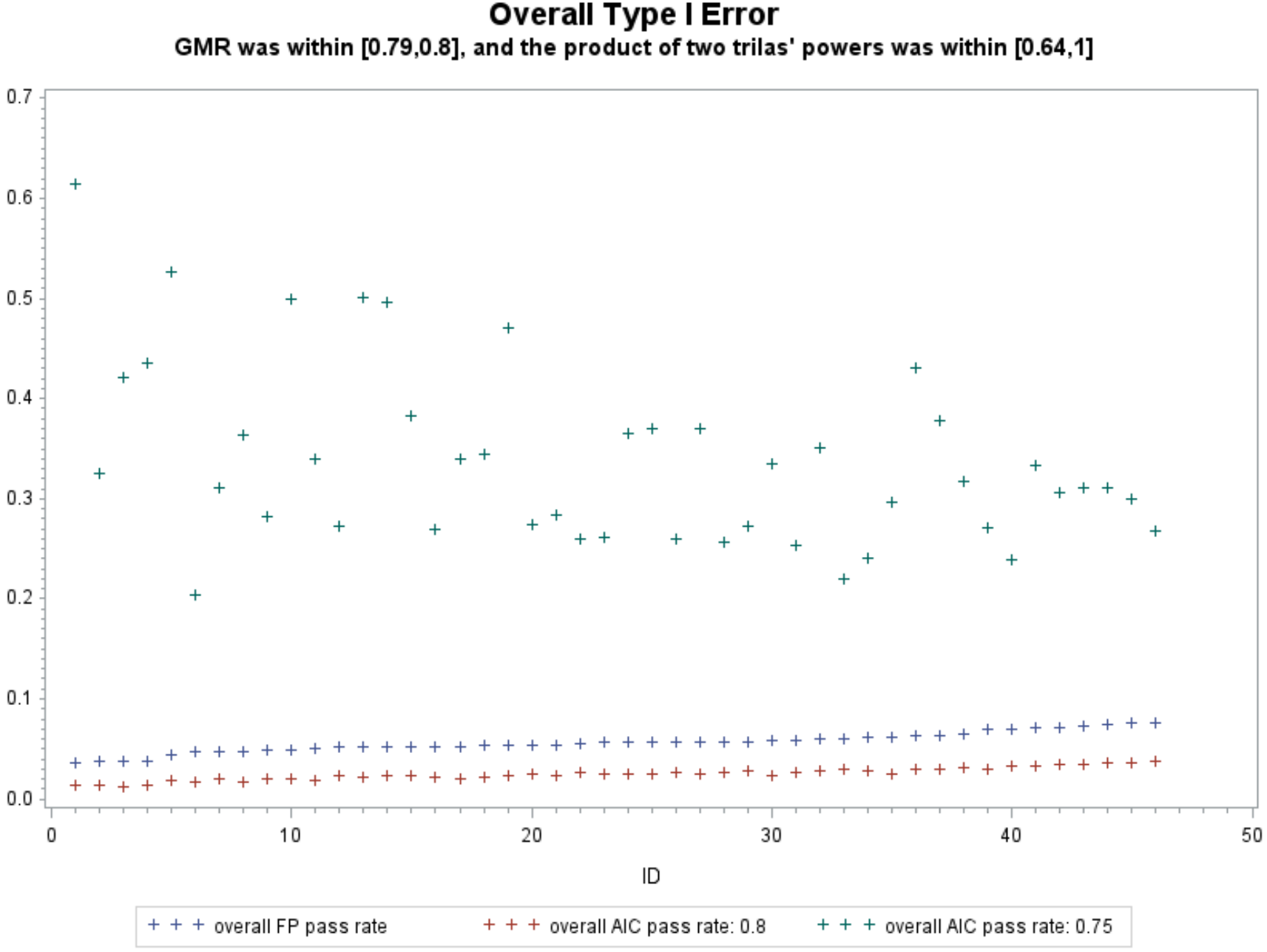
Similarity assumptions

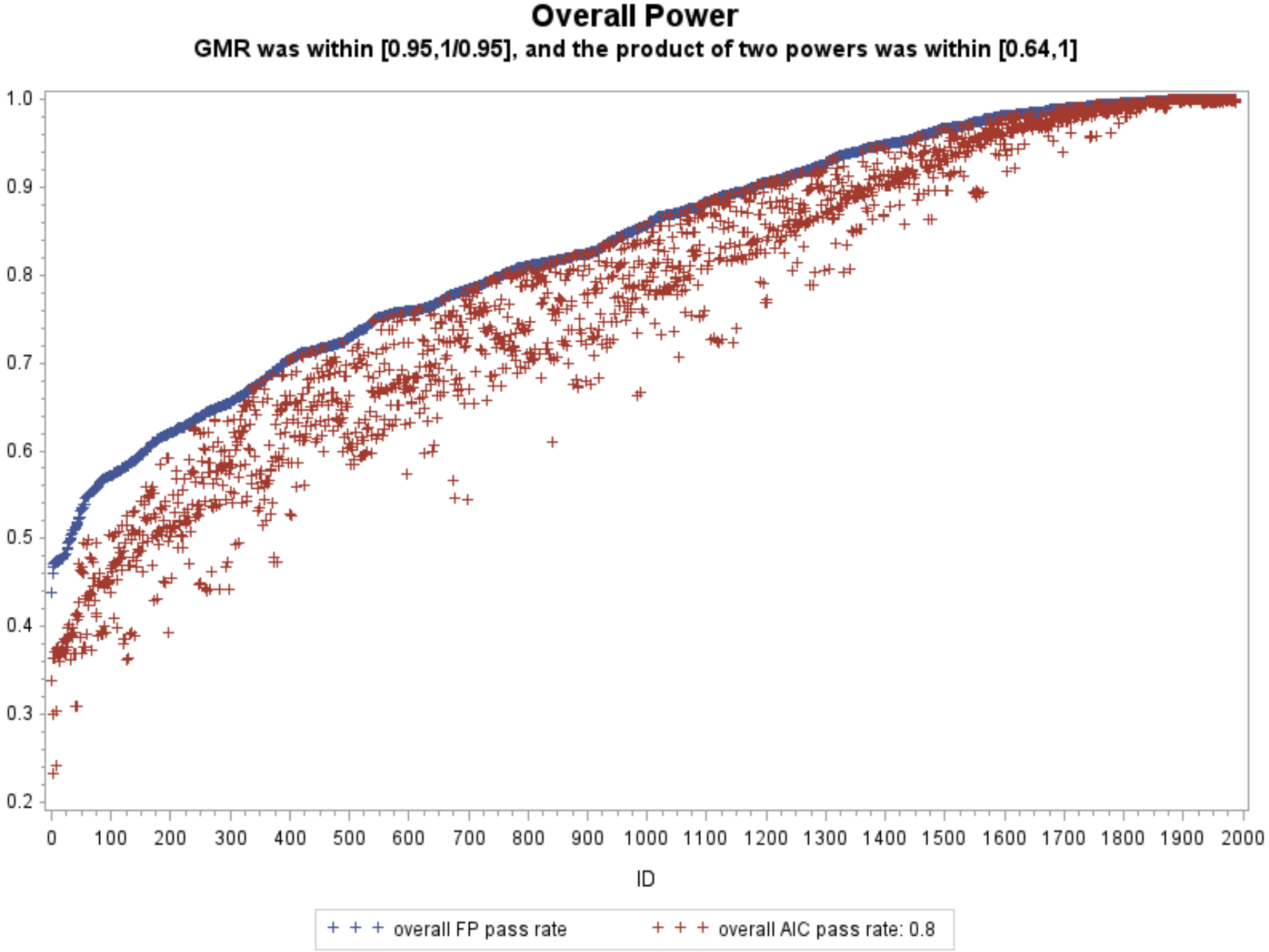
- Clinical similarity and methodological similarity
- The relative effect estimated by the trial of G_A versus B_R is generalizable to patients in the trial of G_B versus B_R , and vice versa.
- Patient characteristics, the mode of drug administration, and parameter measurement
- Methodological quality: similarly biased
- Trials with different designs might not be comparable.
- Basic designs of such bioequivalence studies are generally consistent.

