



MEDICAL UNIVERSITY  
OF VIENNA



# Sex- and group-related problems in BE. A delusion.

Helmut Schütz  
Center for Medical Data Science of  
the Medical University of Vienna  
BEBAC, Vienna



Wikimedia Commons • 2007 Sokoljan • Creative Commons SA 3.0 Unported



# ICH M13A: 2.1.1 Study Population

**Subjects should be at least 18 years of age and preferably have a Body Mass Index between 18.5 and 30.0 kg/m<sup>2</sup>. [...]**

**Subjects should preferably be non-nicotine users [...]**

- No concerns about extrapolating to patients < 18 years.
- No concerns about extrapolating to obese patients (in the U.S. ~42% of the adult population).
- No concern about extrapolating to 23% of adults smoking tobacco.

**If a drug product is intended for use in both sexes, it is recommended the study include male and female subjects.**

- Concerns about a Sex-by-Formulation interaction, i.e., questioning extrapolating from a study performed in either sex to the general population?

# Meta-study (235 mixed-sex datasets)

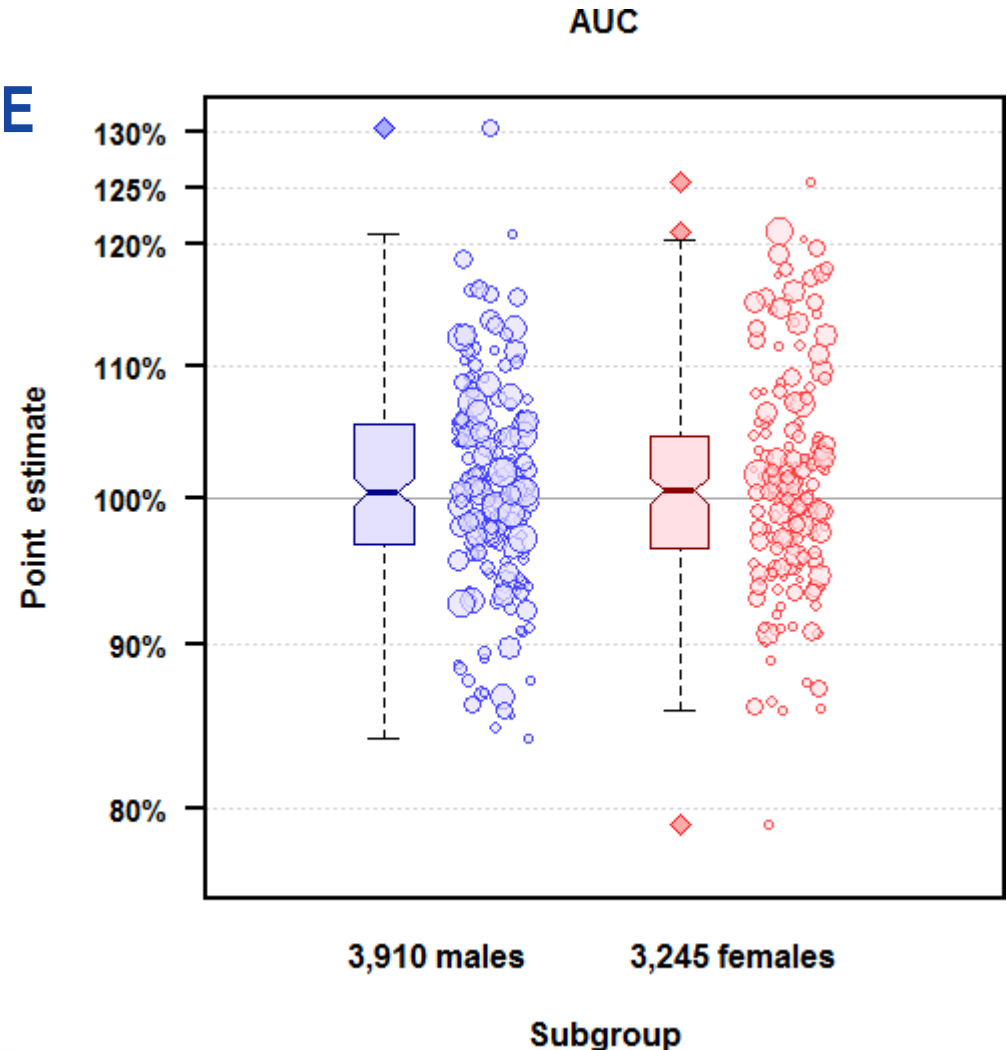
185 datasets passing BE

$n$   $\tilde{x}$  30 (12 – 117)

