Multivariate Analysis of Laboratory Data within Linear Model Framework

Ramona Scheufele Reinhard Meister

Charité, TFH Berlin

supported by

Hypatia Programm

Förderung von Nachwuchswissenschaftlerinnen an der TFH Berlin

Laboratory Data Impact

Laboratory parameters

- classical indicators of physiological status
- well understood within context of clinical chemistry
- important for analysing pharmacological effects in clinical trials

- not often used as primary endpoints
- only reported in a descriptive way
- almost never analysed as multivariate samples

CHARITÉ

Laboratory Data Importance

Laboratory parameters

- modern genome-based approaches don't replace the macroscopic picture of biochemical pathways
- information on biochemical pathways has been accumulated for many many decades
- deviations from expected patterns in the joint distribution of laboratory parameters can give important hints on possible adverse drug effects

CHARITÉ

Example Multicenter RCT on Diabetes

Dataset

- Acknowledgement Dr. Ullrich Munzel kindly provided a *real* dataset
- Treatments $n \sim 40$
  3 Dose groups and 1 Placebo control
- Lab parameters $p = 22$
  Blood, Electrolite/ Liver / Kidney, Lipids
- Centers

Center	1	2	3	4	5
n	17	23	57	31	46

- Covariates
  Treatment, Center, Sex, Age

CHARITÉ

Analysing Laboratory Data: A Multivariate Task

Points to consider

- **Dimensions**
  genome data $p >> n$ (thousands $>>$ hundreds and less)
  lab data p ($<$) n (dozens ($<$) hundreds and less)
- **Distributions**
  univariate normality not really critical $\rightarrow$ use transformations
  checking multivariate normality real challenge, practically impossible
- **A possible solution**
  GlobalAncova implements distribution free multivariate tests in the framework of Linear Models' structure

Hummel M, Meister R, Mansmann U. (2007) GlobalANCOVA: Exploration and Assessment of Gene Group Effects.

Bioinformatics

CHARITÉ

Univariate Linear Model

$$x^{(i)} = m^{(i)} + \xi^{(i)} = \begin{pmatrix} 1 & c_{11} & \cdots & c_{1d} \\ \vdots & \vdots & & \vdots \\ 1 & c_{n1} & \cdots & c_{nd} \end{pmatrix} \times \begin{pmatrix} \beta_{i0} \\ \vdots \\ \beta_{id} \end{pmatrix} + \begin{pmatrix} \xi_1^i \\ \vdots \\ \xi_n^i \end{pmatrix}$$

$$= C \times \beta_i^t + \xi^{(i)}$$

Global Linear Model

$$X = \begin{pmatrix} x^{(1)} \\ \vdots \\ x^{(p)} \end{pmatrix} = \begin{pmatrix} C & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & C \end{pmatrix} \times \begin{pmatrix} \beta_1^t \\ \vdots \\ \beta_p^t \end{pmatrix} + \begin{pmatrix} \xi^{(1)} \\ \vdots \\ \xi^{(p)} \end{pmatrix}$$

$$= \mathbf{C} \times \beta + \xi$$

$$Cov(\xi) = \Sigma$$

Test Statistic

$$RSS = \sum_{ij} (\widehat{\xi}_j^i)^2$$

$$F_{GA} = \frac{RSS_{RM} - RSS_{FM}}{RSS_{FM}} \times \frac{n-q}{q-r}$$

q - Number of parameters in the full model

r - Number of parameters in the reduced model

CHARITÉ

RCT on Diabetes Test for Treatment Effects

call

```
library(GlobalAncova)
eff.treat <- GlobalAncova( xx = dat.post,
            full.formula = ~ TREAT + CEN + SEX + AGE,
            red.formula  = ~ CEN + SEX + AGE,
            model.dat = info,
            perm = 10000)§
```

CHARITÉ

Test Result

```
$effect

          ¨TREATDose 1¨ ¨TREATDose 2¨ ¨TREATDose 3¨

$ANOVA
```

