ICH M13A: Testing for multi-group and multi-center effects in bioequivalence

Statistical considerations and consequences for interpretation

Helmut Schütz
Center for Medical Data Science of the Medical University of Vienna
Faculty of Pharmacy, Universidade de Lisboa, Lisbon
BEBAC, Vienna

ICH M13A: 2.2.3.1 General Considerations

The statistical analysis should take into account sources of variation that can be reasonably assumed to have an effect on the response variable.

- Makes sense, of course.
- However, *is it reasonable* to assume that groups of subjects
  - in the same clinical site,
  - dosed within short interval(s) – quite often less than a week apart,
  - samples analyzed with the same method at the same lab …
- … would differ in their PK outcomes?
- IMHO, that's an insult to the mind.

ICH M13A: 2.2.3.5 Multi-Group Design Studies

BE should be determined based on the overall treatment effect in the whole study population. In general, the assessment of BE in the whole study population should be done without including the Group by Treatment interaction term in the model […] However, the appropriateness of the statistical model should be evaluated to account for the multi-group nature of the BE study. Applicants should evaluate potential for heterogeneity of treatment effect across groups, i.e., Group by Treatment interaction. If the Group by Treatment interaction is significant, this should be reported and the root cause of the Group by Treatment interaction should be investigated to the extent possible.

ICH M13A: 2.2.3.5 cont'd

Substantial differences in the treatment effect for PK parameters across groups should be evaluated. Further analysis and interpretation may be warranted in case heterogeneity across groups is observed.

- Which difference might be *substantial*?
- Is assessement of a Group by Treatment interaction *optional*, *recommended*, or *mandatory*?
- Testing the Group by Treatment interaction at *which* level (0.1 or 0.05)?
- *Which* extent of the 'root cause analysis' might be considered *acceptable*?

Crossover models

Interaction model (I)

Y ~ Group, Sequence, Subject(Group × Sequence), Period(Group), Group × Sequence, Treatment, **Group × Treatment**

Group model (II)

Y ~ Group, Sequence, Subject(Group × Sequence), Period(Group), Group × Sequence, Treatment

Conventional (III)

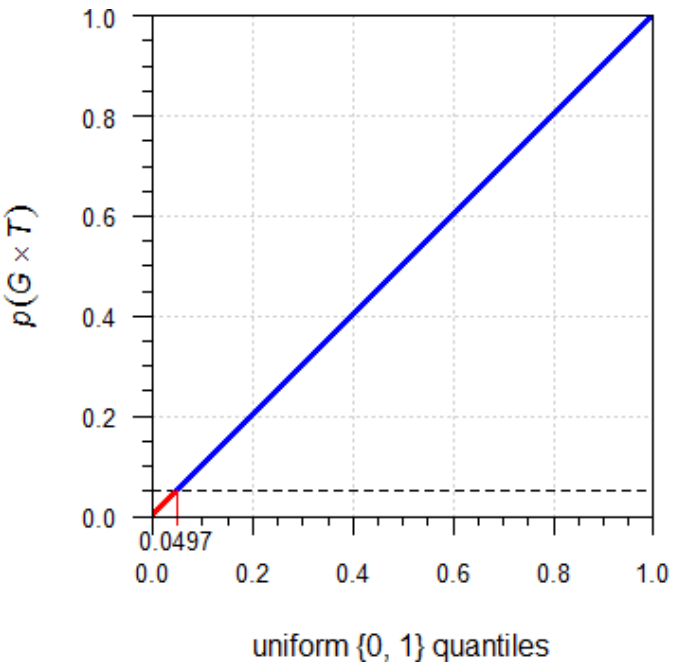
Y ~ Sequence, Subject(Sequence), Period, Treatment

In the interaction model (I) unbiased estimate of the treatment effect is not possible!