$f_{\text{♂}}$   $\tilde{x}$  52.9% (31.9 – 87.5%)

$f_{\text{♀}}$   $\tilde{x}$  47.1% (12.5 – 68.1%)

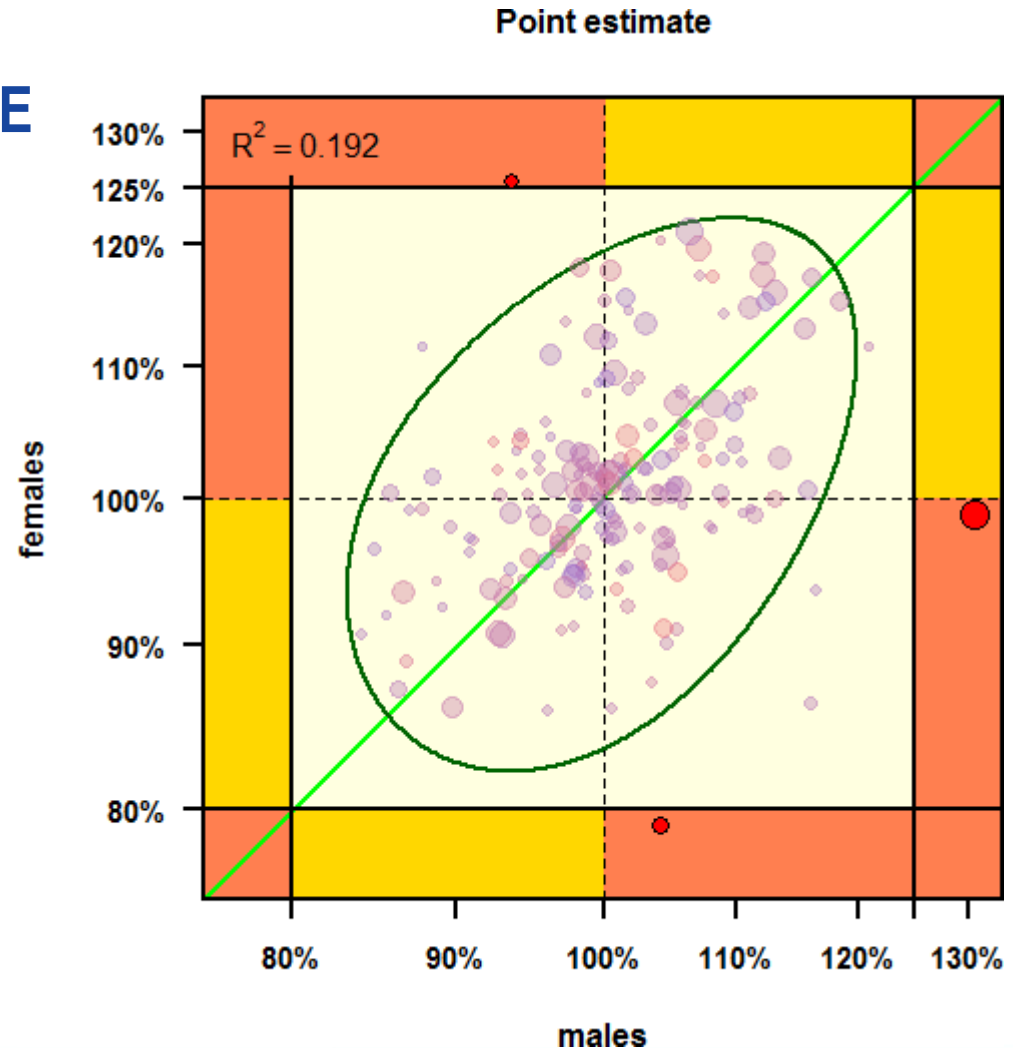
- No evidence that medians of subgroups differ (i.e., notches of box plots overlap).
- Similar within-subject CV of males ( $\tilde{x}$  12.53%) and females ( $\tilde{x}$  12.61%).



# Meta-study (235 mixed-sex datasets)

## 185 datasets passing BE

- Difference in PEs of males and females  $> \pm 20\%$  in 3.24% of datasets.
- Difference in PEs of males and females  $\leq \pm 10\%$  in 77.8% of datasets.
- Discordant Qualitative Interaction in 1.62% of datasets.



