How to analyse multiple ordinal scores in a clinical trial? Multivariate vs. univariate analysis

Céline M. Laffont, Didier Concordet

How to analyse multiple ordinal scores in a clinical trial?

Multivariate vs univariate analysis

C. Laffont & D. Concordet

INRA, UMR 1331, Toxalim, F-31027 Toulouse, France.
Université de Toulouse, INPT, ENVT, UPS, EIP, F-31076 Toulouse, France.

Project funded by **Novartis Pharma AG, Switzerland**

Case study

- In dogs, the effect of robenacoxib (NSAID) on osteoarthritis is assessed using several scores

Posture (at stand)	**Lameness at walk**	**Lameness at trot**	**Pain at palpation**	**Willingness to raise contralateral limb**
Normal	None	None	No pain	No resistance
Slightly abnormal	Mild	Mild	Mild	Mild resistance
Markedly abnormal	Obvious	Obvious	Moderate	Moderate resistance
Severely abnormal	Marked	Marked	Severe	Strong resistance
				Refuses

What is done in practice

- Compute the **sum of scores** and analyse it as a **continuous** variable

Posture

Lameness at walk

Lameness at trot

Willingness to raise contralateral limb

Pain at palpation

Sum of investigator scores: 0-16

Why is this approach not appropriate?

- It ignores the actual metric of each score and assumes that all categories are equidistant

Posture	
Normal	0
Slightly abnormal	1
Markedly abnormal	2
Severely abnormal	3

➢ The distance between **0** and **1** is not the same as the distance between **2** and **3**

Why is this approach not appropriate?

- It ignores the actual metric of each score and assumes that all categories are equidistant

Posture	
Normal	0
Slightly abnormal	1
Markedly abnormal	2
Severely abnormal	3

Lameness at trot	
None	0
Mild	1
Obvious	2
Marked	3

- The distance between **1** an **2** is not the same as the distance between **1** and **2**
- **"Weighted"** sum of scores have been proposed but not ideal

What should be done

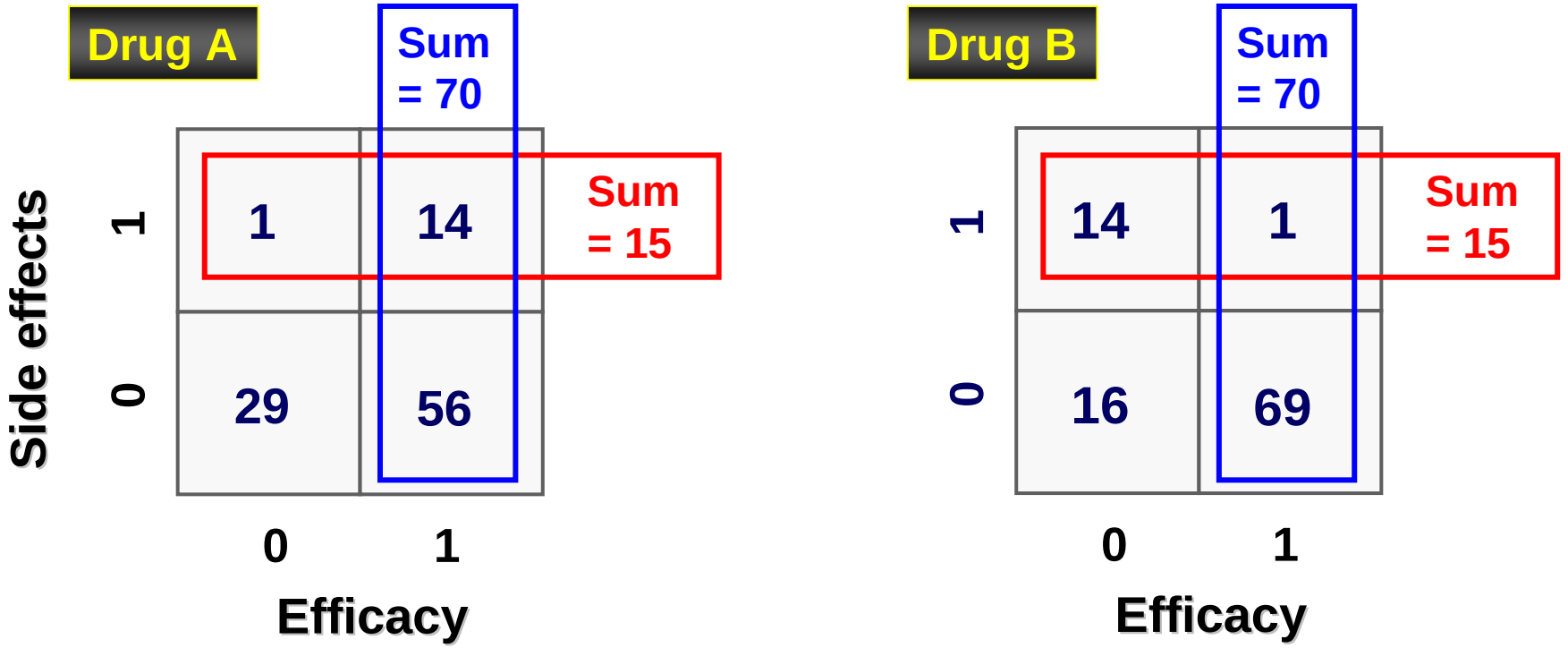
- Analyse the data as ordered categorical data using appropriate models (logit, probit…)
- Many publications on ordinal data analysis
  - Applications to assess drug effect

    pain relief, nicotine craving scores, sedation, diarrhea, neutropenia…
  - Estimation/modelling issues

    AAPS 2004, JPKPD 2001, 2 articles in JPKPD 2004, JPKPD 2008…
- **But published models restricted to the analysis of only *one* score !**

Limits of univariate analyses

- They only estimate marginal distributions



Drugs A and B have the same marginal distributions but different benefit-risk ratios !

Limits of univariate analyses

- Univariate analyses assume scores are independent while in many cases, they should be correlated

