Non-inferiority clinical trials: a multivariate test for multivariate PD

C.M. Laffont [1], M. Fink [2], D. Concordet [1]

[1]INRA, UMR 1331, Toxalim, Toulouse, France;
Université de Toulouse, INPT, ENVT, UPS, EIP, Toulouse, France
[2]Novartis Pharma AG, Basel, Switzerland

Problem

Endpoint 1

Endpoint 2

Endpoint 3

… Endpoint K

Clinician

Multiple PD endpoints are a common feature of clinical trials

Ex: rheumatoid arthritis (ACR), Alzheimer’s Disease (ADAS-cog), schizophrenia (PANSS), depression (HAMD)

Problem

- The objective of the trial is to compare a test drug with a positive or negative control by doing a statistical test

Endpoint 1 **Endpoint 2**
Endpoint 3 **.. Endpoint K**

Statistician

"Summary" variable

- Binary variables (responders: yes/no)
- Sum of scores (categorical)
- A function of "continuous" responses

Problem

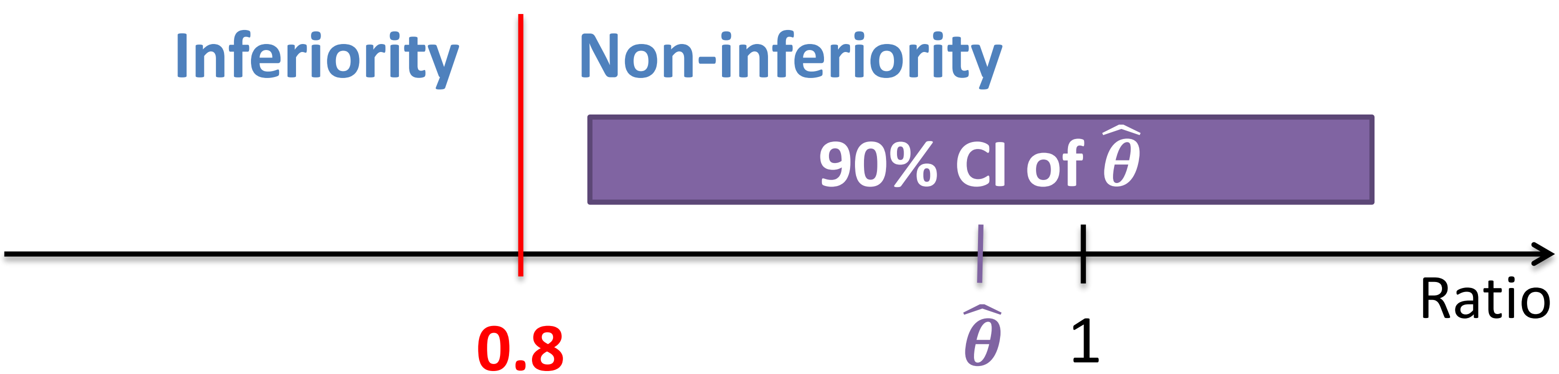
- The problem of the **"summary" variable** is the inevitable **loss of information** associated with the reduction in dimension
- So it seems that keeping all endpoints for the analysis is more appropriate
- Generally, endpoints are not all continuous variables but include categorical data (binary, ordinal, counts), which increases the complexity of the analysis
- In that case, a modeling approach allows to recover the "continuous case", which increases power

Problem

- Multiple endpoints: how to test drug efficacy?
- Here, we will focus on **non-inferiority analysis** which is the most common analysis when a test drug and a positive control are compared
- What does non-inferiority means?
    - In one dimension (one endpoint)
    - In multiple dimensions (multiple endpoints)

Non-inferiority in one dimension

- Let θ be the ratio of effects (drug/control). In case of identical effects, $\theta = 1$
- **Hypotheses:** $H_0 : \theta <$ **non-inferiority margin** (here 0.8)
  $H_1 : \theta \geq$ **non-inferiority margin**
- α is the risk to wrongly conclude non-inferiority (5%)
- $\hat{\theta}$ is the sample estimate. Non-inferiority is concluded when its 90% confidence interval (CI) is above 0.8



Non-inferiority in $K > 1$ dimensions

- Now imagine that we have ***K* endpoints**
- For each endpoint k, θ_k is the effect ratio and $H_{0,k}$ is the null hypothesis
- Global null hypothesis: **2 possible definitions of inferiority**

$$H_0 = \bigcup_k H_{0,k}$$

"Union": non-inferiority must be demonstrated on ***all*** endpoints

$$H_0 = \bigcap_k H_{0,k}$$

"Intersection": non-inferiority must be demonstrated on ≥ 1 endpoint

Non-inferiority in $K > 1$ dimensions

2 strategies

Multiple univariate tests

- Test each endpoint separately
- Compile the results of univariate analyses

Single multivariate test

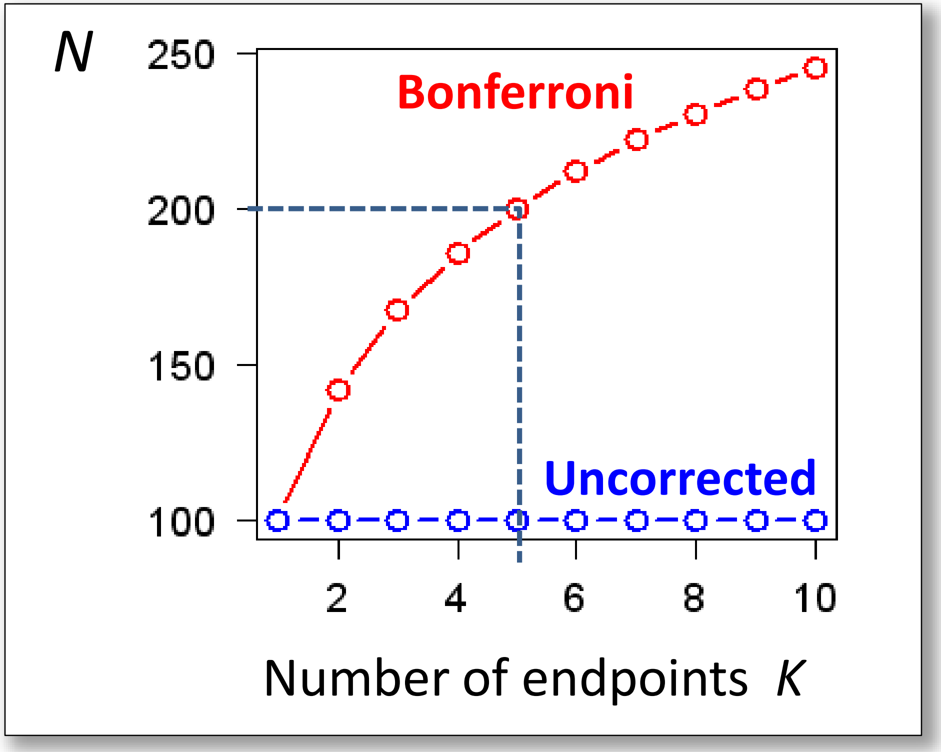
Objective: evaluate the gain in power for a single multivariate test vs. the compilation of univariate tests