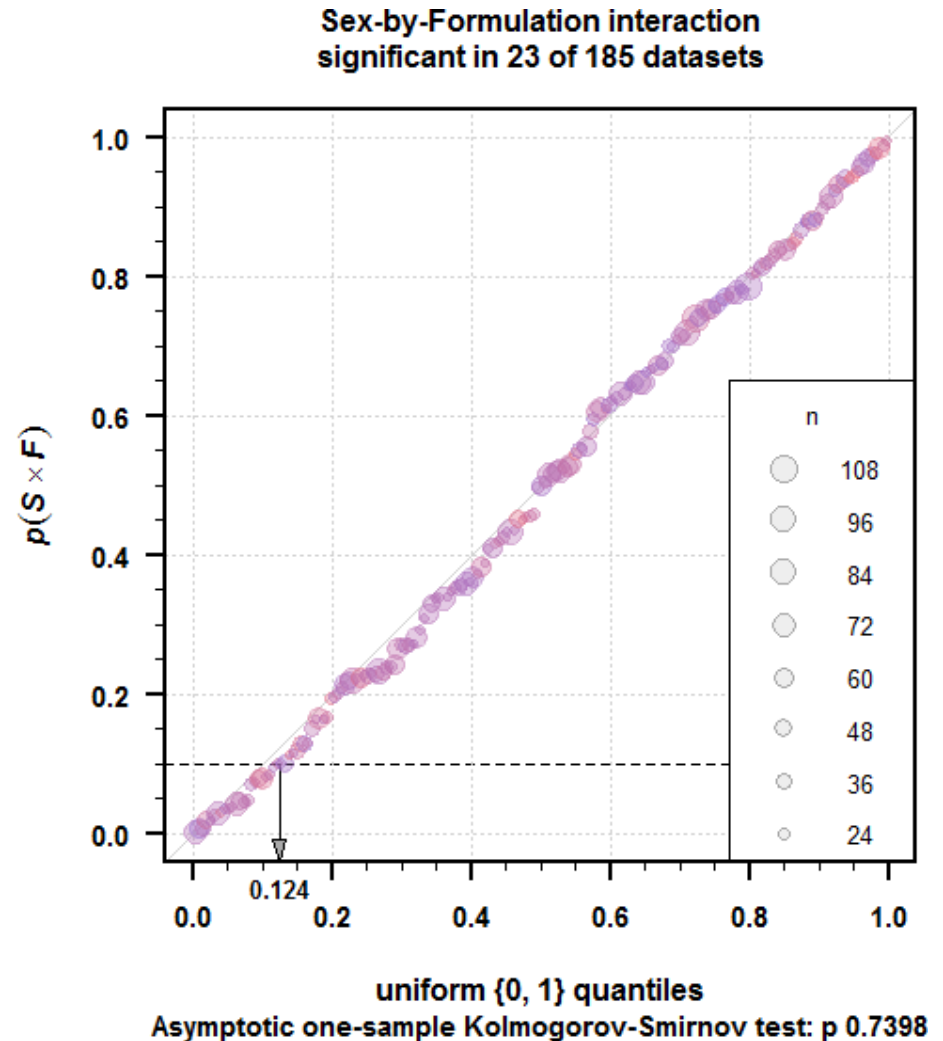
# Meta-study (235 mixed-sex datasets)

## 185 datasets passing BE

- Significant ( $p < 0.1$ )  $S \times F$  interaction in 12.4% of datasets.

Confirmed conclusions of González-Rojano *et al.* 2019

“There is no evidence to require studies in both sex groups, combined or separately.



# 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.**

- Is assessement of a Group by Treatment interaction mandatory?
- Significance level 0.1 or 0.05?
- Which difference might be 'substantial'?

## **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.**

# Crossover models

## Interaction model (I)

$Y \sim \text{Group, Sequence, Subject}(\text{Group} \times \text{Sequence}),$   
 $\text{Period}(\text{Group}), \text{Group} \times \text{Sequence}, \text{Treatment},$   
 $\text{Group} \times \text{Treatment}$

## Group model (II)

$Y \sim \text{Group, Sequence, Subject}(\text{Group} \times \text{Sequence}),$   
 $\text{Period}(\text{Group}), \text{Group} \times \text{Sequence}, \text{Treatment}$

## Conventional (III)

$Y \sim \text{Sequence, Subject}(\text{Sequence}), \text{Period}, \text{Treatment}$

