Stratified medicine using the partykit package

Heidi Seibold, Achim Zeileis and Torsten Hothorn

Epidemiology, Biostatistics and Prevention Institute

University of Zurich

University of Zurich[UZH]

July, 2015

What is stratified medicine?

Identifying subgroups of patients for whom the treatment has a different effect than for others.

Effect is:

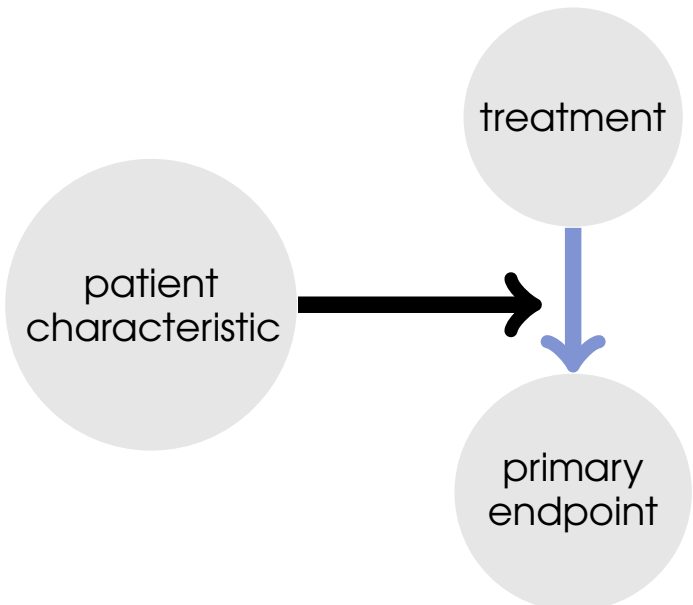
- Stronger
- Lower
- Contrary

Subgroups are defined by patient characteristics.

What is stratified medicine?

Goal: Find subgroups

$\hat{=}$ covariate $\times$ treatment interactions



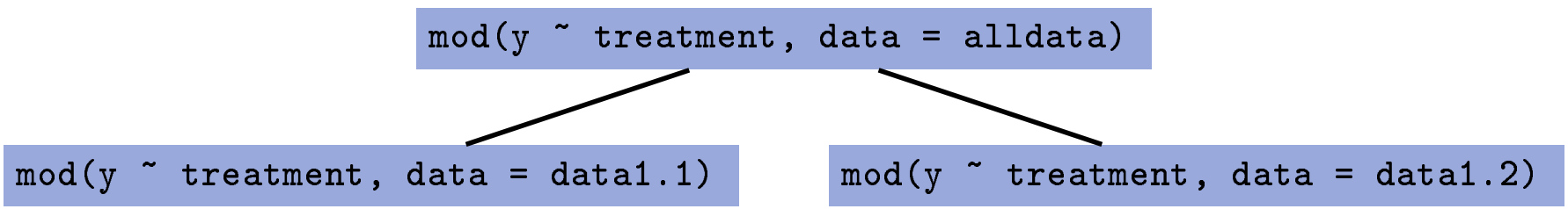
MOB

Model-based recursive partitioning (1)

Start with a parametric model

⇒ Find partitions based on patient characteristics.

⇒ Fit separate models in partitions.



`mod` can be e.g. `lm`, `glm`, `survreg`, ...

Partitioning

How to find the Partitions?

Test independence between **score function** and **each patient characteristic**.
$\rightarrow$ Detect instabilities in the intercept and treatment parameter.

- Partition if global p-value smaller than significance level
- Use as split variable the one corresponding to the smallest p-value

Implementation

Example: glm

```
my.glm <- function(data, weights, family = binomial) {

  mod <- glm(y ~ treatment, data = data,
             subset = weights > 0, family = family)

  # extract score function
  ef <- sandwich::estfun(mod)
  ret <- matrix(0, nrow = nrow(data), ncol = ncol(ef))
  ret[weights > 0,] <- as.matrix(ef)
  ret
}


tree <-  ctree(y + treatment ~ z1 + z2 + z3,
               data, ytrafo = my.glm)
```

Data

PRO-ACT database (2)

- Amyotrophic lateral sclerosis (ALS) patients
- Data of several clinical trials
- Treatment of interest:
  Riluzole, only approved drug against ALS (no cure)
- Primary endpoints of interest:
  ALS Functional Rating Scale (ALSFRS)

Question: Do all patients benefit the same from Riluzole?