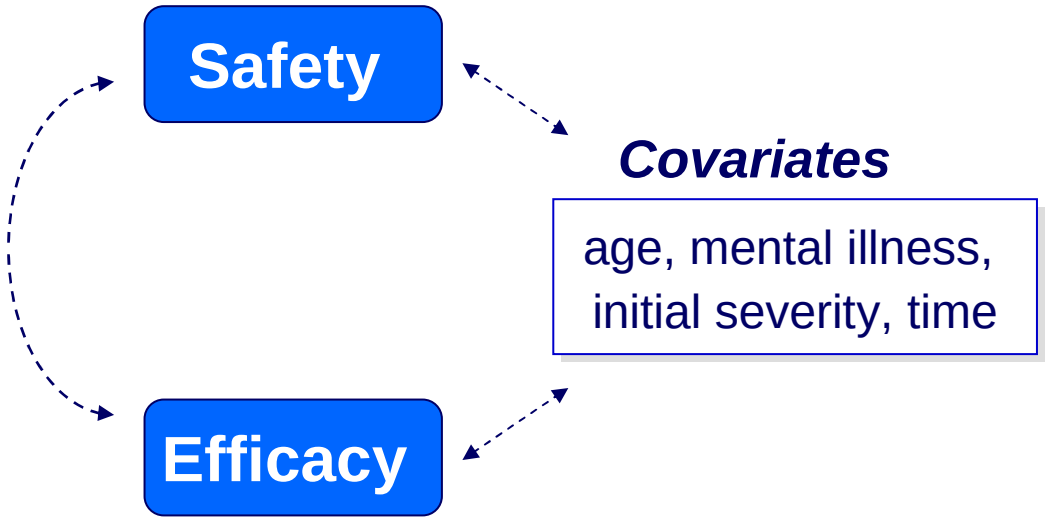
Posture (at stand)	**Lameness at walk**	**Lameness at trot**	**Pain at palpation**	**Willingness to raise contralateral limb**
Normal	None	None	No pain	No resistance
Slightly abnormal	Mild	Mild	Mild	Mild resistance
Markedly abnormal	Obvious	Obvious	Moderate	Moderate resistance
Severely abnormal	Marked	Marked	Severe	Strong resistance
				Refuses

Multivariate analysis: background

- Very few approaches exist to analyse jointly several ordinal scores
- In 2007, Todem et al. proposed a **probit mixed effects model** for longitudinal bivariate data

  ***Statist. Med.* 26:1034**

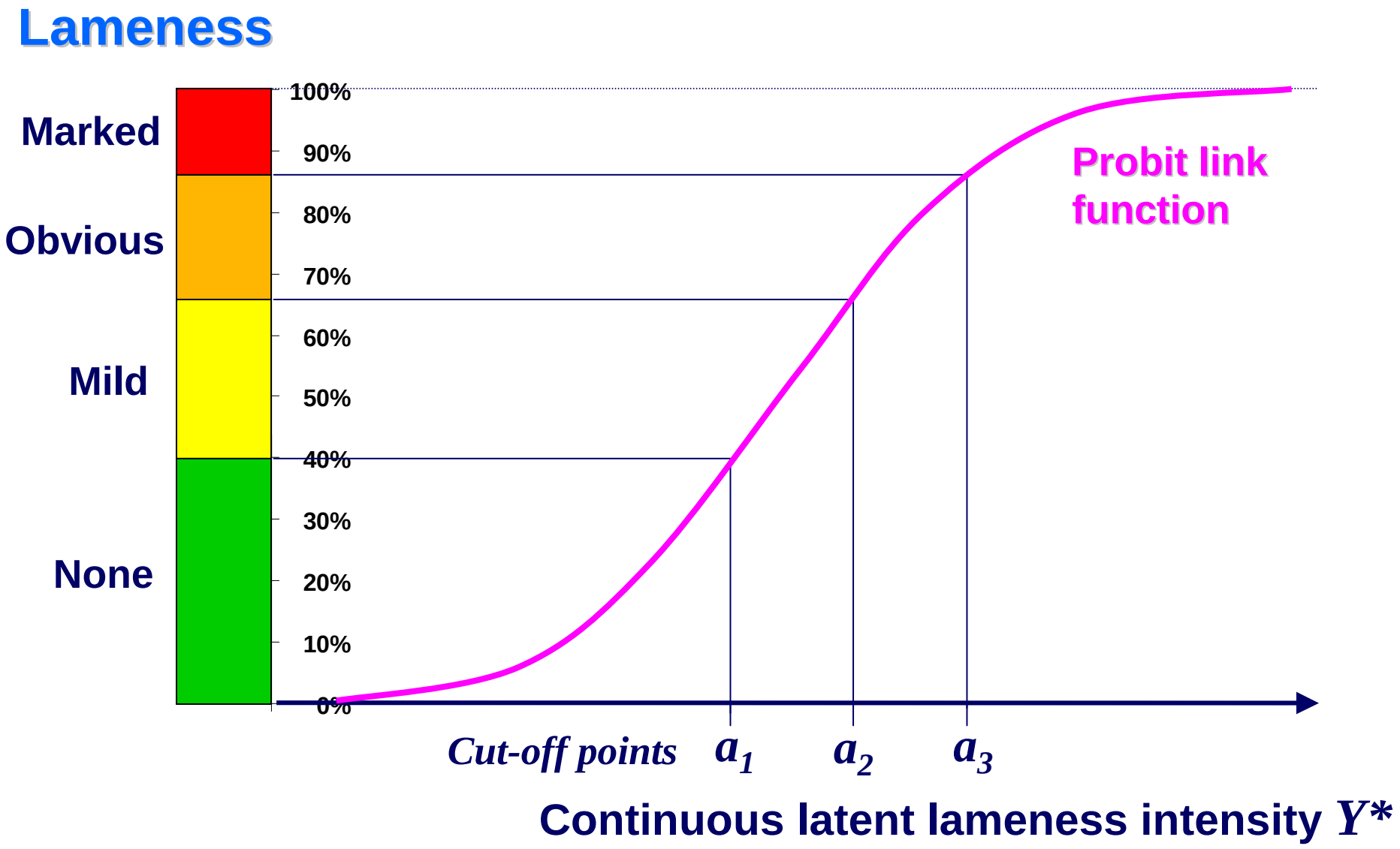
Application:
Fluvoxamine data



Objectives

- **Extend this previous model (Todem et al.)**
  - To analyse more than two scores (model estimation issue)
  - To apply to population PK/PD data
- **Identify similarities between scores**
  - Are some scores redundant?

Model based on latent variable approach



Model based on latent variable approach

- The K scores $Y_1, Y_2 \ldots Y_K$ are obtained by categorisation of K latent variables $Y_1^*, Y_2^* \ldots Y_K^*$

$$Y_{kij}^* = f_k(\beta_k, x_{kij}) + \eta_{ki} + \varepsilon_{kij}$$

f_k : (non)linear function for response $k = 1 \ldots K$

x_{kij} : covariates for subject i, response k and time t_{kij}

β_k : fixed effects for response k

η_{ki} : random effects for inter-individual variability

ε_{kij} : random effects for intra-individual variability

Modelling correlations between scores

- The correlations between the scores across time are modeled as correlations between latent variables Y^*

$$\begin{pmatrix} \eta_{1i} \\ \dots \\ \eta_{Ki} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}, \Omega \right) \qquad \begin{pmatrix} \varepsilon_{1ij} \\ \dots \\ \varepsilon_{Kij} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}, \Gamma \right) \qquad \eta_{ki} \perp \varepsilon_{kij}$$

$$\varepsilon_{kij} \perp \varepsilon_{kij'}$$

- η : overall correlation between scores within subjects
- ε : correlation within subjects at a given occasion