	SSQ	DF	MS
Effect	4.645021	66	0.07037911
Error	332.690789	3674	0.09055275

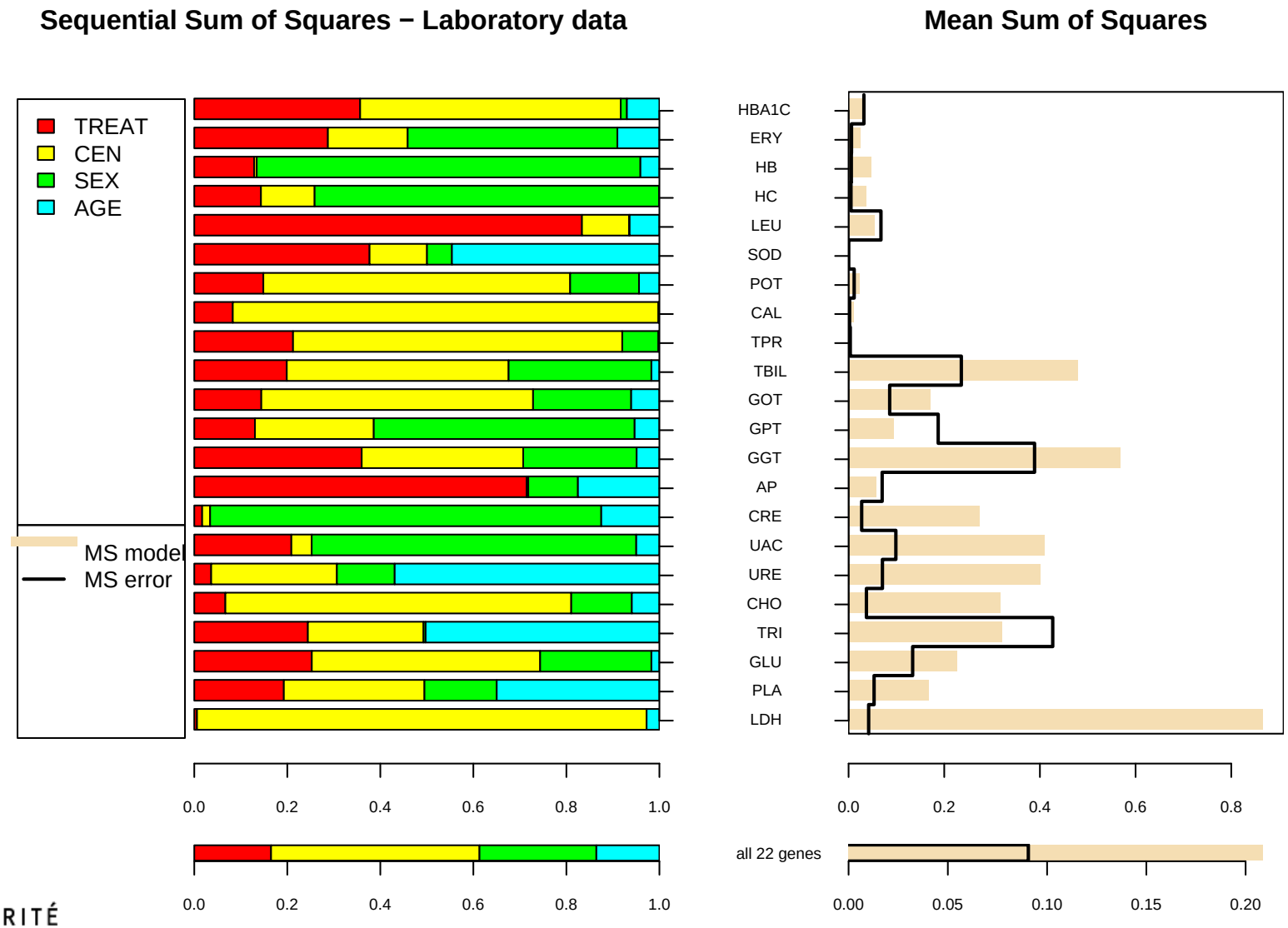
```
$test.result

                      [,1]
F.value   0.7772168
p.perm    0.7355000

§

No significant treatment effect
```