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

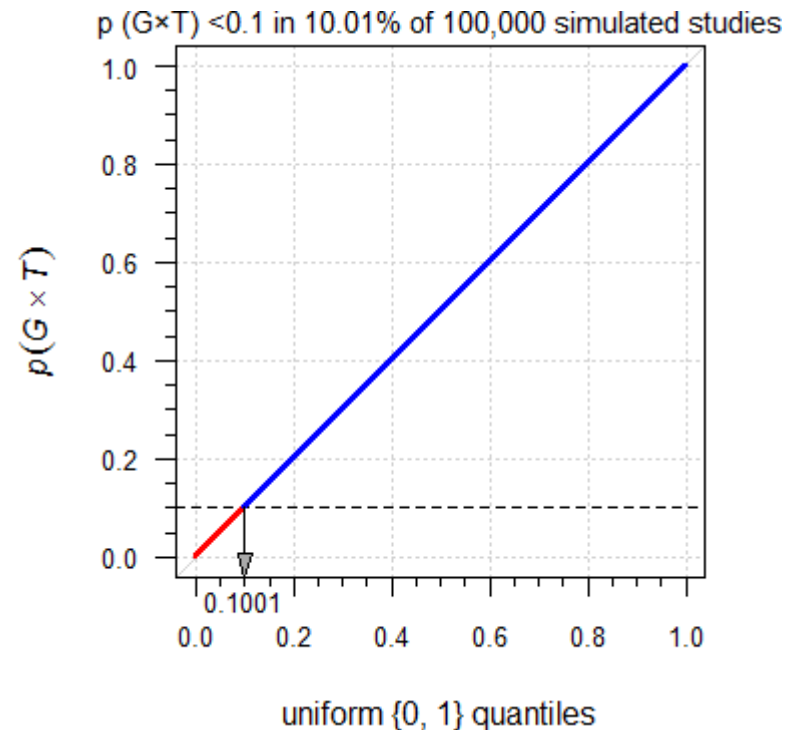


# $10^5$ simulated studies ( $n_1 = n_2 = 24$ )

**No Group by Treatment interaction:**

$$\text{GMR}_1 = \text{GMR}_2 = 1.0000$$

- Significant  $G \times T$  interaction in 10.01% of simulated studies.
- At the level of the test → **false positives!**