Parameter estimation

- **Likelihood function**

$$L(y_i/\eta_i)=\prod_{j=1}^{ni}\prod_{m_1=1}^{c1}\cdots\prod_{m_K=1}^{cK}\left[P\left(Y_{1ij}=m_1;\ \ldots\ ;Y_{Kij}=m_K\middle|\eta_i\right)\right]^{I(y_{1ij}=m_1)\times\ldots\times I(y_{Kij}=m_K)}$$

$$L(y_i)=\int L(y_i|\eta_i)P(\eta_i)d\eta_i$$

$$L(y,\theta)=\prod_{i=1}^{N}L(y_i)$$

where $\eta_i=(\eta_{1i},\ldots,\eta_{1K})$

Need to compute the multivariate normal cdf Φ_K

No closed form (nonlinear mixed effects model)

Parameter estimation

- **No current software can be used**
  - Own program written in C++
- **Approximation of the multivariate normal cdf Φ_K**
  - Gauss-Legendre quadratures (8 nodes)
- **Stochastic EM algorithm (SAEM-like)**
  - Efficient for ordinal data analysis

    (Kuhn & Lavielle, CSDA 2005 ; Savic et al. AAPS 2011)
  - Metropolis-Hastings algorithm
  - Gauss-Newton/gradient method for optimisation

Method evaluation with simulation studies

- **Bivariate analysis**

2 scores Y_1 and Y_2 (3 and 4 categories resp.)

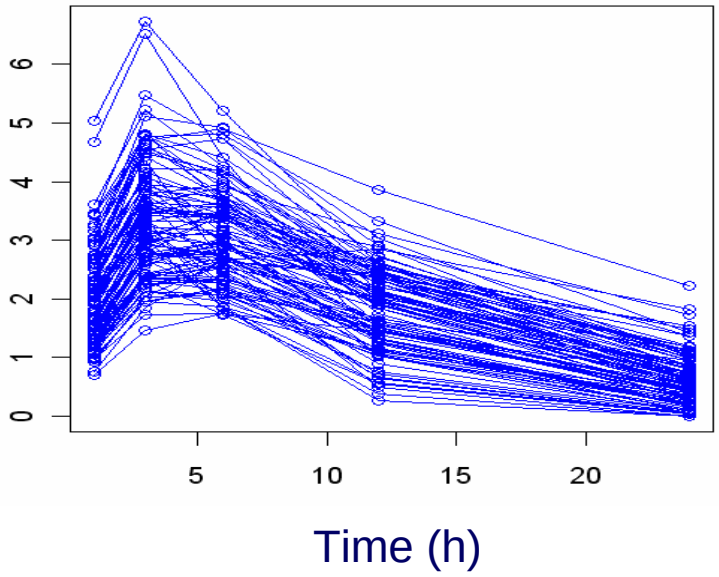
$$Y_{1ij}^{*} = Slope \times C_{ij} + \eta_{1i} + \varepsilon_{1ij}$$

$$Y_{2ij}^{*} = \frac{E_{\max} \times C_{ij}}{EC_{50} + C_{ij}} + \eta_{2i} + \varepsilon_{2ij}$$

$$\text{corr}\,(\eta_{1i}, \eta_{2i}) = 0.8$$

$$\text{corr}\,(\varepsilon_{1ij}, \varepsilon_{2ij}) = 0.8$$

Drug conc. C_{ij} (1 comp. oral)



100 subjects, 5 obs. /subject at 1, 3, 6, 12 and 24 h

Method evaluation with simulation studies

- **Trivariate analysis**

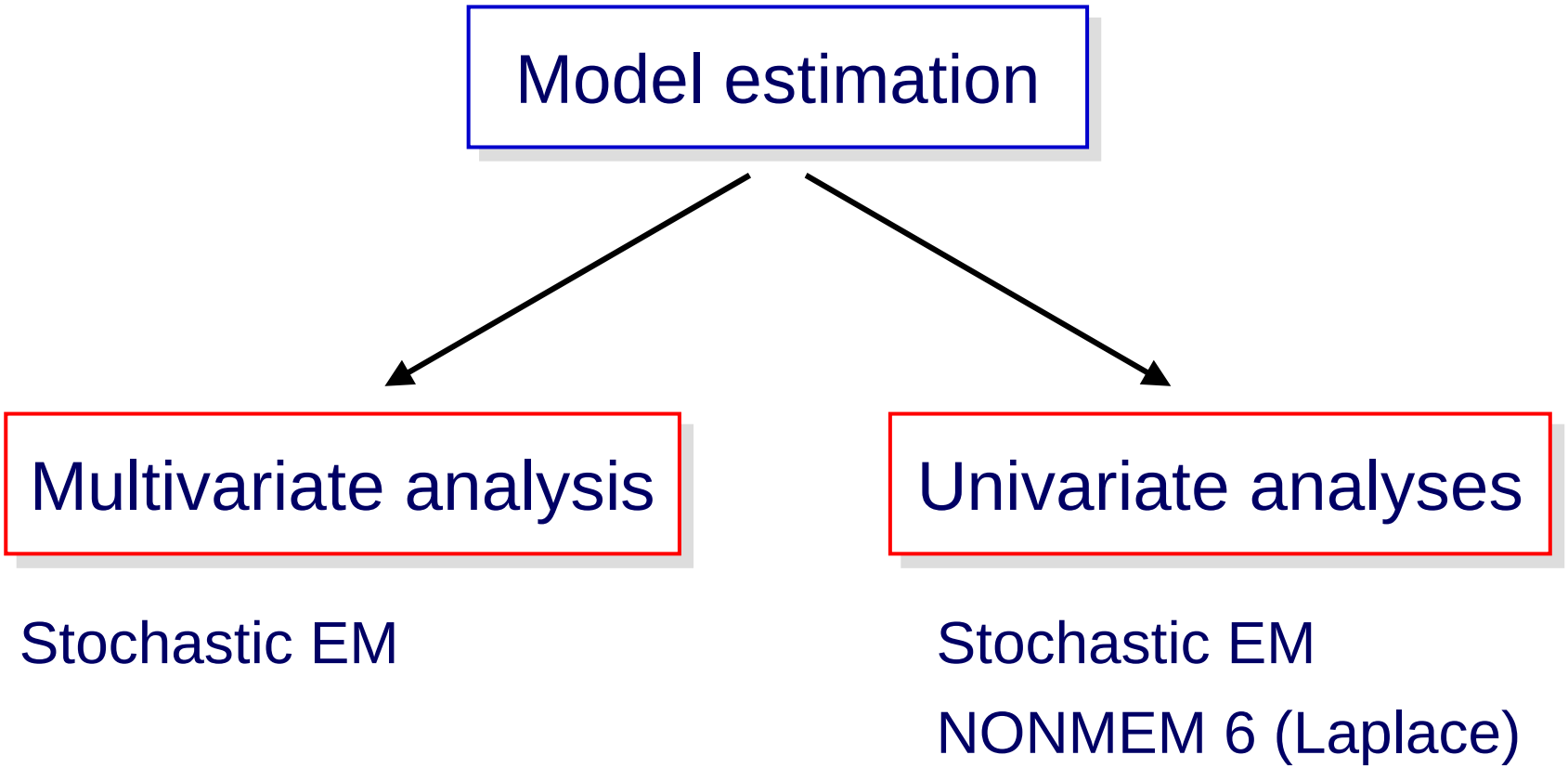
3 scores Y_1, Y_2 and Y_3 (3 categories each)

$$Y_{kij}^{*} = Slope_k \times Dose_j + \eta_{ki} + \varepsilon_{kij} \qquad k = 1...3 \ , \ j = 1...4$$

$$Corr(\eta) = \begin{pmatrix} 1 & & \\ 0.90 & 1 & \\ 0 & -0.10 & 1 \end{pmatrix} \qquad Corr(\varepsilon) = \begin{pmatrix} 1 & & \\ 0.85 & 1 & \\ 0 & -0.10 & 1 \end{pmatrix}$$