ALS Functional Rating Scale: ALSFRS

Model

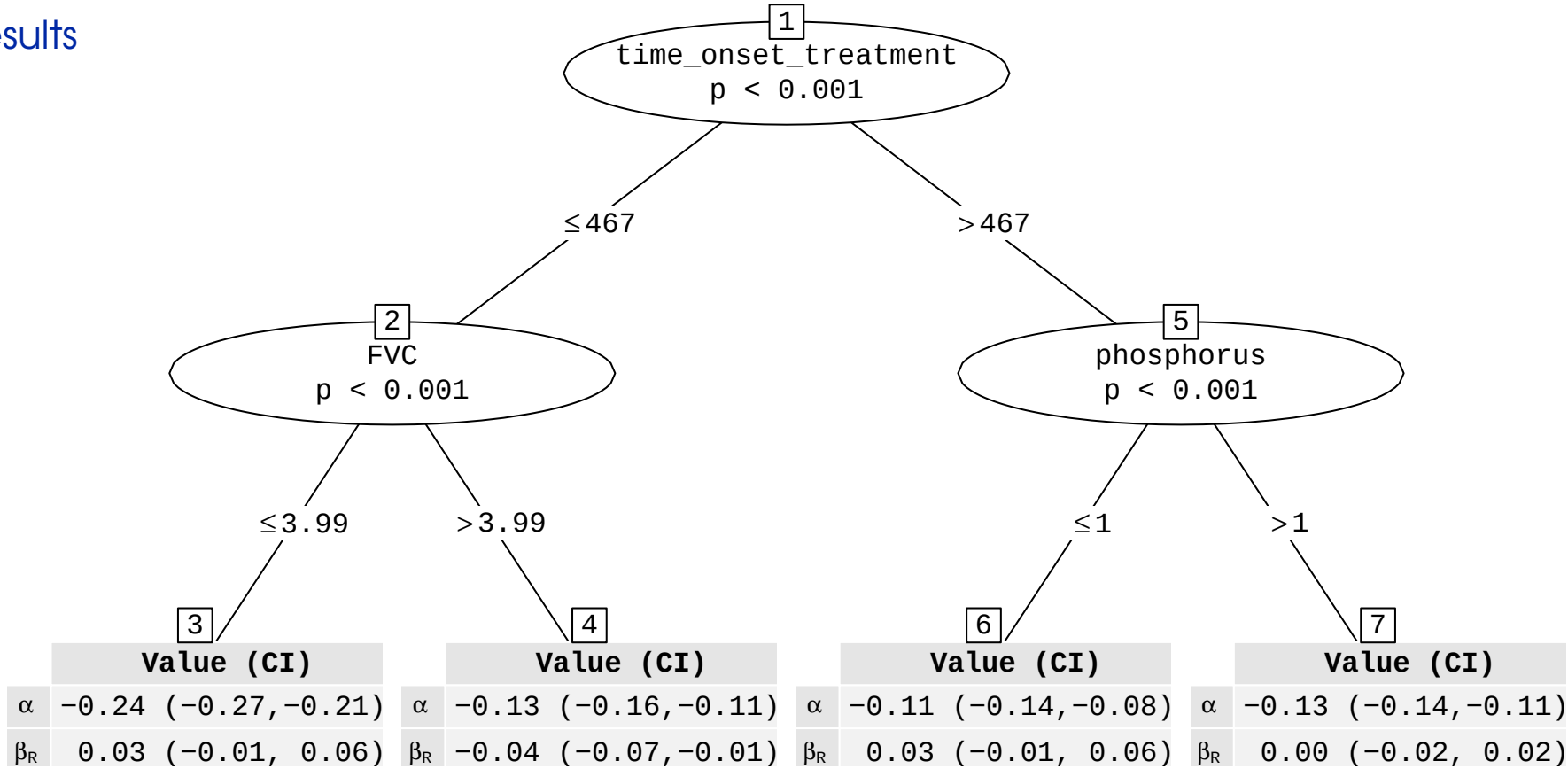
Gaussian model for the ALSFRS after six months taking into account the ALSFRS at the start of the treatment.

```
glm(ALSFRS6 ~ Riluzole + offset(log(ALSFRS0)),
    family = gaussian(link = "log"))
```

	Value (CI)
α	−0.16 (−0.17,−0.15)
β_R	0.01 (0.00, 0.03)

ALSFRS example

Results

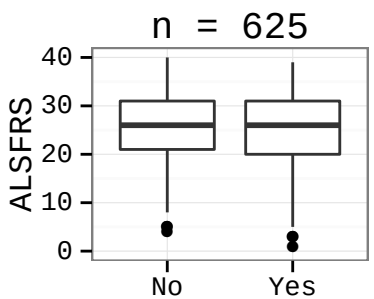


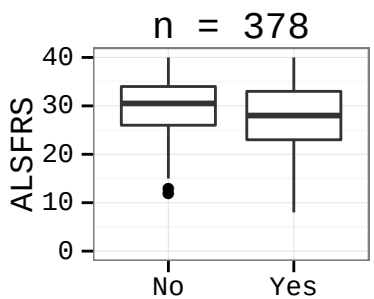
	Value (CI)
α	−0.24 (−0.27,−0.21)
β_R	0.03 (−0.01, 0.06)

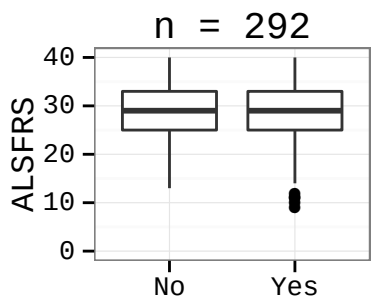
	Value (CI)
α	−0.13 (−0.16,−0.11)
β_R	−0.04 (−0.07,−0.01)

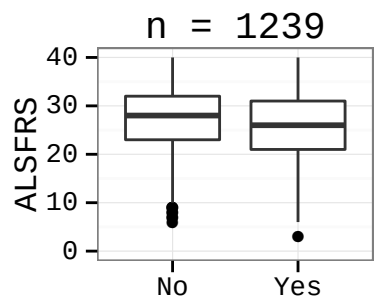
	Value (CI)
α	−0.11 (−0.14,−0.08)
β_R	0.03 (−0.01, 0.06)

	Value (CI)
α	−0.13 (−0.14,−0.11)
β_R	0.00 (−0.02, 0.02)









Summary/Outlook

- Interpretability
- Easy to use, flexible software
- More complex models possible, e.g.
  $\rightarrow$ More than two treatment options
  $\rightarrow$ Further covariate in the model
  $\rightarrow$ Multivariate outcomes
- Personalized instead of stratified medicine with cforest instead of ctree

Literature

A. Zeileis, T. Hothorn, K. Hornik
Model-Based Recursive Partitioning
Journal of Computational and Graphical Statistics, 2008.

Massachusetts General Hospital, Neurological Clinical Research Institute
Pooled Resource Open-Access ALS Clinical Trials Database
`https://nctu.partners.org/ProACT/`, 2013.

Contact me: heidi.seibold@uzh.ch