10^5 simulated studies (homoscedasticity)

No Group by Treatment interaction:

Group	n	GMR	CV_w
1	24	1.0000	0.335
2	24	1.0000	0.335
	48	1.0000	0.355

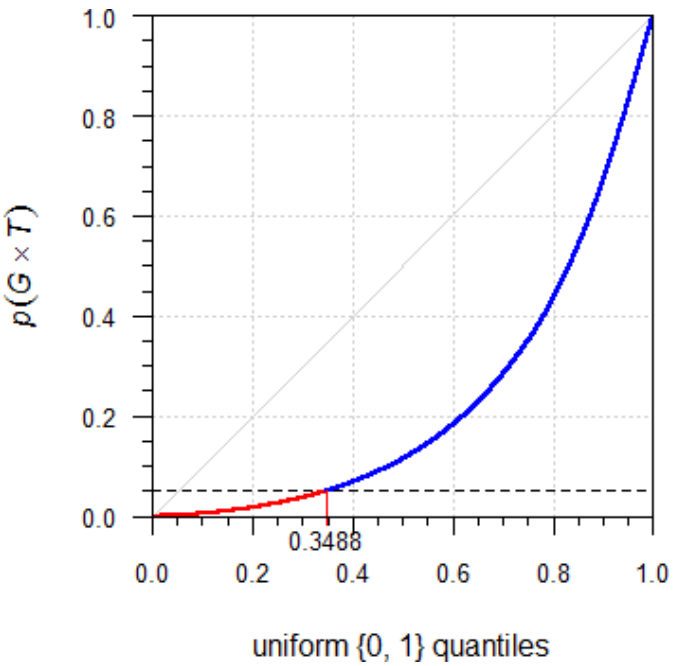


- Significant *G* × *T* interaction detected in 4.97% of simulated studies.
- At the level (0.05) of the test → **false positives!**
- 'Root cause analysis' of an effect that might happen by *pure chance* is useless.

10^5 simulated studies (heteroscedasticity)

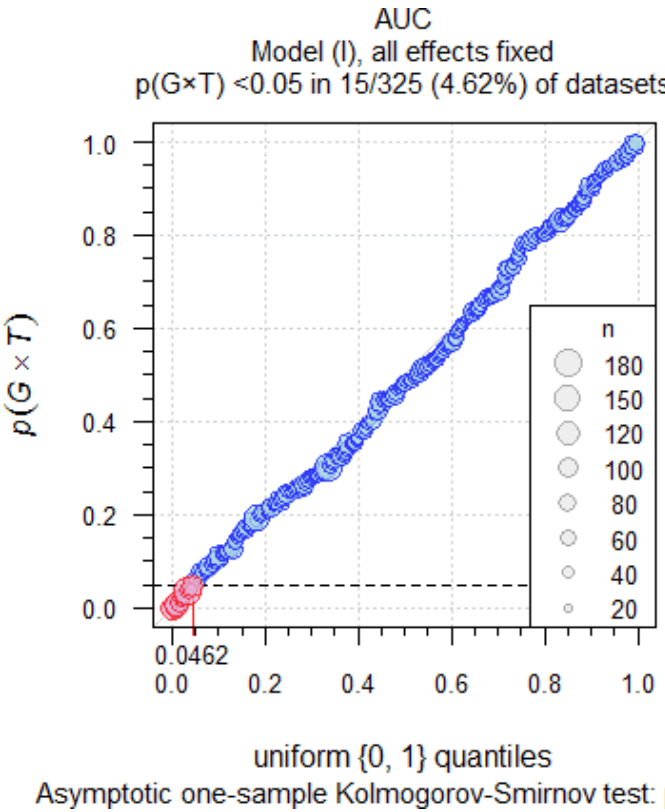
True Group by Treatment interaction:

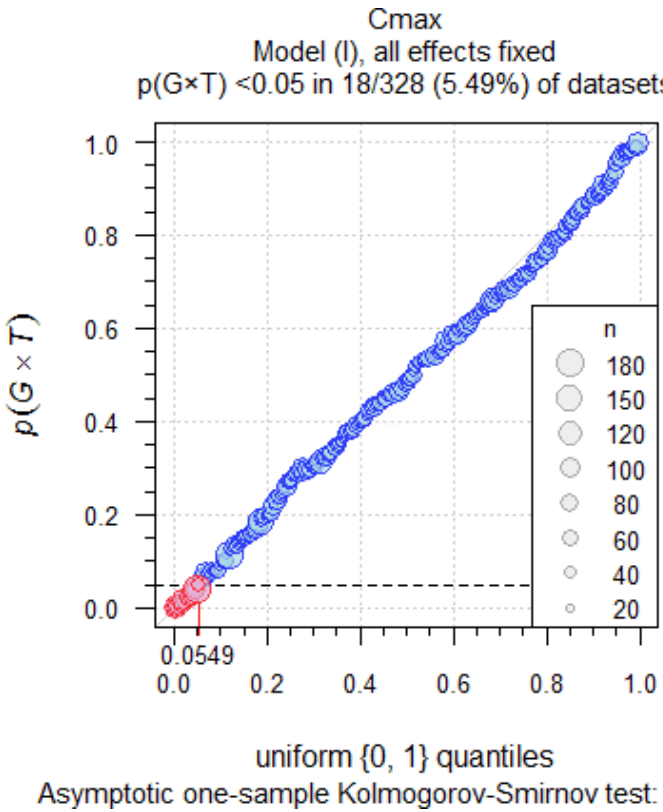
Group	n	GMR	CV_w
1	38	1.0605	0.369
2	10	0.8000	0.298
	48	1.0000	0.355



- Significant *G* × *T* interaction detected in 34.9% of simulated studies.
- As expected, well above the level of the test.
- In 65.1% of simulated studies the true *G* × *T* interaction **is not detected!**

Meta–study (325 datasets AUC, 328 C_{max})





$G \times T$ interaction 'detected' at approximately the level (0.05) of the test; in well-controlled trials likely false positives.

ICH M13A: 2.2.3.5 cont'd

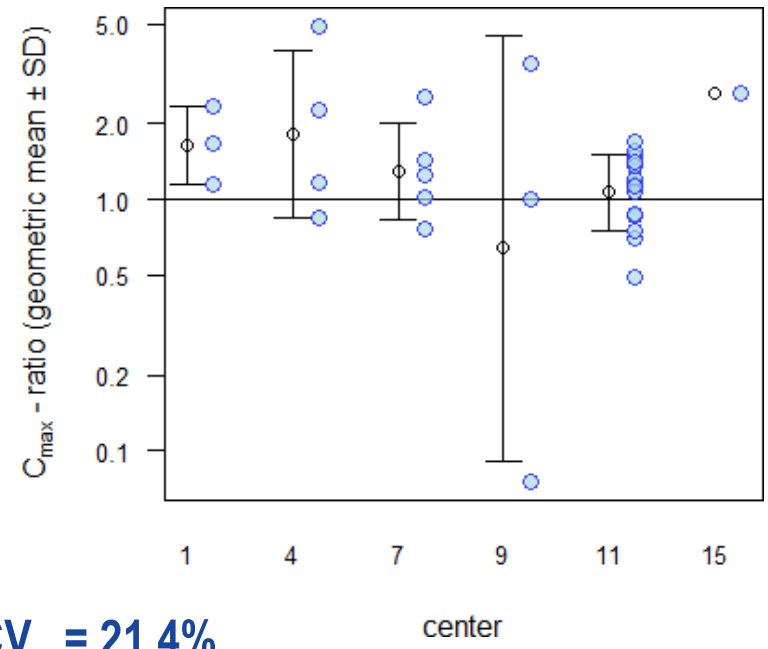
In multicentre BE studies, when there are very few subjects in some sites, these subjects may be pooled into one group for consideration in the statistical analysis. Rules for pooling subjects into one group should be pre-specified and a sensitivity analysis is recommended.

- No specific model recommended in the guideline.
- Sensitivity analysis of *what*?
  - Means and / or variances of centers' PK responses?
  - Exploratory by graphical methods (scatter or box plots) sufficient?
  - Statistical test?
    - *p*-value of Center by Treatment interaction (Model I)?
    - Level of the test (0.1 or 0.05)?