200 subjects, 4 obs./subject, i.e. one per dose: 2.5, 5, 10, 20 mg

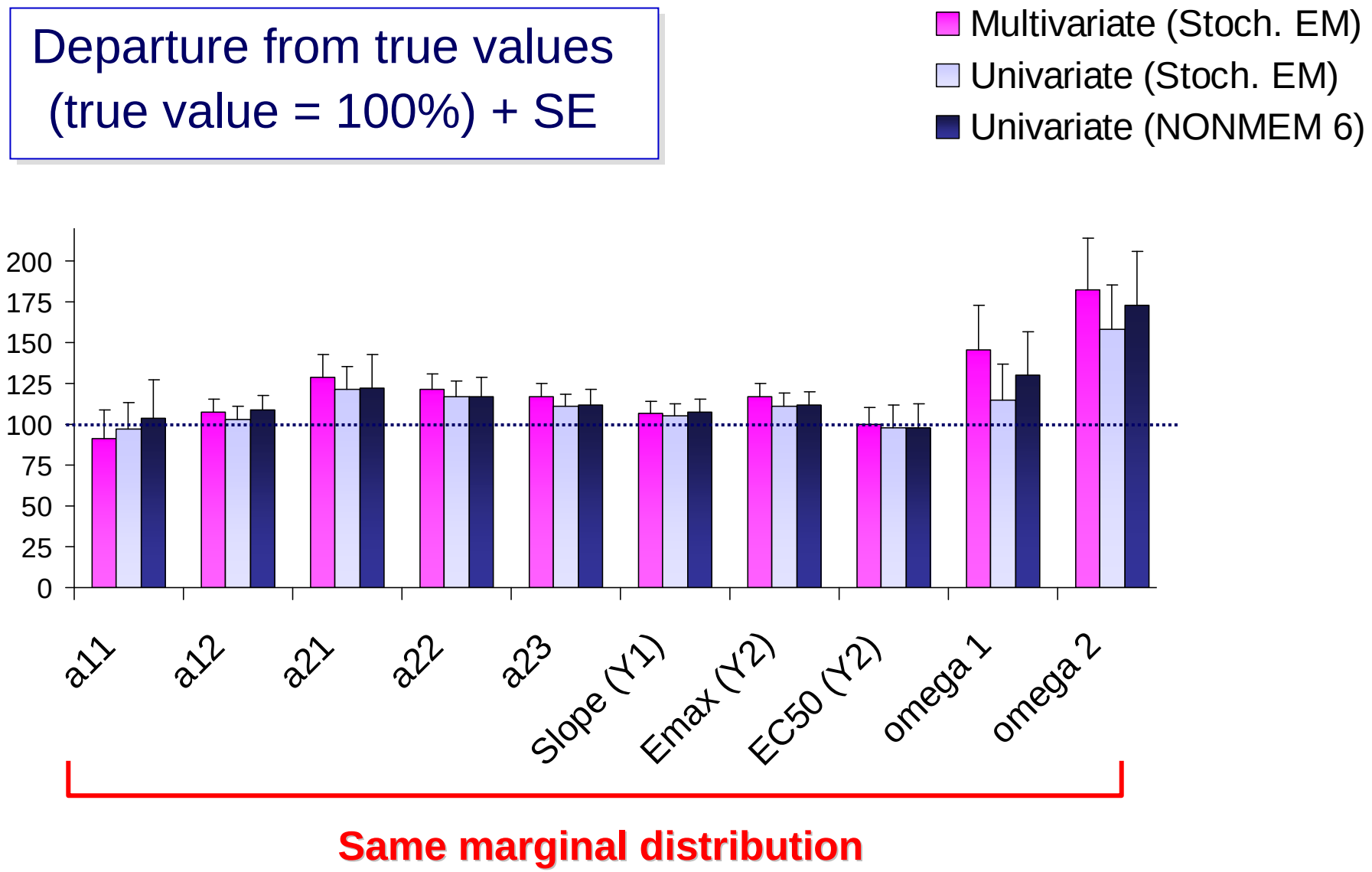
Method evaluation with simulation studies



Results:
Bivariate analysis

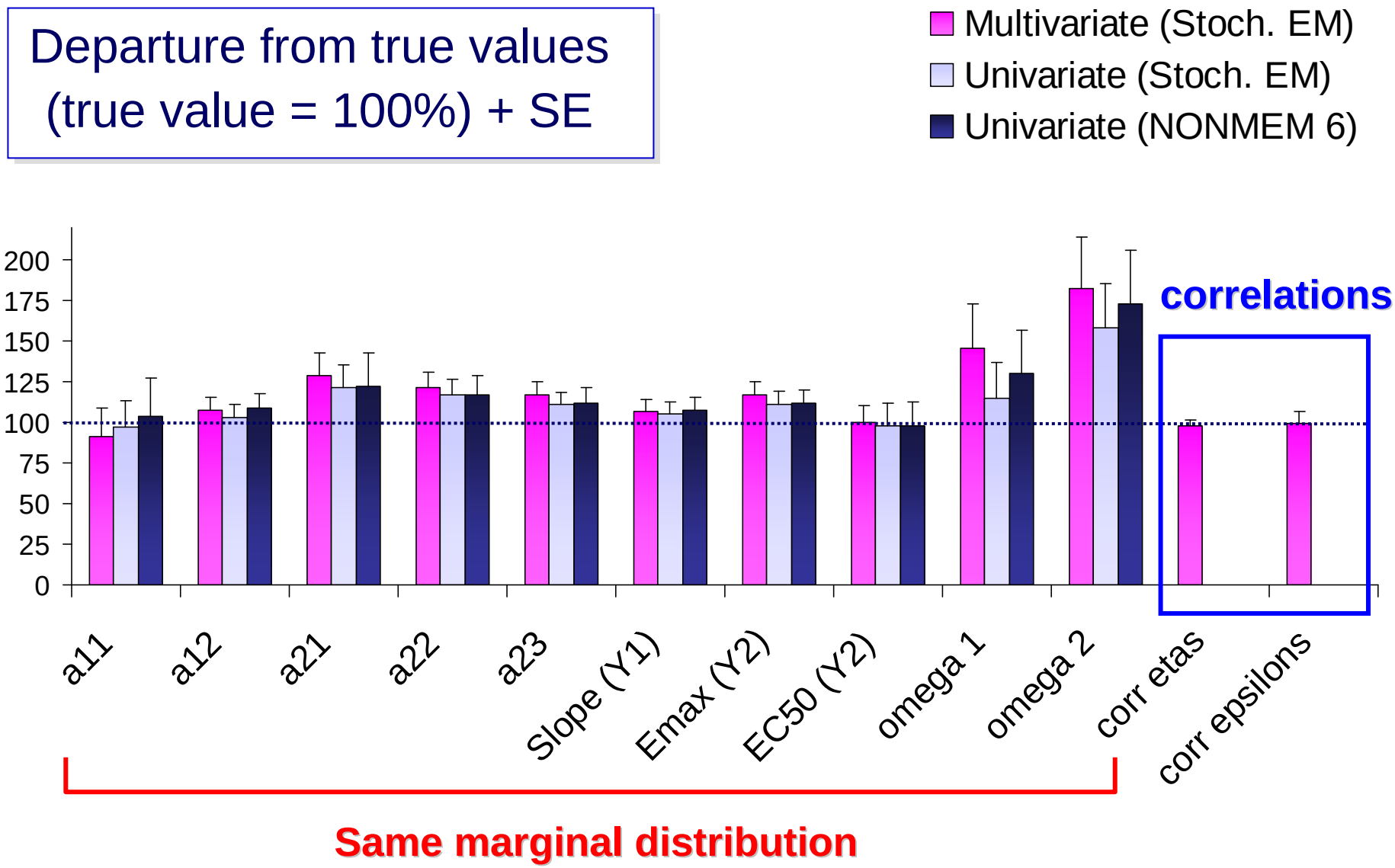
Bivariate analysis : parameter estimates