RCT on Diabetes Plot.all()



RCT on Diabetes Test for Center Effect

call

```
eff.cen <- GlobalAncova ( xx = dat.post,
           formula.full = ~ TREAT + CEN + SEX + AGE,
           formula.red  = ~ TREAT + SEX + AGE,
           model.dat = info,
           perm = 10000)§
```

CHARITÉ

RCT on Diabetes Test for Center Effect

Test Results

$effect

¨CEN2¨ ¨CEN3¨ ¨CEN4¨ ¨CEN5¨

$ANOVA

	SSQ	DF	MS
Effect	23.64064	88	0.26864363
Error	319.79803	3608	0.08863582

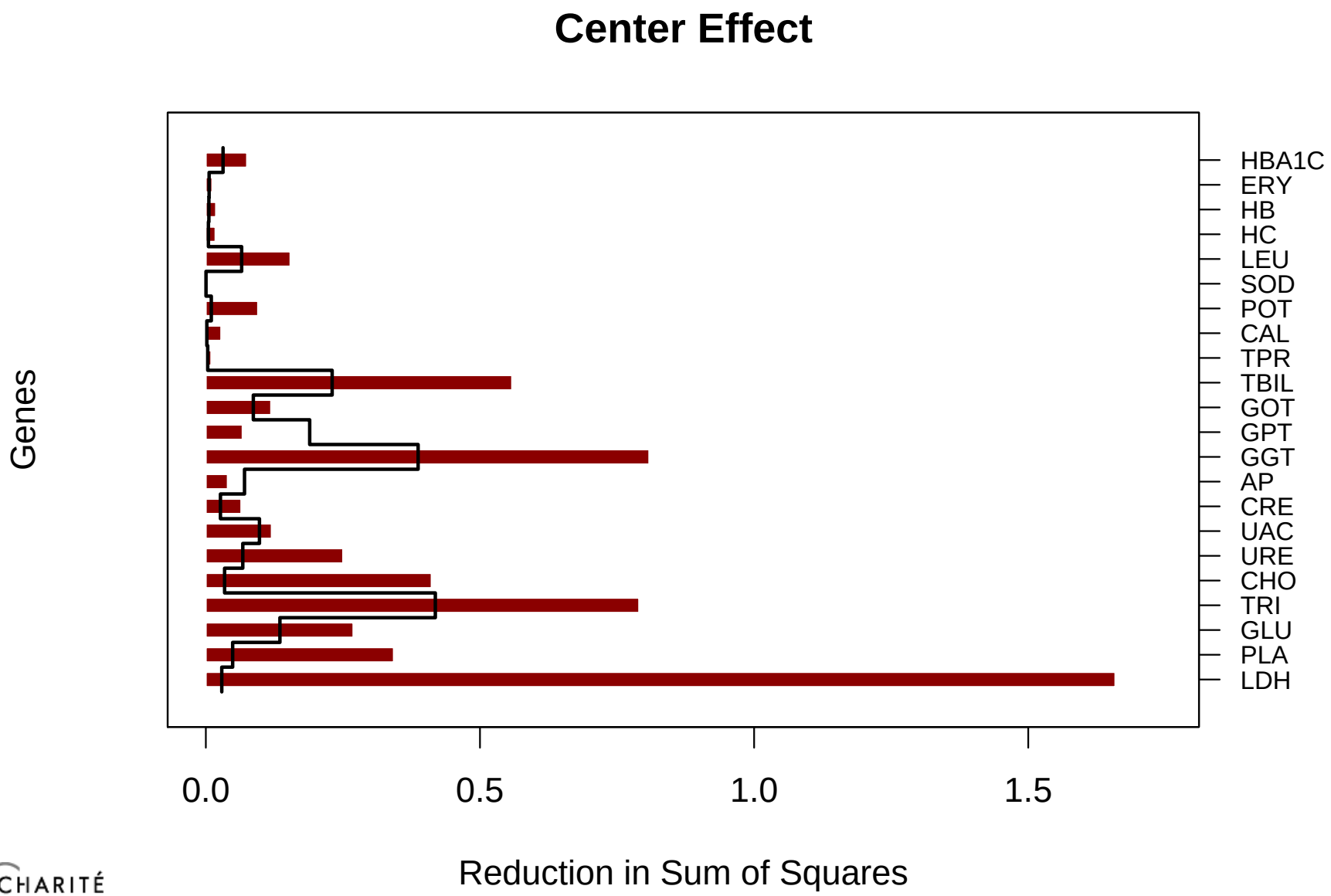
$test.result

	[,1]
F.value	3.03087
p.perm	0.00000