Case study (6 centers)

Extreme C_{max}-ratios and / or between-subject variances only in small centers.

- Per protocol analysis by Model III
  121.01% (96.13 – 152.33%),
  n = 30, CV_w = 55.6%
- Significant Center by Treatment interaction (Model I *p* = 0.0006)
- Analysis of the largest center
  103.80 (89.87 – 119.90%), n = 14, CV_w = 21.4%
- Root cause analysis
  - Likely improper sample handing in small centers → stability problems.
  - Since no deviations documented, not accepted in referral (2018).



Conclusions

- Inclusion of a group-term may substantially compromise power (in the meta-analysis by 6%); it is impossible to detect a true Group by Treatment interaction by statistics, i.e., a subsequent ‘investigation of a root cause’ is futile.
- Statistically significant does not imply clinically relevant.
- Multi-center studies are problematic – should be avoided if ever possible.

The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.

John W. Tukey

A mathematician is a blind man in a dark room looking for a black cat which isn’t there.

attr. to Charles Darwin

ICH M13A: Testing for multi-group and multi-center – effects in bioequivalence

Helmut Schütz
Center for Medical Data Science
MEDICAL UNIVERSITY OF VIENNA
1090 Vienna, Austria
helmut.schuetz@meduniwien.ac.at
Faculty of Pharmacy, Universidade de Lisboa
1649-004 Lisbon, Portugal
helmut@edu.ulisboa.pt
BEBAC
1070 Vienna, Austria
helmut.schuetz@bebac.at

»Λόγον ἔχεις;« »ἔχω.« »τί οὖν οὐ χρᾷ;«

References

1. Alosh M, Fritsch K, Huque M, Mahjoob K, Pennello G, Rothmann M, Russek-Chen E, Smith E, Wilson S, Yiu L. *Statistical Considerations on Subgroup Analysis in Clinical Trials*. Stat Pharm Res. 2015; 7(4): 286–304. https://doi.org/10.1080/19466315.2015.1077726.
2. Sun W, Schuirmann D, Grosser S. *Qualitative versus Quantitative Treatment-by-Subgroup Interaction in Equivalence Studies with Multiple Subgroups.* Stat Pharm Res. 2022; Early View 31 Oct 2022. https://doi.org/10.1080/19466315.2022.2123385.
3. Glerum PJ, Neef C, Burger DM, Yu Y, Maliepaard M. *Pharmacokinetics and Generic Drug Switching: A Regulator's View*. Clin Pharmacokin. 2020; 59: 1065–9. https://doi.org/10.1007/s40262-020-00909-8.
4. Benet LZ. *Why Do Bioequivalence Studies in Healthy Volunteers?* 1st MENA Regulatory Conference on Bioequivalence, Biowaivers, Bioanalysis and Dissolution. Amman, Jordan, 23 September 2013. https://web.archive.org/web/20220818171344/http://www.rbbbd.com/pdf/presentations/Leslie%20Z%20Benet/lecture%201%20Why%20Do%20Bioequivalence%20Studies%20%20in%20Healthy%20Volunteers.pdf.
5. Bolton S, Bon C. *Pharmaceutical Statistics. Practical and Clinical Applications*. New York: informa healthcare; 5th edition 2010. p. 629.
6. Bae K-S, Kang S-H. *Bioequivalence data analysis for the case of separate hospitalization*. Transl Clin Pharmacol. 2017; 25(2): 93–100. https://doi.org/10.12793/tcp.2017.25.2.93.
7. ICH. *Bioequivalence for Immediate Release Solid Oral Dosage Forms. M13A*. Draft version. 20 December 2022. https://database.ich.org/sites/default/files/ICH_M13A_Step2_draft_Guideline_2022_1125.pdf#page=20.
8. FDA, CDER. *Statistical Approaches to Establish Bioequivalence. Guidance for Industry*. Revision 1 Draft. Silver Spring. December 2022. https://www.fda.gov/media/163638/download#page=25.
9. Schütz H. *Group 'Effect'. To Pool or Not to Pool?* September 18, 2023. https://bebac.at/articles/Group-Effect.phtml.