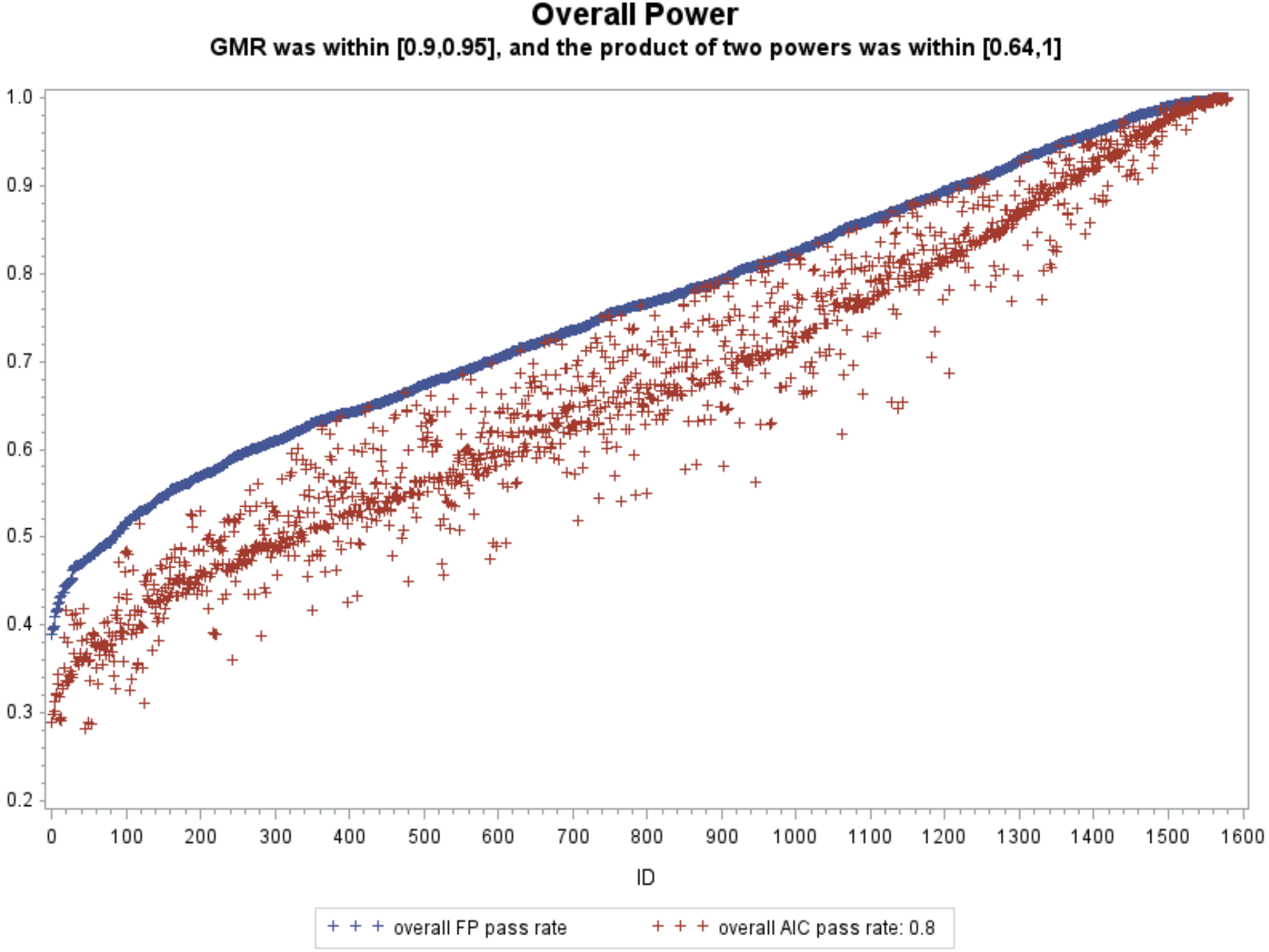
Simulation study

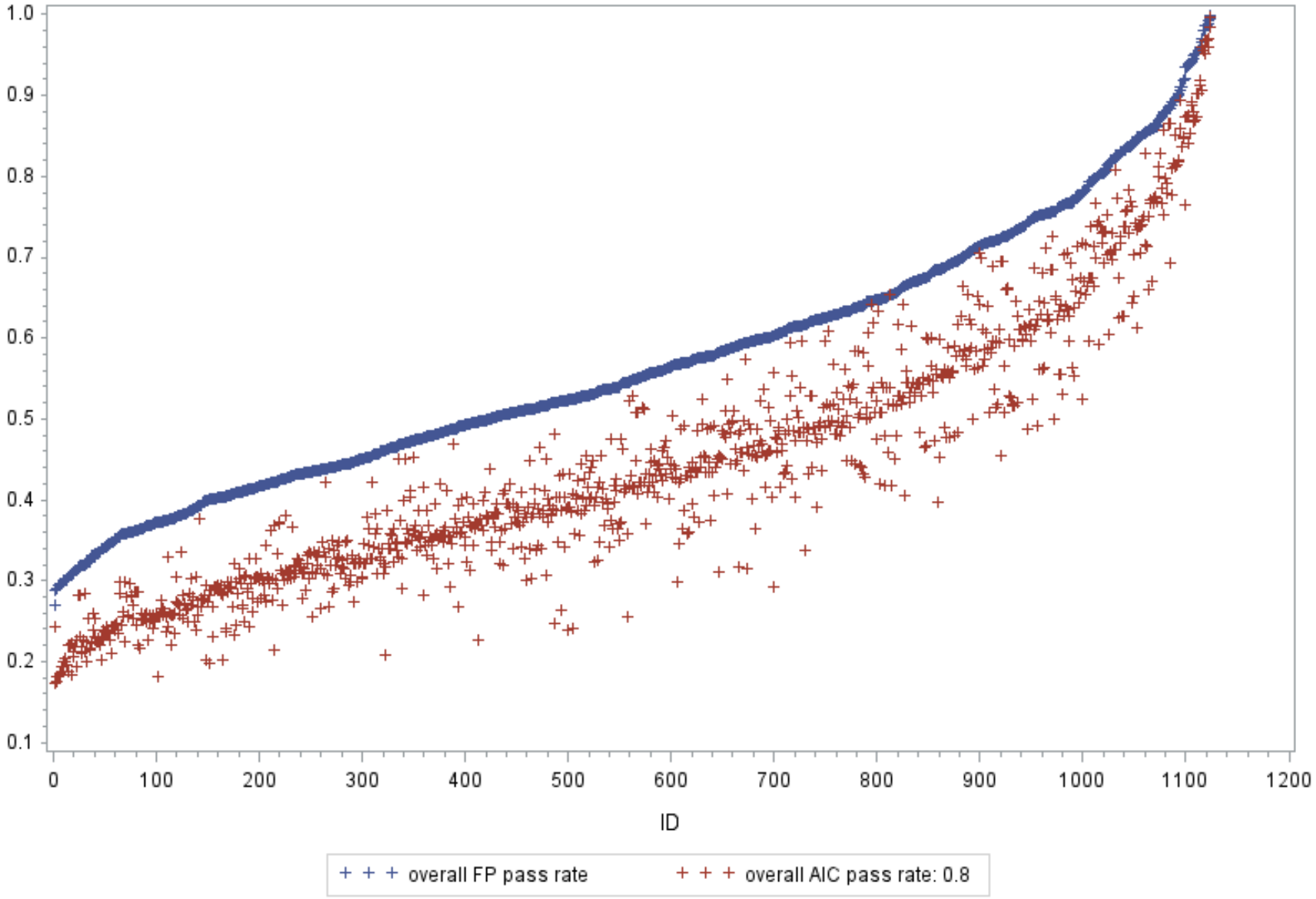
- **A variety of scenario** with different parameter specifications under **2×2 crossover design** were considered. (each point in the plot represents a scenario)
- 100,000 repetitions
- Overall type I error and power were compared: regarding testing

  $\{\delta_L \leqslant \mu_I - \mu_R \leqslant \delta_U,\ I = A, B \text{ and } \delta'_L \leqslant \mu_A - \mu_B \leqslant \delta'_U\}$.
- The adjusted indirect comparison (AIC) method with $\delta'_L = log(0.75)$ (green) is inappropriate regarding the type I error. The proposed method (FP: blue) has larger power than the AIC method with $\delta'_L = log(0.8)$ (red).