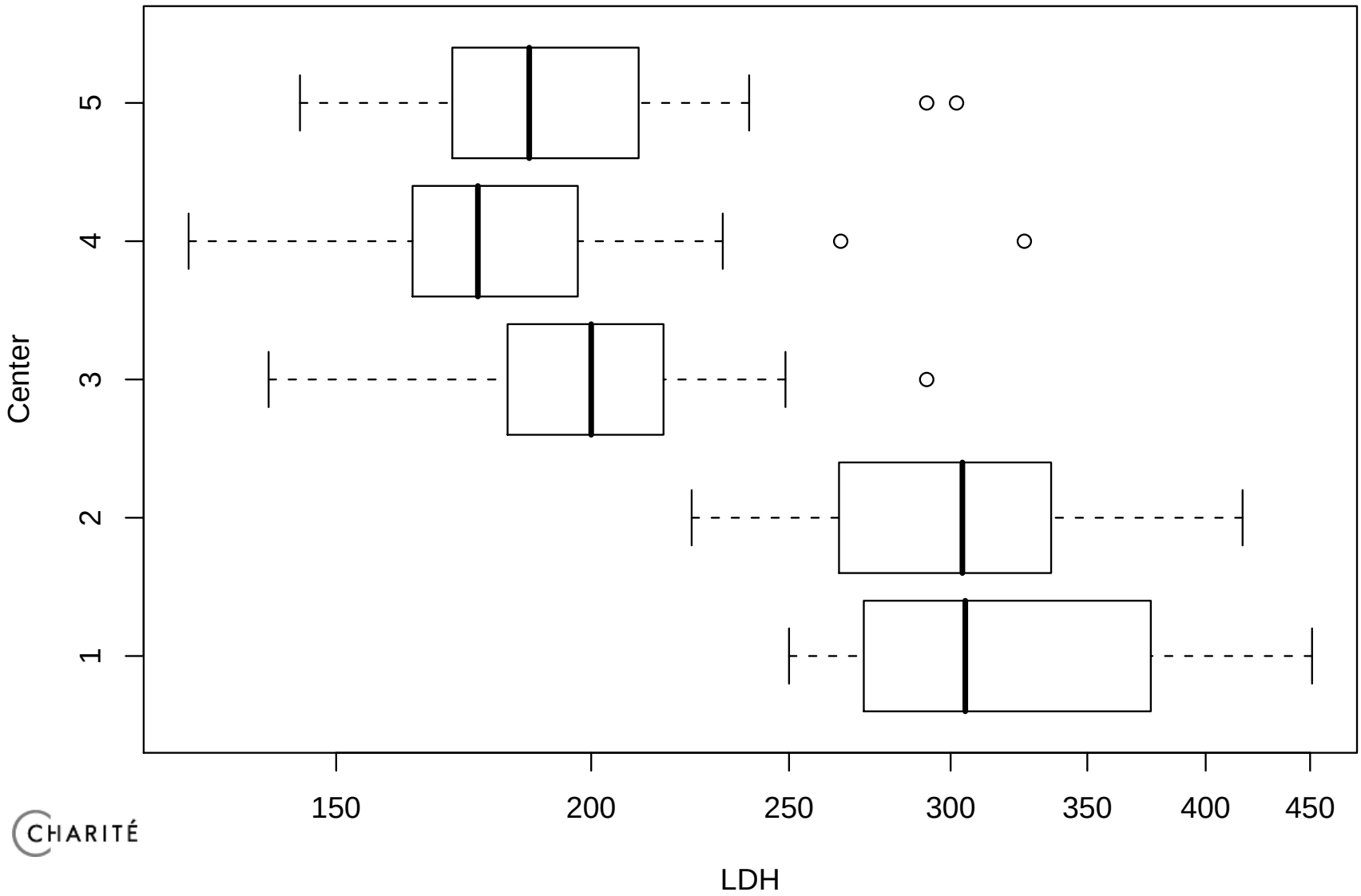
§

Significant center effect

RCT on Diabetes Test For Center Effect



RCT on Diabetes Center Effect - LDH



RCT on Diabetes LDH normally distributed?

RCT on Diabetes Center Effect Adjusted for Baseline

Adjustment

- Adjustment for (genewise) covariates possible
- Changes from baseline values

call

```
eff.cen <- GlobalAncova ( xx = dat.post - dat.pre,
           formula.full = ~ TREAT + CEN + SEX + AGE,
           formula.red  = ~ TREAT + SEX + AGE,
           model.dat = info,
           perm = 10000 )
```