Departure from true values
(true value = 100%) + SE



Bivariate analysis : parameter estimates

Departure from true values
(true value = 100%) + SE



Bivariate analysis : VPC

- **Joint probability**

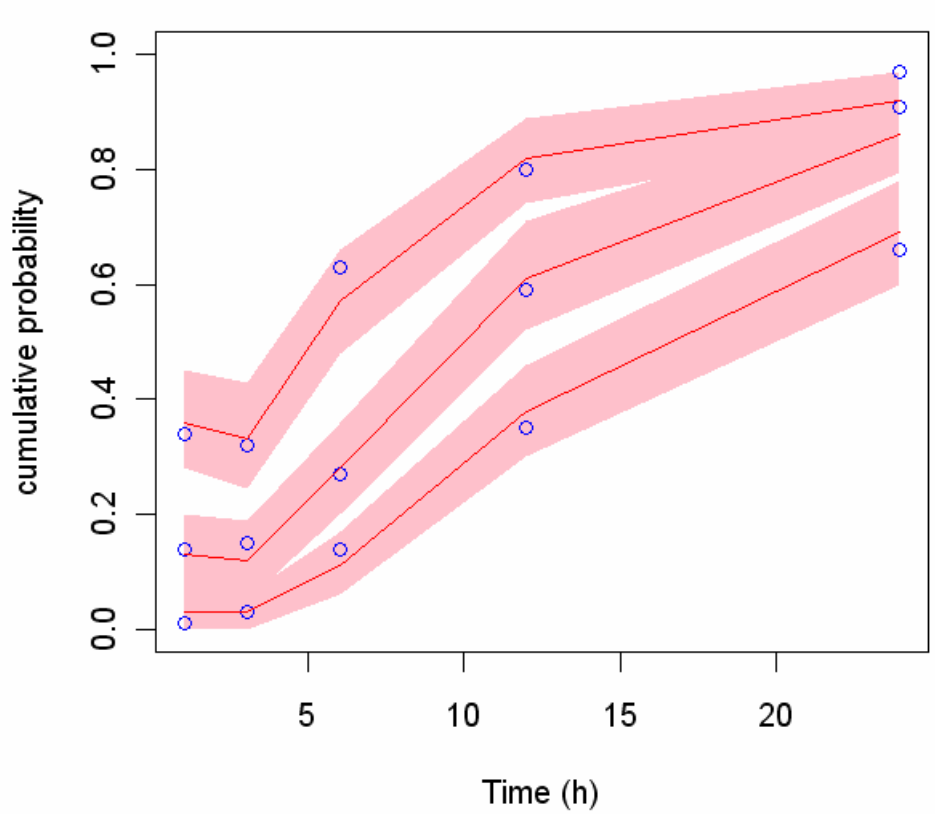
$P(Y_1 \leq 2\ ;\ Y_2 \leq 1,\ 2 \text{ or } 3)$