Overall Type I Error
GMR was within [0.79,0.8], and the product of two trilas' powers was within [0.64,1]
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
ID
overall FP pass rate
overall AIC pass rate: 0.8
overall AIC pass rate: 0.75


Overall Power
GMR was within [0.95,1/0.95], and the product of two powers was within [0.64,1]
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
ID
overall FP pass rate
overall AIC pass rate: 0.8


Overall Power
GMR was within [0.9,0.95], and the product of two powers was within [0.64,1]
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
ID
overall FP pass rate
overall AIC pass rate: 0.8


Overall Power
GMR was within [0.85,0.9], and the product of two powers was within [0.64,1]
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
ID
overall FP pass rate
overall AIC pass rate: 0.8

Real example analysis (malaria generics)

- Three bioequivalence studies conducted independently (WHO public assessment reports)
- Fixed dose artemether/lumefantrine 20/120 mg tablets (Coartem®/Riamet®) from Novartis Pharma (Basel, Switzerland)
- 55, 64, and 58 adult men
- Single-center, open-label, randomized, two-period, two-treatment, two-sequence, crossover studies under non-fasting conditions
- Three measures: C_{max}, $AUC_{0\text{-}t}$, $AUC_{0\text{-}inf}$
- The Benjamini–Hochberg procedure (the target FDR of 0.1)

Real example analysis (malaria generics)

Table IV. The 90% confidence interval of pharmacokinetic parameters of artemether and lumefantrine in fixed dose combination generics, the fiducial probability with $\delta_U' = \delta_U = \log(1.25)$, and the Benjamini–Hochberg-adjusted p-value

		The 90% confidence interval		
		Ajanta	Ipca	Cipla
Artemether	C_{max}	86.3–99.1	101.0–118.9	86.7–103.7
	AUC_{0-t}	93.3–104.7	102.4–116.6	90.1–106.9
	AUC_{0-inf}	93.5–104.6	102.8–117.0	90.1–106.9
Lumefantrine	C_{max}	80.3–95.4	87.6–99.5	89.3–108.2
	AUC_{0-t}	82.1–97.3	87.3–103.2	88.7–108.1
	AUC_{0-inf}	82.5–97.8	87.6–103.0	88.7–108.1
		The fiducial probability (the adjusted p-value)		
		Ajanta versus Ipca	Ajanta versus Cipla	Ipca versus Cipla
Artemether	C_{max}	0.795 (0.2048)*	0.996 (0.0238)	0.857 (0.1517)*
	AUC_{0-t}	0.989 (0.0255)	0.999 (0.0055)	0.962 (0.0623)
	AUC_{0-inf}	0.989 (0.0255)	0.999 (0.0055)	0.957 (0.0623)
Lumefantrine	C_{max}	0.955 (0.0623)	0.894 (0.1197)*	0.992 (0.0255)
	AUC_{0-t}	0.975 (0.0459)	0.943 (0.0688)	0.991 (0.0255)
	AUC_{0-inf}	0.980 (0.0405)	0.950 (0.0642)	0.992 (0.0255)

*Not significant by the type 1 method with the target false discovery rate of 0.1.

Real example analysis (HIV/AIDS generics)

- Combivir® (lamivudine/zidovudine) 150 mg/300 mg tablet: an antiviral medication
- Reverse transcriptase inhibitors and helps keep the HIV virus from reproducing in human body
- Indicated for the treatment of HIV-1 infection in combination with at least one other antiretroviral agent
- 62, 31, and 43 subjects
- C_{max}, $AUC_{0\text{-}t}$, $AUC_{0\text{-}inf}$
- Randomized, open label, two-treatment, two-period, two-sequence, single-dose, and crossover designs under fasting conditions

Real example analysis (HIV/AIDS generics)

Table V. The 90% confidence interval of pharmacokinetic parameters of lamivudine and lidovudine in fixed dose combination generics, the fiducial probability with $\delta_U' = \delta_U = \log(1.25)$, and the Benjamini–Hochberg-adjusted p-value