First strategy: multiple univariate tests

- We need to penalize for the multiplicity of the tests to keep a global α risk of 5%
- Bonferroni correction: we use $\dfrac{\alpha}{K}$ instead of α

CI are larger than without Bonferroni correction, which requires to increase the number of subjects N

$N \times 2$ when $K = 5$ endpoints



Second strategy: single multivariate test

- Instead of multiple univariate CIs, we compute **a multivariate confidence region**
- We assume that the sample size is sufficiently large so that the estimator $\hat{\theta}$ is **normally distributed** (common assumption in pop PK/PD)

$$\hat{\theta} \sim N\left(\theta, Var(\hat{\theta})\right)$$

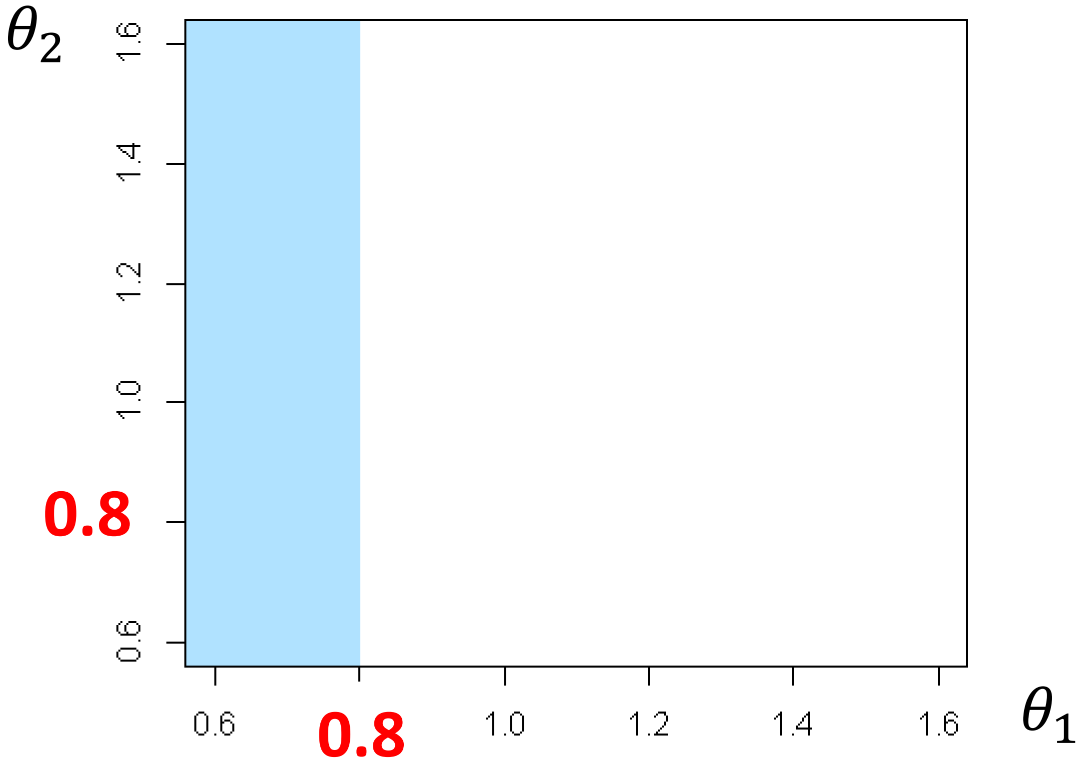
$$Var(\hat{\theta}) \approx \frac{1}{N} I(\theta)^{-1}$$

I : Fisher Information

- The 90% confidence region is an ellipsoid with equation:

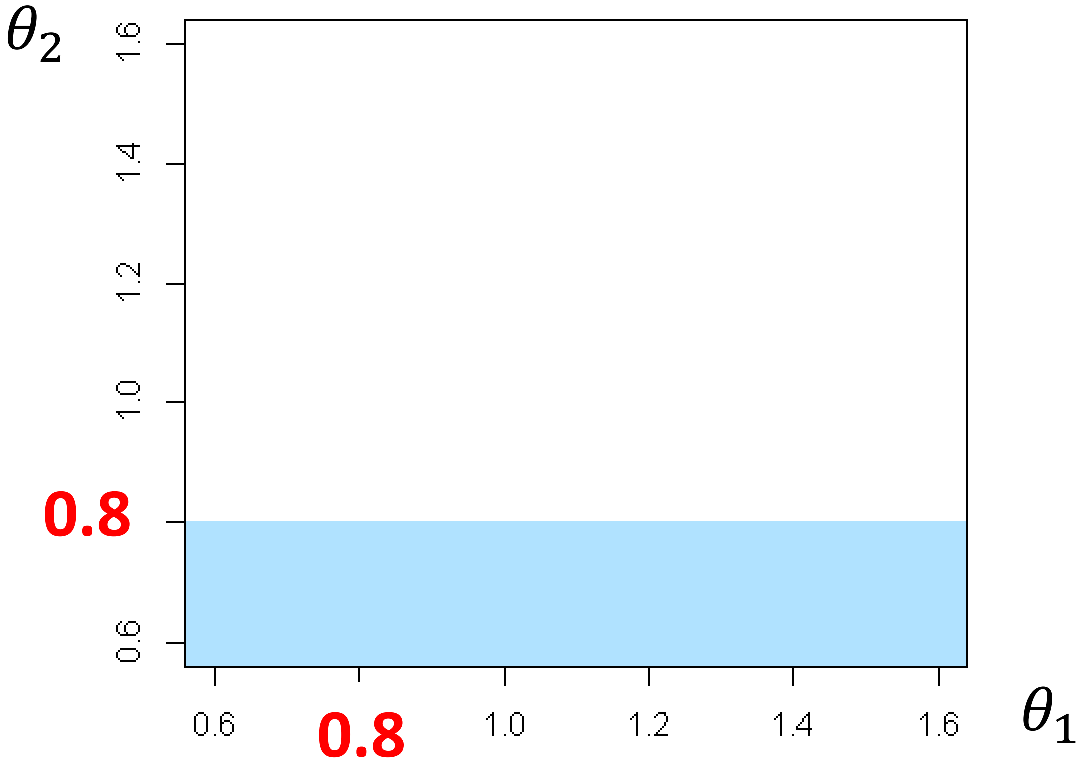
$$\left(\theta - \hat{\theta}\right)' Var\left(\hat{\theta}\right)^{-1} \left(\theta - \hat{\theta}\right) \leq \chi^2_{0.90,\, K\, df}$$