observations

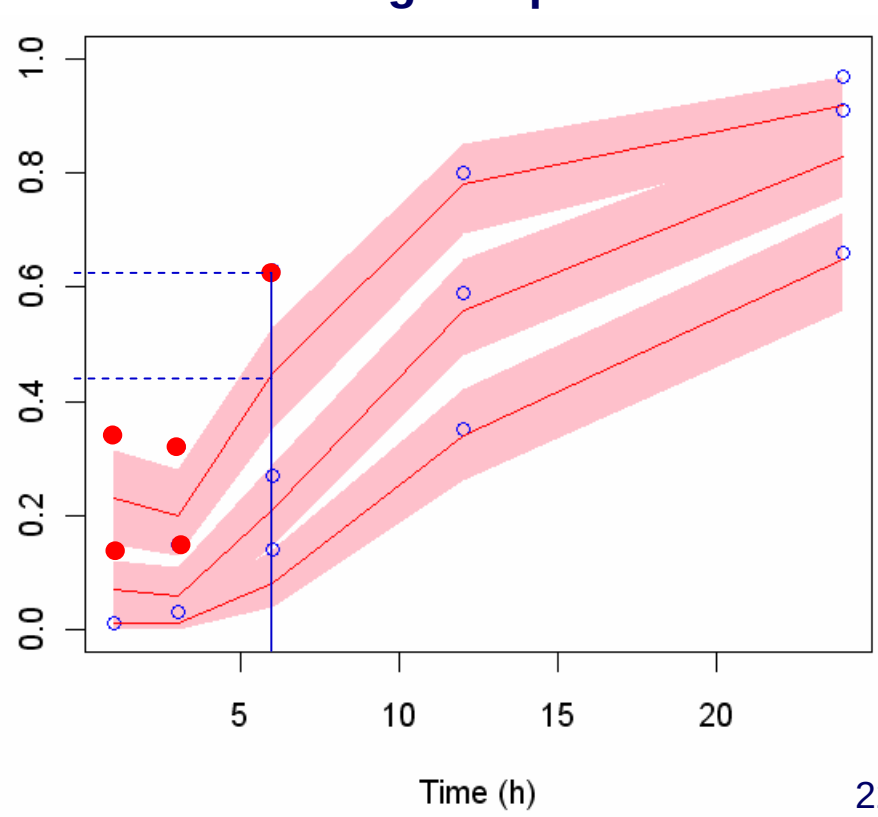
95% CI for model predictions

median

Bivariate analysis



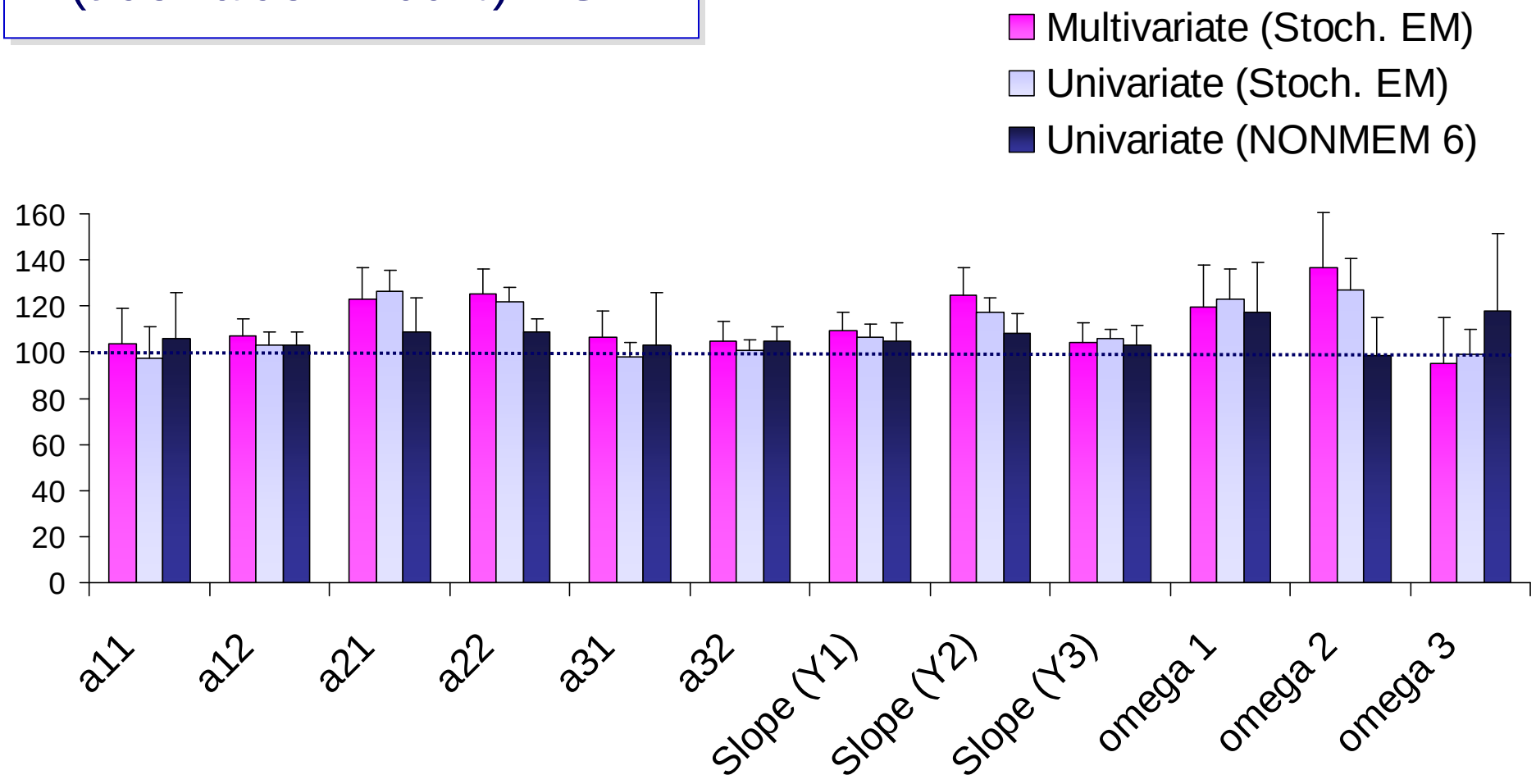
Univariate analyses assuming independence