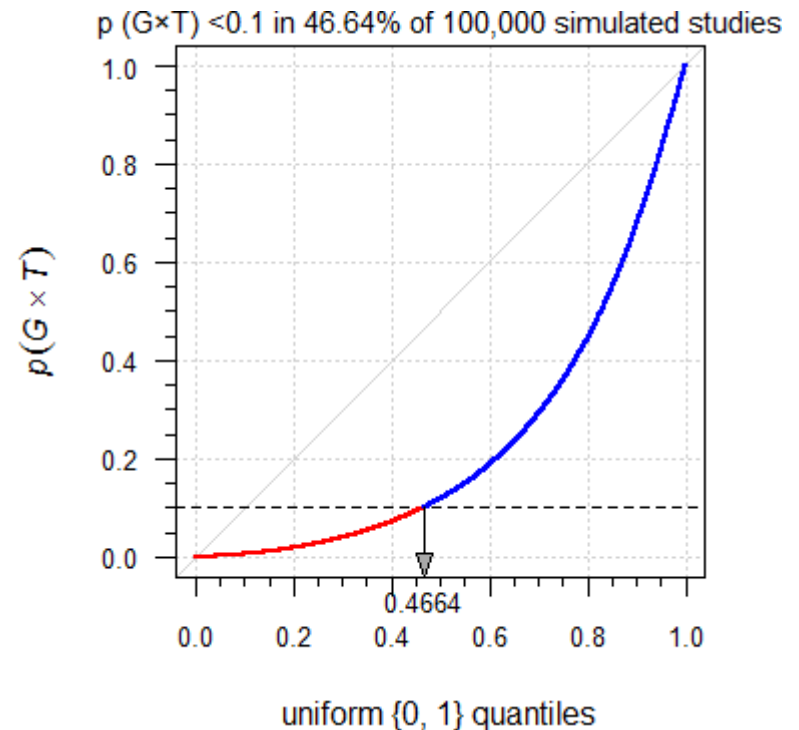


# $10^5$ simulated studies ( $n_1 = n_2 = 24$ )

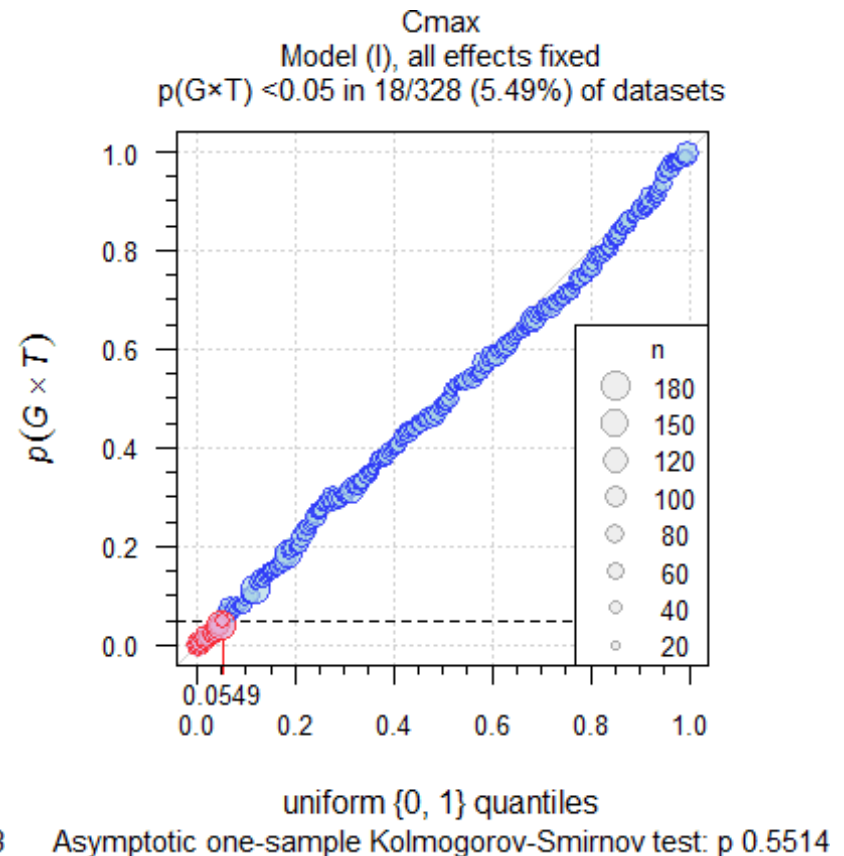
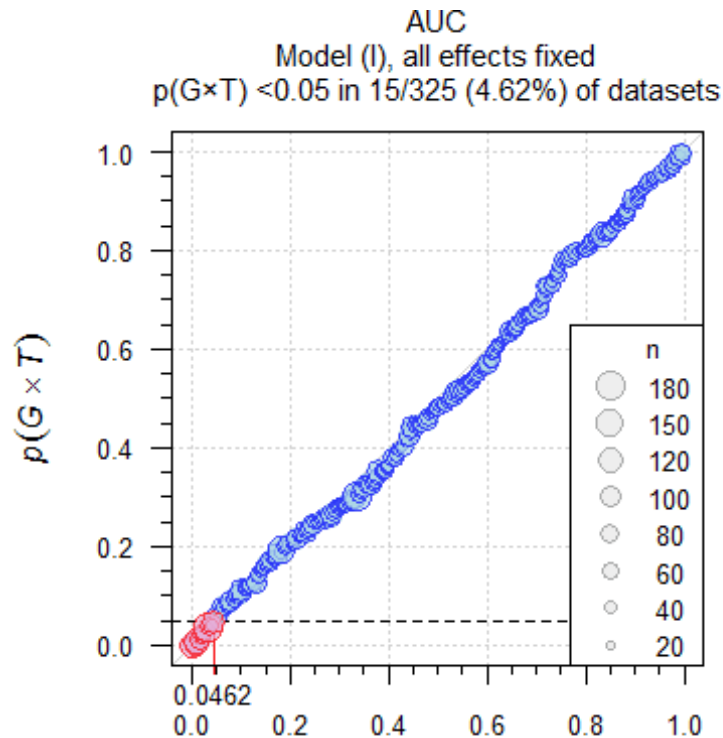
**True Group by Treatment**  
interaction:

$GMR_1 = 0.9000$ ,  $GMR_2 = 1.1111$   
(pooled GMR 1.0000)

- Significant  $G \times T$  interaction in 46.64% of simulated studies.
- As expected, high above the false positive rate.
- But: In 53.36% the true  $G \times T$  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.

# Conclusions

- No empiric evidence that extrapolation of results from studies in healthy subjects (in either sex...) to the patient population is problematic.
- Inclusion of a Group-term may substantially compromise power.
- Impossible to detect a true Group-by-Treatment by statistics, i.e., subsequent 'investigation of a root cause' is futile.

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

# Sex– and group–related problems in BE. A delusion.

Thank You!



**Helmut Schütz**

Center for Medical Data Science



MEDICAL UNIVERSITY  
OF VIENNA

1090 Vienna, Austria

[helmut.schuetz@meduniwien.ac.at](mailto:helmut.schuetz@meduniwien.ac.at)

**BEBAC**

1070 Vienna, Austria

[helmut.schuetz@bebac.at](mailto:helmut.schuetz@bebac.at)

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