Single test: 2 endpoints



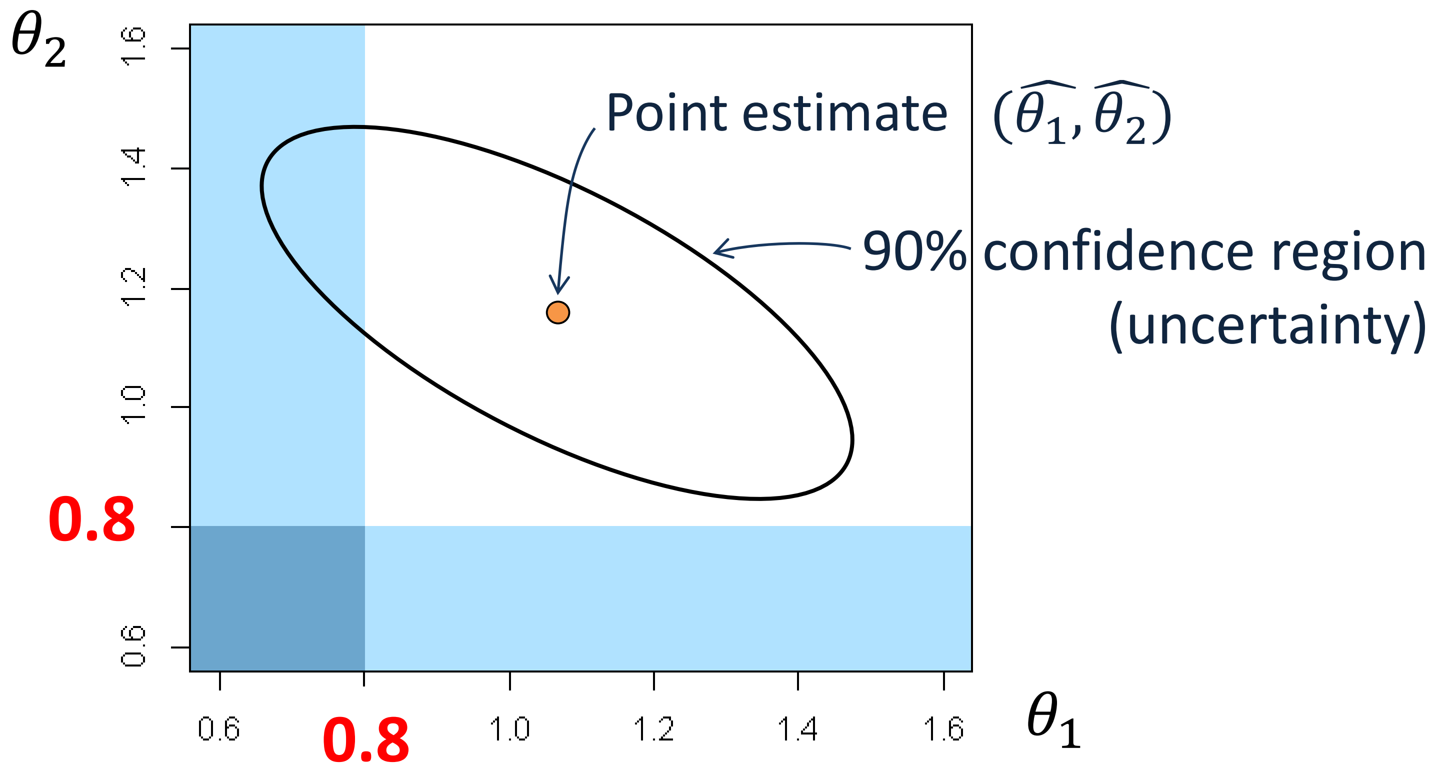
Inferiority on θ_1 ($H_{0,1}$)