		The 90% confidence interval		
		Ranbaxy	Strides	Matrix
Lamivudine	C_{max}	96.8–109	90.63–110.52	82.4–101.7
	AUC_{0-t}	97.8–105	96.31–108.29	88.0–100.6
	AUC_{0-inf}	98.2–105	96.42–107.92	88.8–100.7
Zidovudine	C_{max}	85.5–108	81.15–112.51	83.8–114.1
	AUC_{0-t}	93.7–102	92.77–105.41	97.3–109.7
	AUC_{0-inf}	93.7–102	92.63–105.15	97.4–109.5
		The fiducial probability (the adjusted p-value)		
		Ranbaxy versus strides	Ranbaxy versus matrix	Strides versus matrix
Lamivudine	C_{max}	0.996 (0.0057)	0.928 (0.0834)	0.929 (0.0834)
	AUC_{0-t}	1.000 (0.0000)	0.999 (0.0016)	0.995 (0.0072)
	AUC_{0-inf}	1.000 (0.0000)	1.000 (0.0010)	0.997 (0.0041)
Zidovudine	C_{max}	0.909 (0.0966)	0.926 (0.0834)	0.866 (0.1344)*
	AUC_{0-t}	1.000 (0.0001)	1.000 (0.0004)	0.999 (0.0012)
	AUC_{0-inf}	1.000 (0.0001)	1.000 (0.0004)	0.999 (0.0012)

*Not significant by the type 1 method with the target false discovery rate of 0.1.

Conclusion

- Compared to the existing methods, proposed methods have two aspects of advantages: clinical meaningful and more power.
- Extension to simultaneous comparison of three generics and multiple testing
- **Similarity assumptions**
- Further research: average bioequivalence → population bioequivalence and individual bioequivalence

Reference

1. Chow SC, Liu JP. *Design and Analysis of Bioavailability and Bioequivalence Studies* 3rd ed. CRC Press, 2009.

2. FDA. *Guidance for Industry: Statistical Approaches to Establishing Bioequivalence*. Center for Drug Evaluation and Research, U.S. Food and Drug Administration: Rockville, MD, 2001.

3. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* 1987; **15**(6):657–680.

4. Chow SC, Endrenyi L, Lachenbruch PA, Mentre F. Scientific factors and current issues in biosimilar studies. *Journal of Biopharmaceutical Statistics* 2014; **24**:1138–1153.

5. Chow SC, Song FY, Chen M. Some thoughts on drug interchangeability. *Journal of Biopharmaceutical Statistics* 2016; **26**:178–186.

6. Anderson S, Hauck WW. The transitivity of bioequivalence testing: potential for drift. *International Journal of Clinical Pharmacology and Therapeutics* 1996; **34**(9):369–374.

7. Bialer M, Midha KK. Generic products of antiepileptic drugs (AEDs): a perspective on bioequivalence and interchangeability. *Epilepsia* 2010; **51**:941–950.

8. Privitera M. Generic antiepileptic drugs: current controversies and future directions. *Epilepsy Curr* 2008; **8**:113–117.

9. van Gelder T, Gabardi S. Methods, strengths, weaknesses, and limitations of bioequivalence tests with special regard to immunosuppressive drugs. *Transplant Int* 2013; **26**(8):771–777.

10. FDA. *Draft Guidance for Industry: Bioequivalence Studies with Pharmacokinetic Endponits for Drugs Submitted under an ANDA*. Center for Drug Evaluation and Research, U.S. Food and Drug Administration: Rockville, MD, 2013.

11. Song F. *What is indirect comparison?*, February 2009.

12. Garca-Arieta A, Potthast H, Leufkens H,Welink J, et al. Assessment of the interchangeability between generics. *Generics and Biosimilars Initiative Journal (GaBI Journal)* 2016; **5**(2):55–9.

13. Gwaza L, Gordon J, Welink J, Potthast H, et al. Statistical approaches to indirectly compare bioequivalence between generics: a comparison of methodologies employing artemether/lumefantrine 20/120 mg tablets as prequalified by WHO. *European Journal of Clinical Pharmacology* 2012; **68**:1611.

14. Endrenyi L, Tóthfalusi L. Adjusted indirect comparisons between generics – bioequivalence and interchangeability. *Generics and Biosimilars Initiative Journal (GaBI Journal)* 2016; **5**(2):53–4.

15. Fisher RA. The fiducial argument in statistical inference. *Annals of Human Genetics* 1935; **6**(4):391–398.

16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995; **57**(1):289–300.

17. Benjamini Y. Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2010; **72**(4):405416.

18. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 2001; **29**:1165–1188.

19. Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**(3):479–498.

20. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *PNAS, Proceedings of the National Academy of Sciences* 2003; **100**(16):9440–9445.

Thanks!

Questions