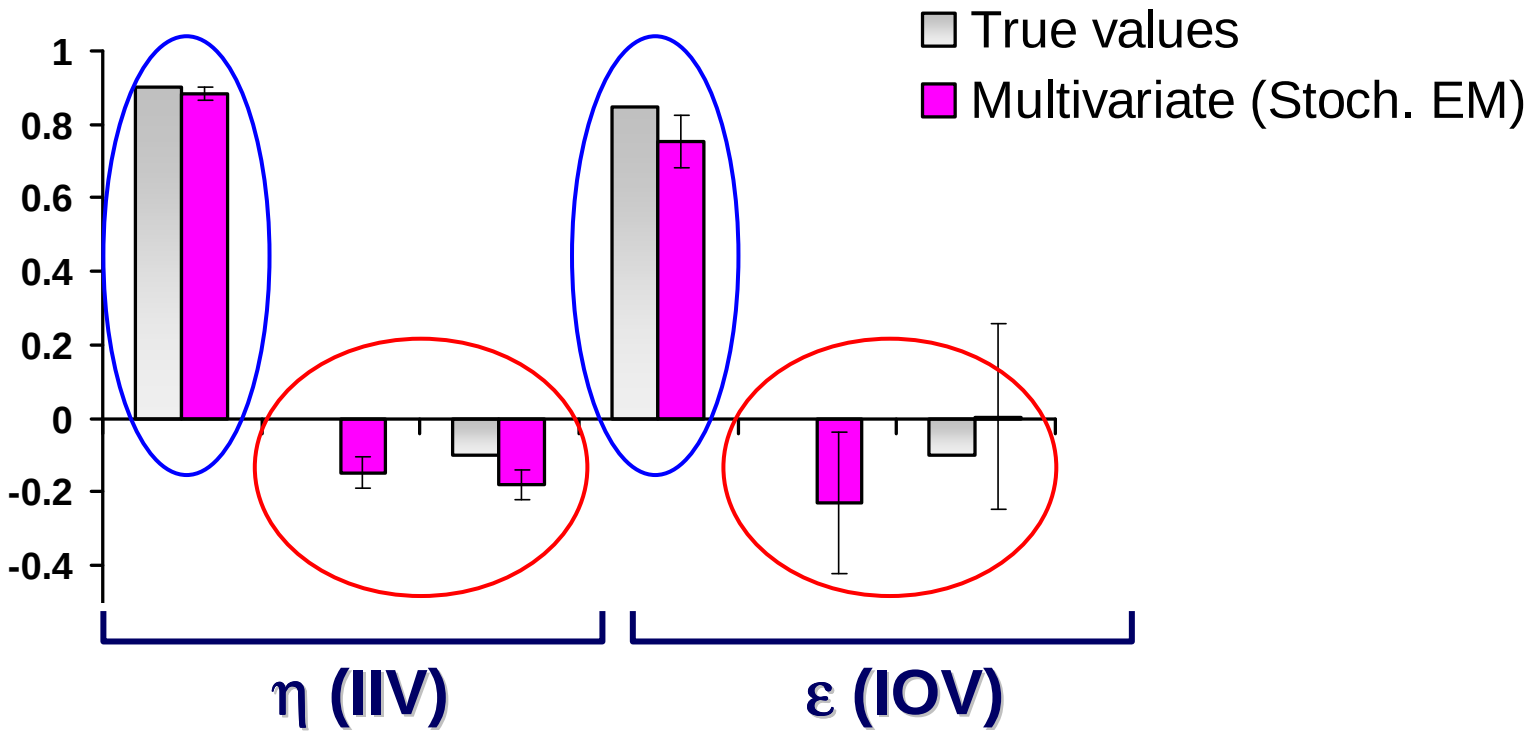
Results:
Trivariate analysis

Estimation of marginal distribution

Departure from true values
(true value = 100%) + SE



Estimation of correlations



- The multivariate analysis allows to catch:
  - the high correlations between scores 1 and 2
  - the poor correlation of score 3 with the others