Single test: 2 endpoints



Inferiority on θ_2 $(H_{0,2})$

Single test: 2 endpoints

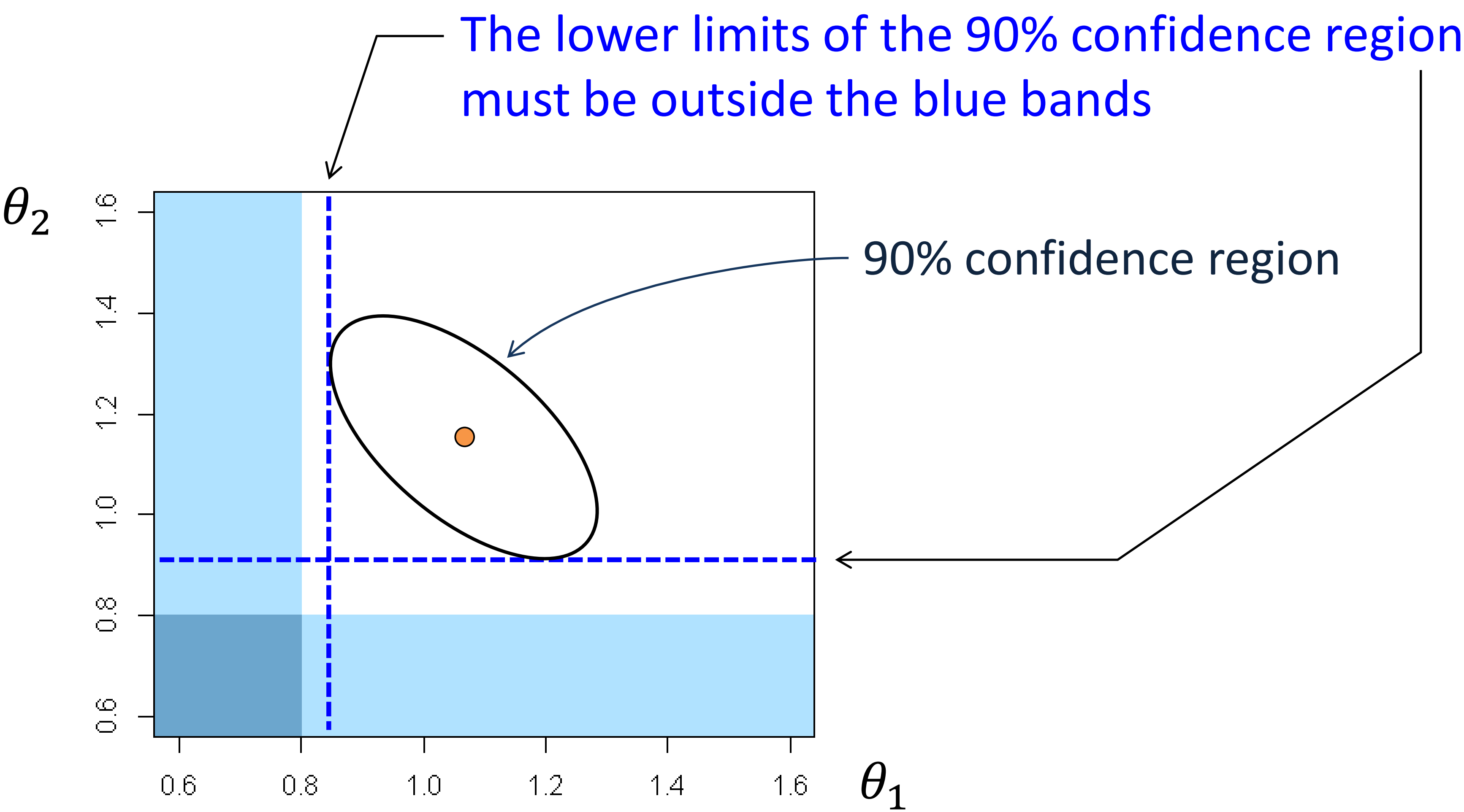


+ Inferiority on θ_1 **or** θ_2 ("**Union**" H_0)

Inferiority on θ_1 **and** θ_2 ("**Intersection**" H_0)