CHARITÉ

RCT on Diabetes Center Effect Adjusted for Baseline

Test Results

```
$effect

                    ¨CEN2¨ ¨CEN3¨ ¨CEN4¨ ¨CEN5¨
```

$ANOVA

	SSQ	DF	MS
Effect	5.588282	88	0.0635032
Error	209.899348	3608	0.0581761

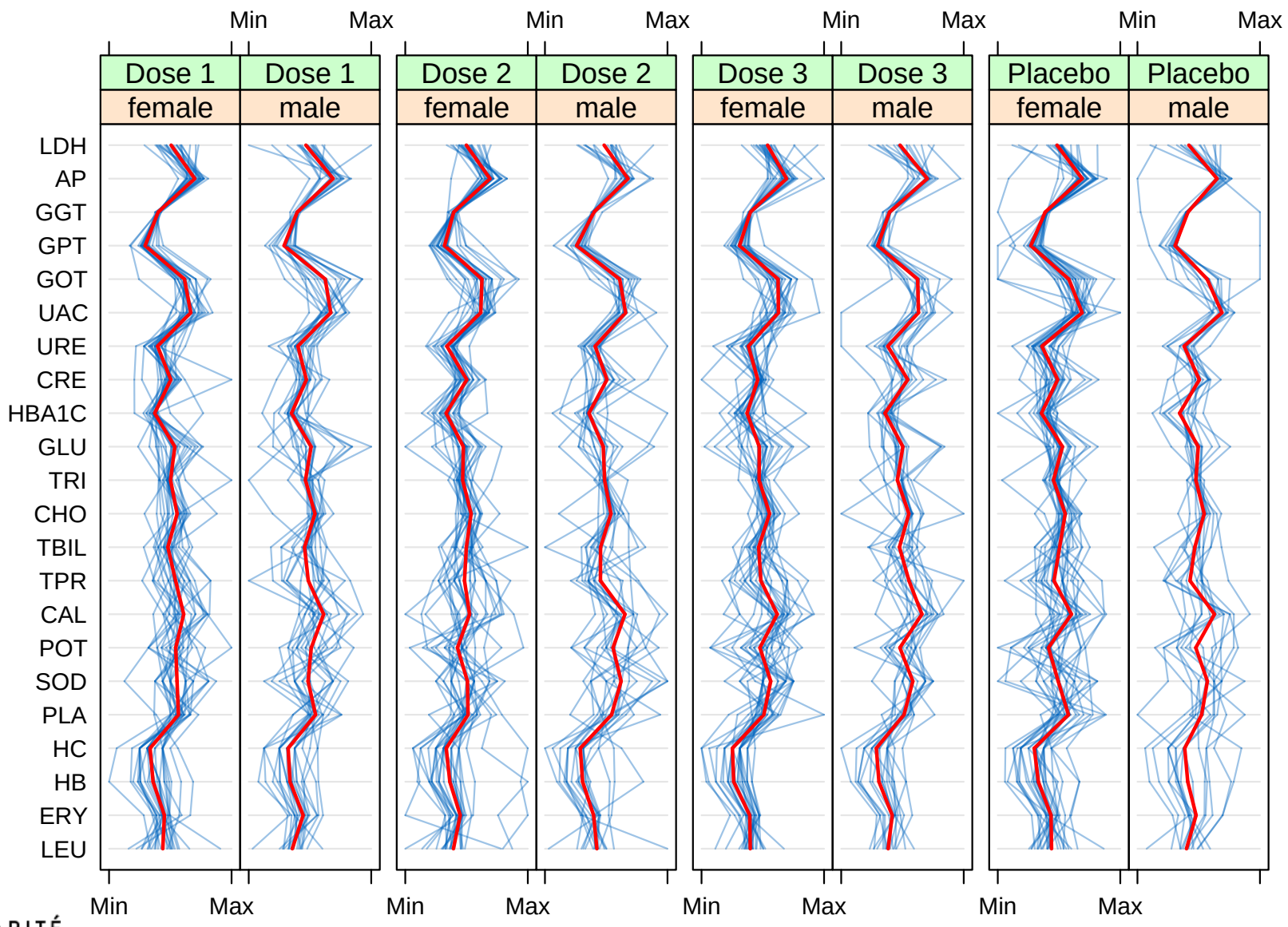
$test.result

	[,1]
F.value	1.091569
p.perm	0.324800

§

Center effect no more significant