Trivariate analysis: VPC

- **Joint distribution**

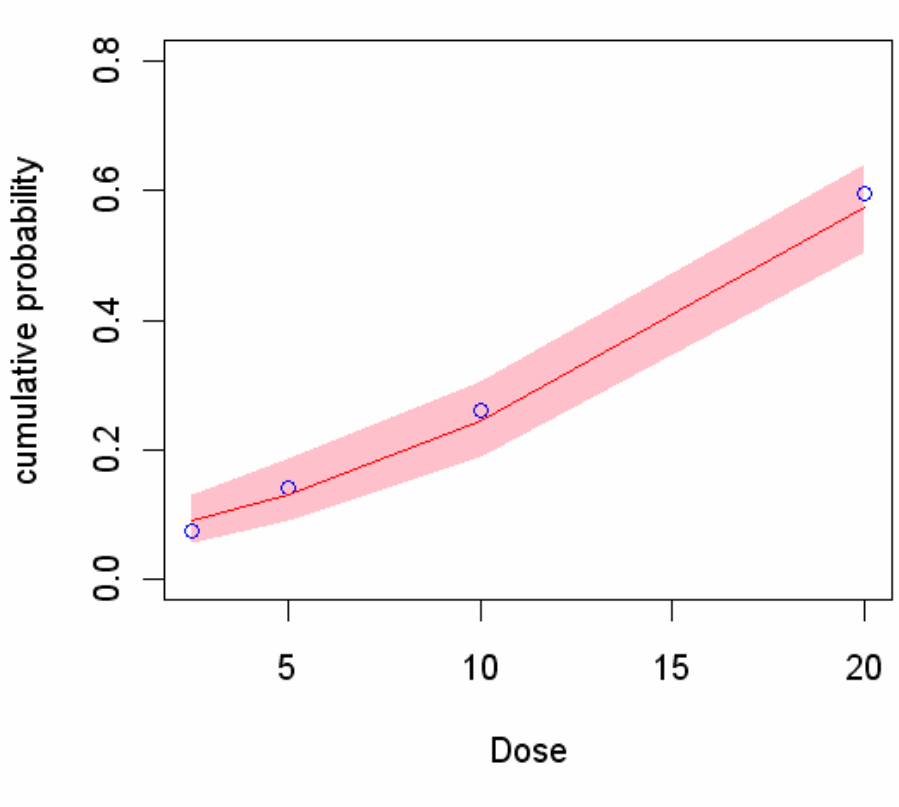
$$P(Y_1 = 3 \; ; \; Y_2 = 3)$$

observations

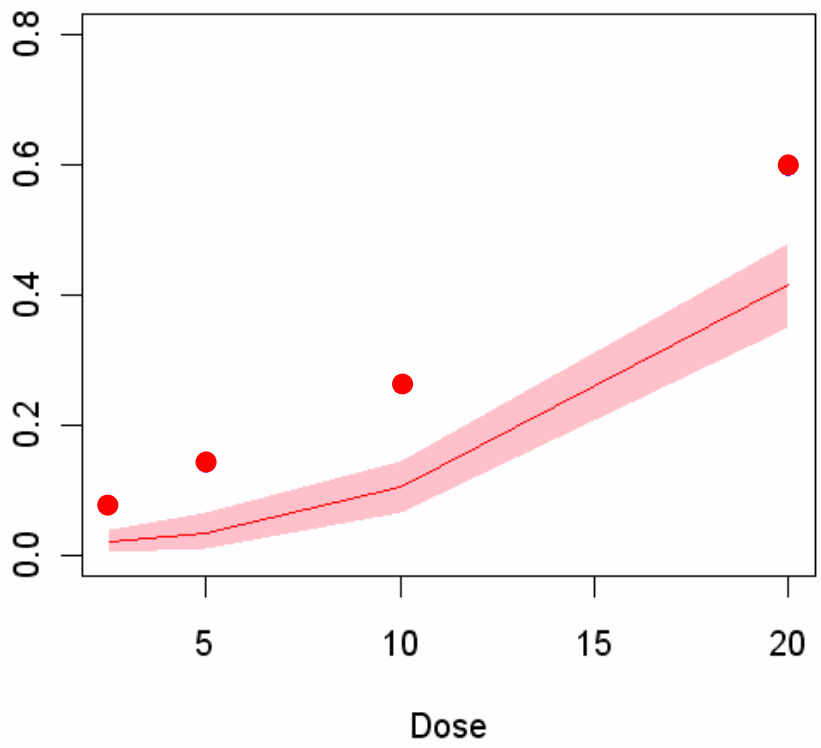
95% CI for model predictions

median

Trivariate analysis



Univariate analyses assuming independence



Trivariate analysis: VPC

- **Joint distribution**

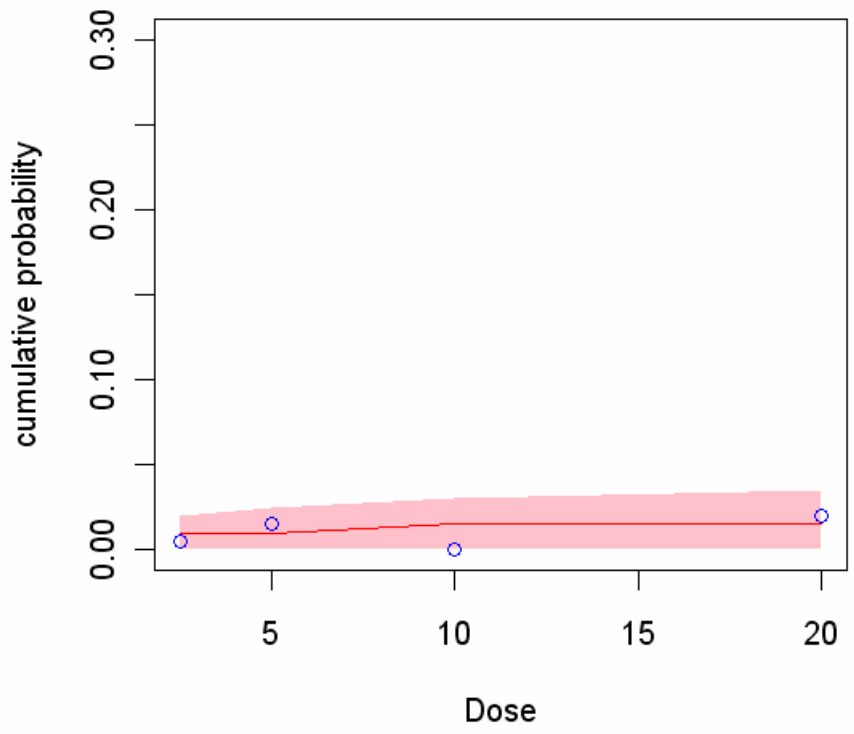
$$P(Y_1 = 3 \; ; \; Y_2 = 1)$$

observations

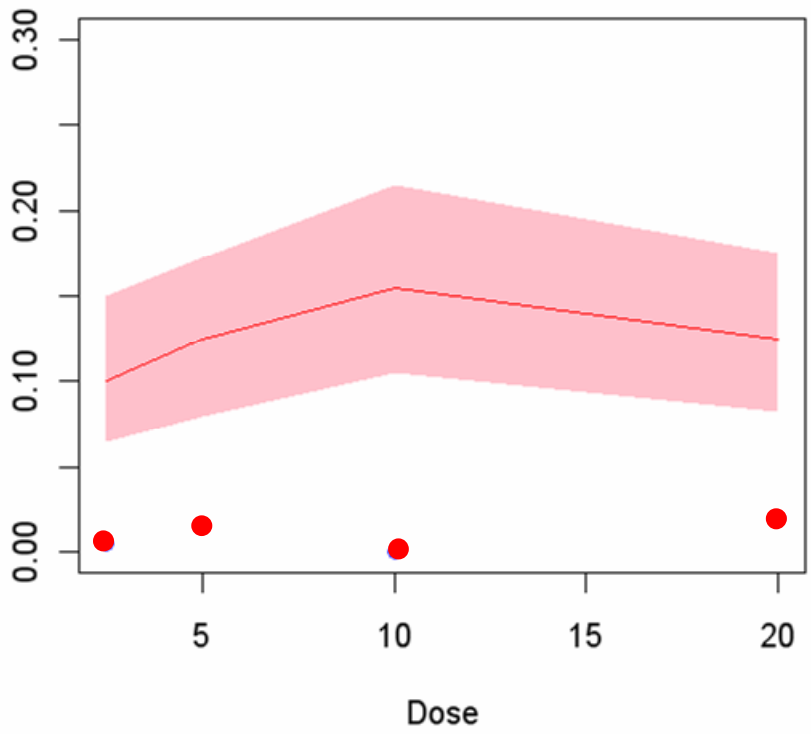
95% CI for model predictions

median

Trivariate analysis



Univariate analyses assuming independence



Objectives

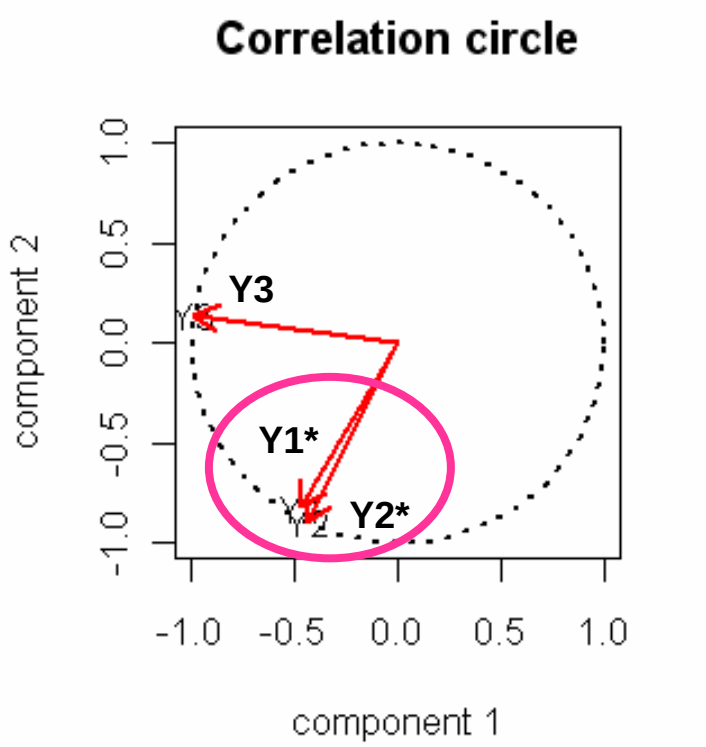
- **Generalise this previous model (Todem et al.)**
  - To apply to population PK/PD data
  - To analyse more than two scores in practice (model estimation issue)

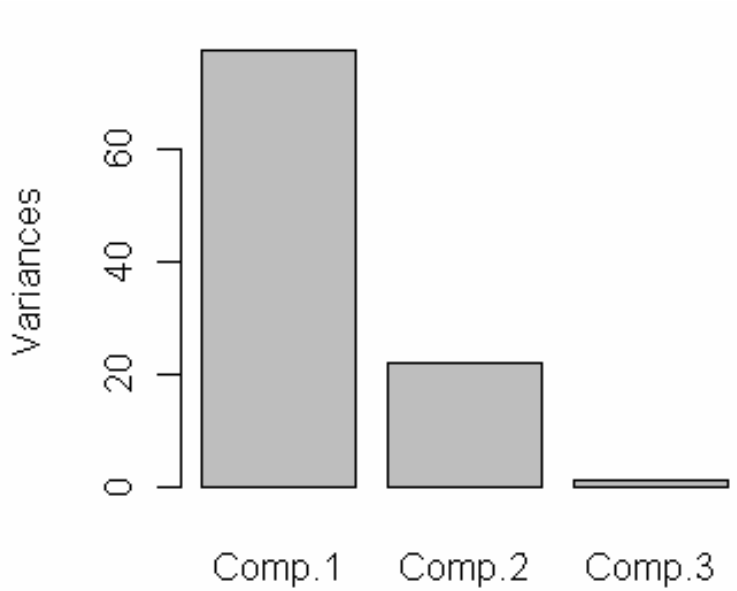
- **Identify similarities between scores**
  - Are some scores redundant?

Principal Component Analysis (PCA)

- To identify scores that document a same physio-pathological process and possible redundancies

  Ex: trivariate analysis





Conclusion

Univariate analyses

Pros
- Rapid - Easy to understand and interpret
Cons
- Assess marginal distributions only - Can lead to some bias and wrong conclusions

Multivariate analysis

Pros
- Avoid bias and wrong conclusions in clinical trials - Identification of redundancies between scores (PCA)
Cons
- Computation time (bivar. = 3h; trivar. = 18h) - Homemade program