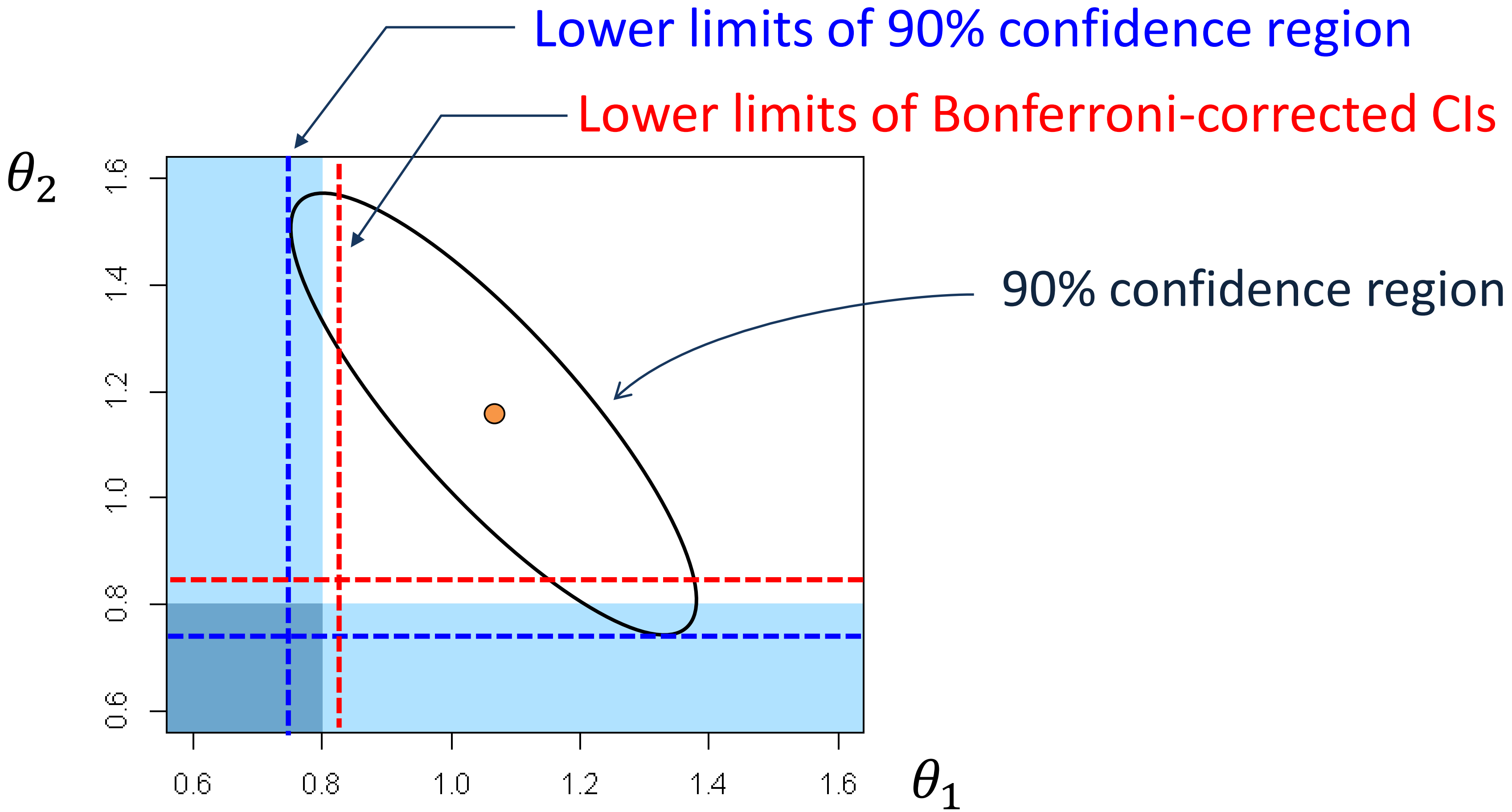
Comparison of the two strategies

- Case of "Union" non-inferiority analysis



"Union" non-inferiority

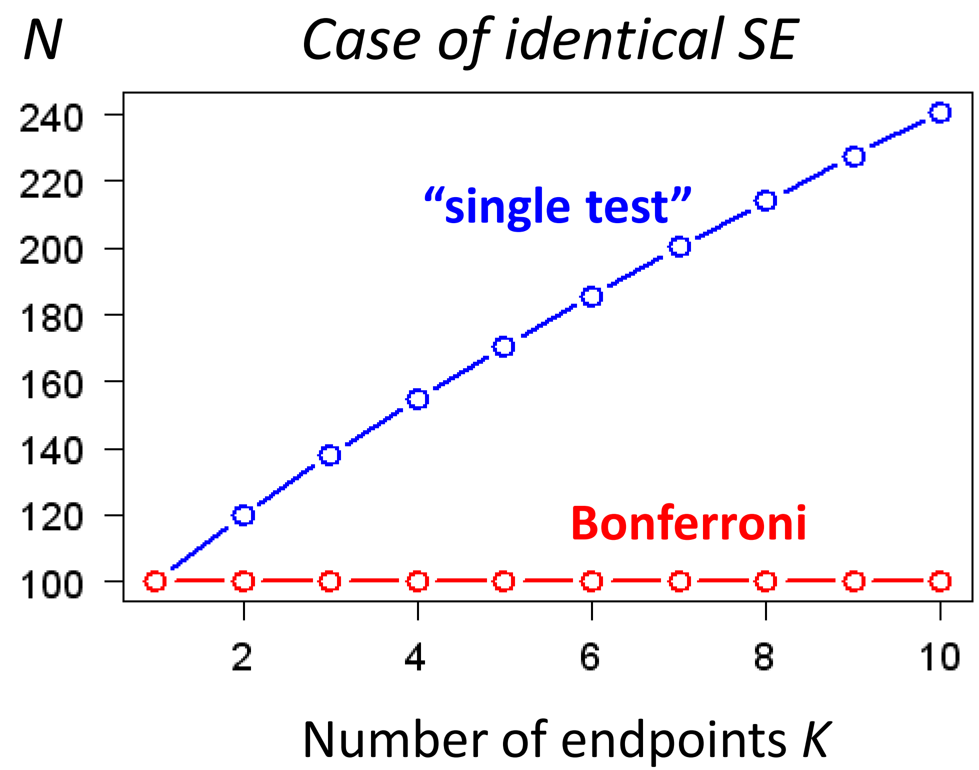
- … but they will always lie below the lower limits of Bonferroni-corrected CIs



"Union" non-inferiority

- So for "Union" non-inferiority, we gain nothing !

To achieve the same power, a single multivariate test requires to increase the number of subjects *N*



$$Ratio_N = \frac{\chi^2_{0.90,K\ df}}{\left[\Phi^{-1}\left(1-\frac{0.05}{K}\right)\right]^2}$$

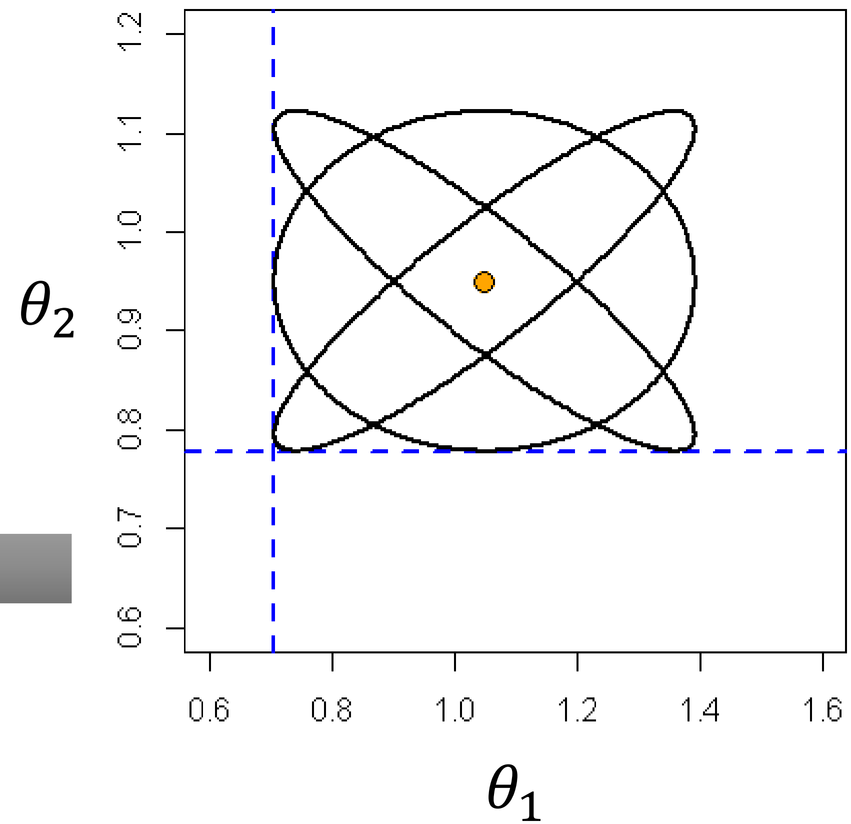
$\Phi = \text{cdf } N(0,1)$

"Union" non-inferiority: influence of correlations

- Correlation has no impact
- The lower limits of the 90% confidence region correspond to T^2 intervals (Hotelling's T^2)