RCT on Diabetes Laboratory Profiles Adjusted for Baseline



CHARITÉ

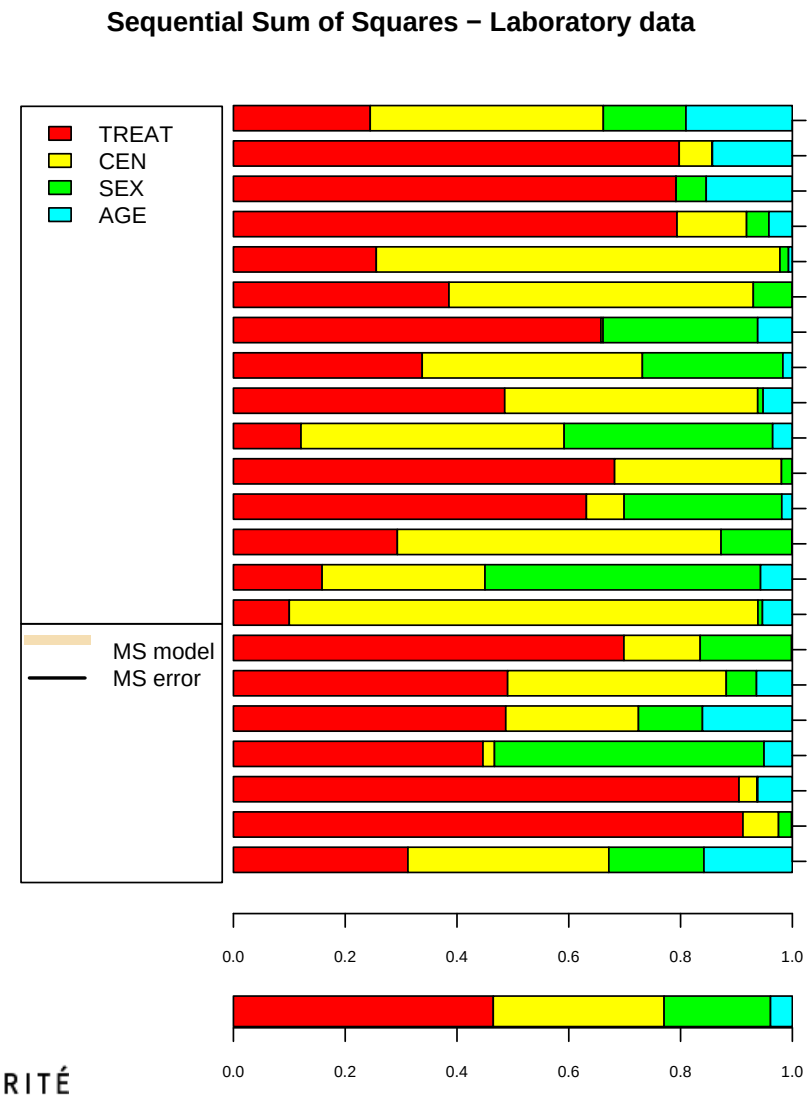
Summary

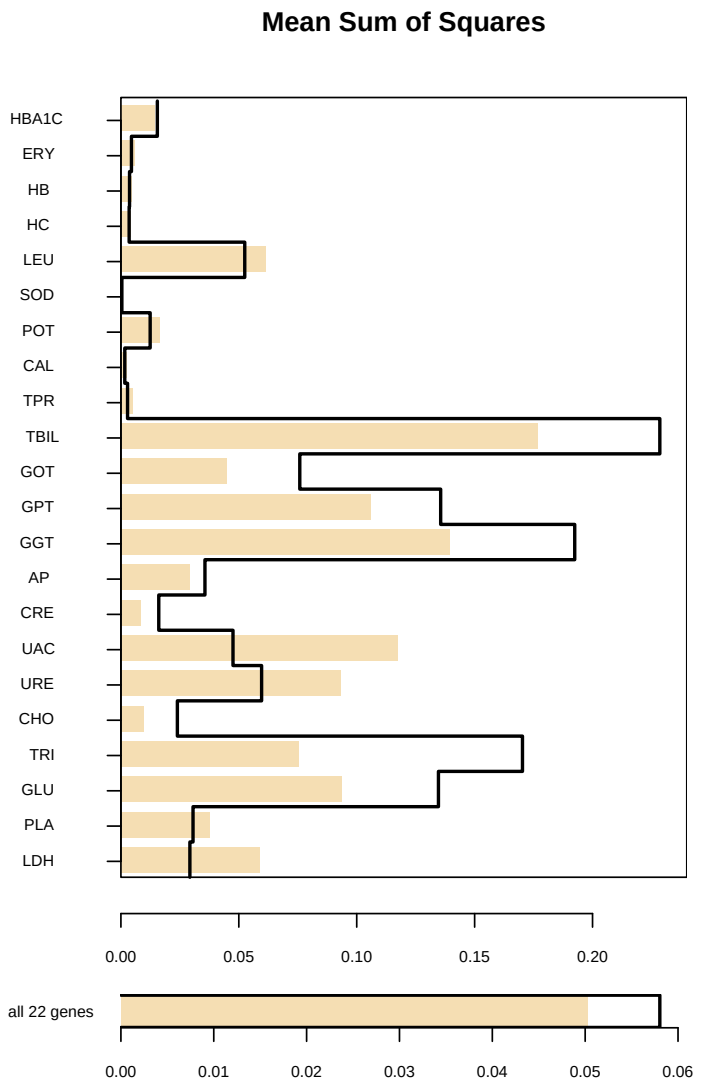
Conclusion

- Multivariate analysis of laboratory data
  - yealds important results
  - can be conveniently performed with GlobalAncova
- Adjustment is neccessary
- try it: www.bioconductor.org

CHARITÉ

1 more slide


Sequential Sum of Squares − Laboratory data
TREAT
CEN
SEX
AGE
MS model
MS error
0.0
0.2
0.4
0.6
0.8
1.0


Mean Sum of Squares
HBA1C
ERY
HB
HC
LEU
SOD
POT
CAL
TPR
TBIL
GOT
GPT
GGT
AP
CRE
UAC
URE
CHO
TRI
GLU
PLA
LDH
all 22 genes
0.00
0.05
0.10
0.15
0.20
0.00
0.01
0.02
0.03
0.04
0.05
0.06

yet another slide