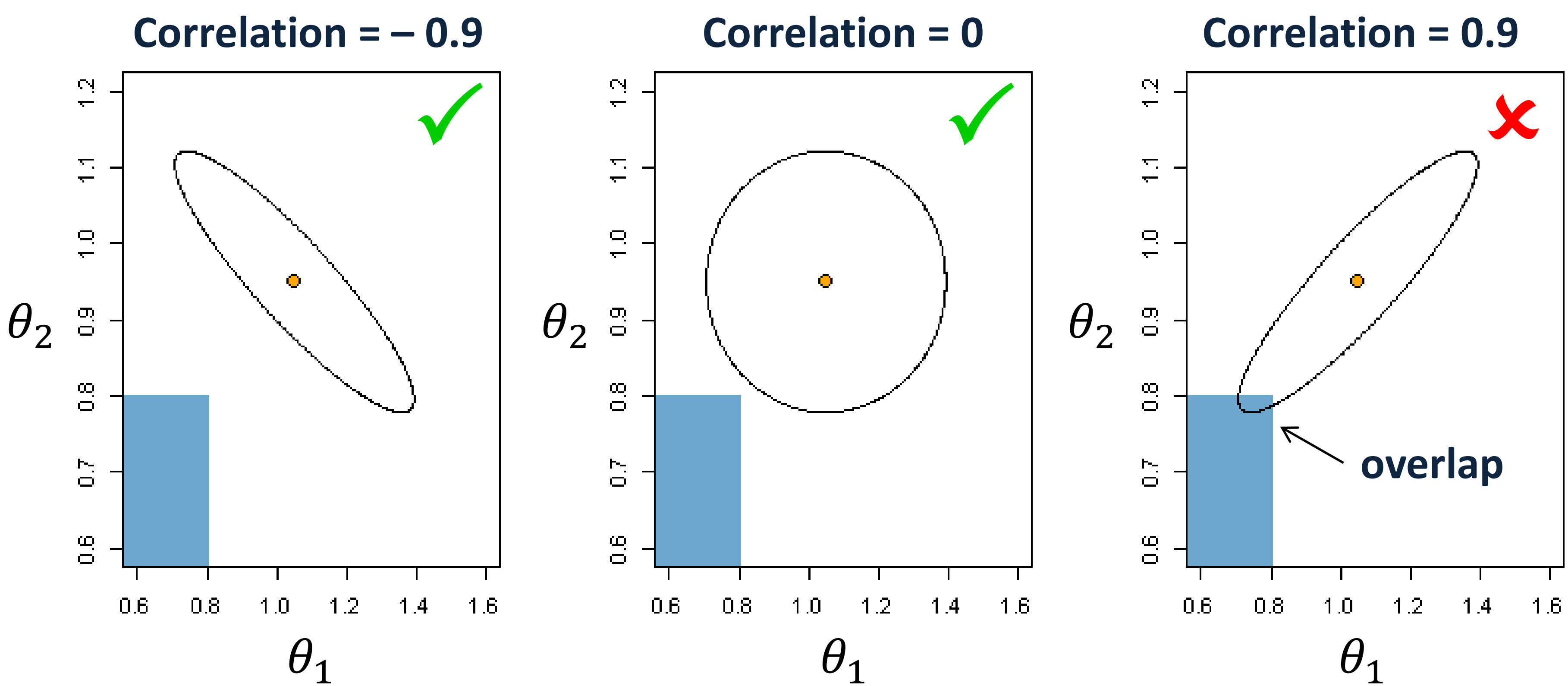
Bivariate case

Correlation changes but the lower limits (blue) are the same



What about "Intersection" non-inferiority ?

- **No definitive answer** depending on the values of SE, on the correlations and on the number of endpoints
- **Influence of correlations**



"Intersection" non-inferiority

- **Influence of the number of endpoints *K***
  - The structure of the variance-covariance matrix of $\hat{\theta}$ is very important as this gives the shape of the confidence region (ellipsoid)
  - We chose to illustrate this aspect using the following settings

$$\hat{\theta} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$Var(\hat{\theta}) = SE^2 \times \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{pmatrix}$$

No difference between treatments

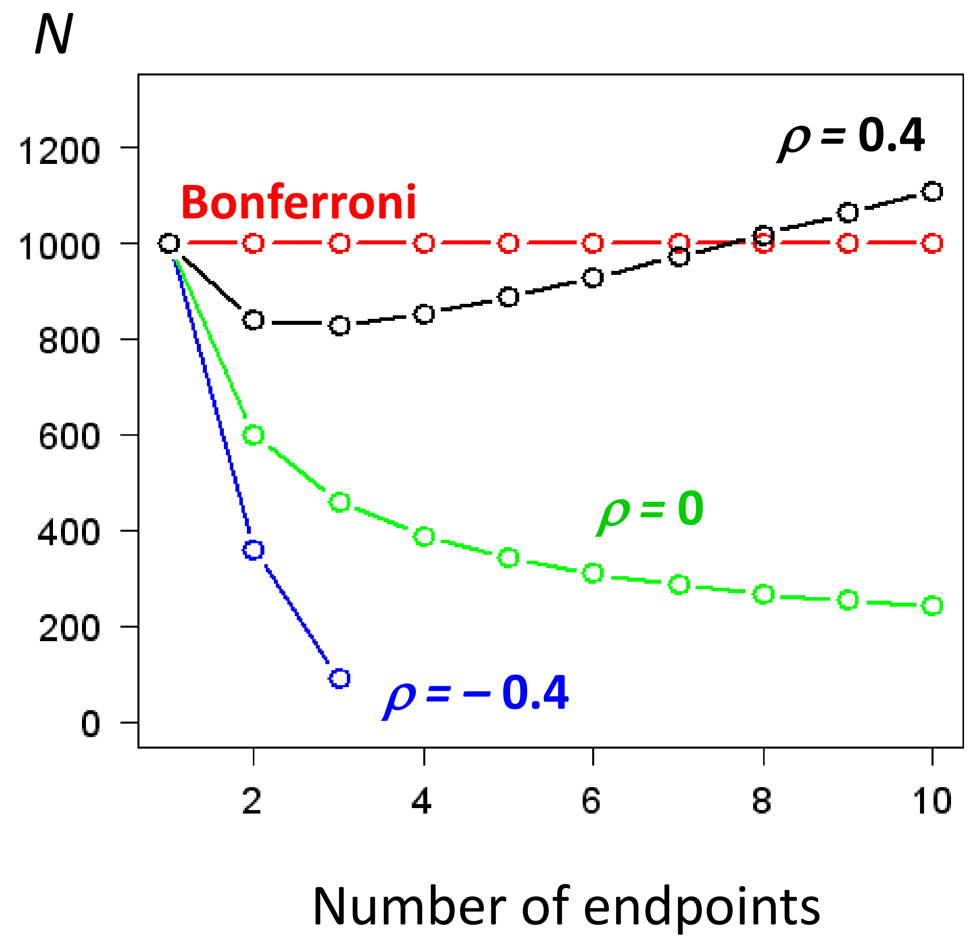
Identical standard errors

“Intersection” non-inferiority

- **Influence of the number of endpoints *K***

In each case, we compute the number of subjects *N* to achieve non-inferiority

$$Var(\hat{\theta}) = SE^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$



Application to real clinical data

- Robenacoxib to treat chronic osteoarthritis in dogs
- 3 randomized blinded clinical trials with positive control
- 4 endpoints: **ordinal scales** coded as 0 (normal)-1-2-3 (severe)

6220 observations

Total of 294 dogs

Robenacoxib: *N* = 232
Control: *N* = 62

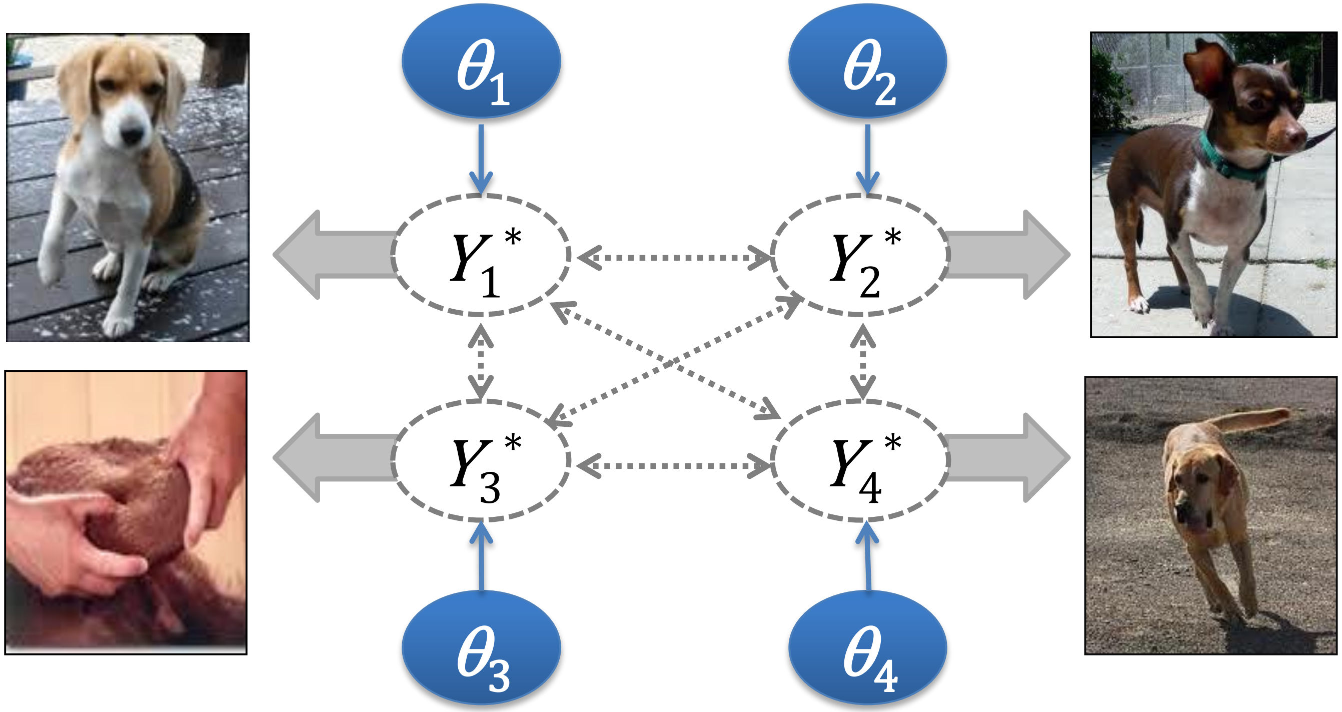
"Posture at a stand"

"Lameness at walk"

"Pain at palpation"

"Lameness at trot"

Application: joint mixed effects model



- Each scale k = categorization of a latent continuous variable Y_k^*
- All correlations between the latent variables were assessed
  Laffont al. PAGE 21 (2012) Abstr 2548
- $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$: ratios for robenacoxib efficacy vs. control

Application: methods and results

- "Intersection" non-inferiority concluded when no overlap between H_0 region and multivariate 90% confidence region
- Evaluation by Monte Carlo simulations ($K = 4$)
- $\mathrm{Var}(\hat{\theta})$ obtained from the joint model analysis

"Intersection" non-inferiority was demonstrated with the single multivariate test, not with multiple univariate tests

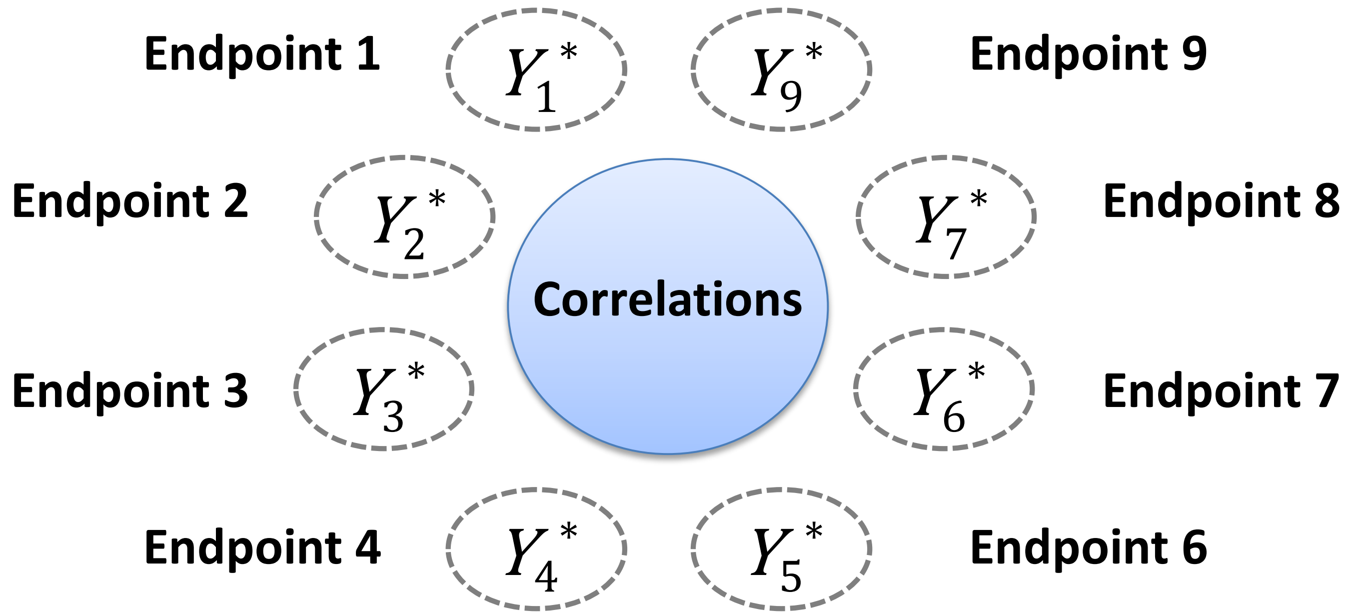
Lower bounds of Bonferroni-corrected CI ranged between 0.76 and 0.78, all < 0.8

Single multivariate test: conclusion

- It is usually claimed that a single multivariate test is more powerful to show a significant difference ($\theta \neq \theta_0$)

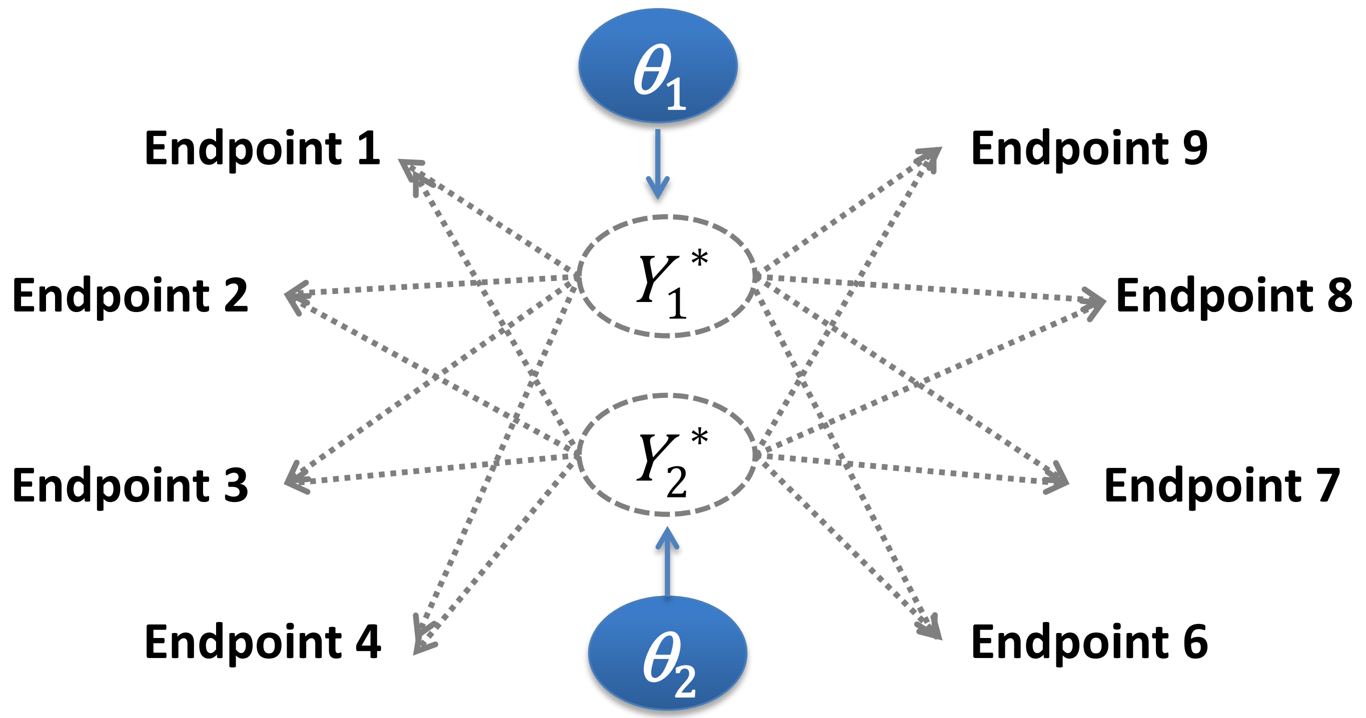
- **For non-inferiority, things are a bit more complicated**
  - For "**Union**" test, we systematically loose power compared to simple Bonferroni-corrected CIs
  - For "**Intersection**" test, no definitive answer, but what is the relevance of "intersection" non-inferiority?
  - An increase in dimension appears to be a problem!
  - Fortunately, there are modeling techniques that can help in reducing dimension without loosing information (compared to "summary" variables), but this is another story…

Surprising question… Thank you!



General case

Factor analysis



In this example, all the information is summarized by only 2 latent variables