



Model-based Recursive Partitioning for Subgroup Analyses

Torsten Hothorn; joint work with Heidi Seibold and Achim Zeileis

Subgroup analyses

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

Effect is:

- Stronger
- Lower
- Contrary

than the average treatment effect.

Suitable models promise better prediction of treatment effect and thus individualised treatments.

Situations of interest for subgroup analyses

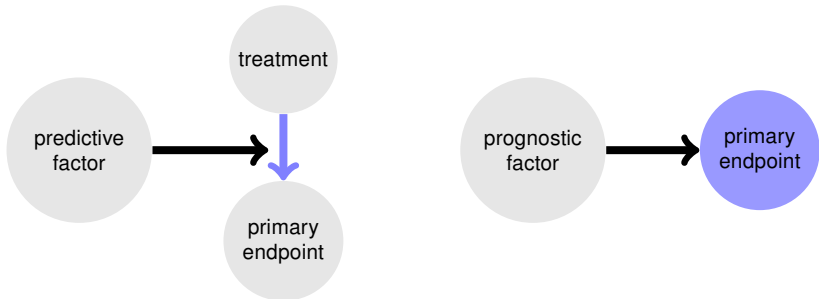
EMA (2014, 1):

1. “The clinical data presented are overall statistically persuasive with therapeutic efficacy demonstrated globally. It is of interest to verify that the conclusions of therapeutic efficacy (and safety) apply consistently across subgroups of the clinical trial population.”
2. “The clinical data presented are overall statistically persuasive but with therapeutic efficacy or benefit/risk which is borderline or unconvincing and it is of interest to identify post-hoc a subgroup, where efficacy and risk-benefit is convincing.”
3. “The clinical data presented fail to establish statistically persuasive evidence but there is interest in identifying a subgroup, where a relevant treatment effect and compelling evidence of a favourable risk-benefit profile can be assessed.”

Subgroup analyses

Goal: Find predictive factors

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



Covariate \times treatment interactions

- Subgroups known: incorporate subgroup \times treatment interaction in linear predictor.
- (Few) Categorical covariates at few levels: incorporate covariate \times treatment interaction in linear predictor.
- Many, potentially numeric covariates: Interactions hard to interpret, difficult to derive subgroups.
- Common approach: Automated interaction detection.
- BUT: Ordinary trees don't know treatment effect parameters.
- We need both parametric models AND trees.

Solution: Model-based recursive partitioning (MOB).

MOB Basics

MOB: Model-based recursive partitioning (Zeileis, Hothorn, Hornik, JCGS, 2008, 3)

Start with model $\mathcal{M}((Y, \mathbf{X}), \vartheta)$ with

$$\vartheta = \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \nu \end{pmatrix} \begin{array}{l} \text{intercept(s)} \\ \text{treatment effect} \\ \text{other parameter(s) of interest} \\ \text{nuisance parameter(s),} \end{array}$$

which fits data $(Y, \mathbf{X}) \in (\mathcal{Y}, \mathcal{X})$.

MOB Basics

Estimation

$$\hat{\vartheta} = \arg \min_{\vartheta} \sum_{i=1}^N \Psi((y, \mathbf{x})_i, \vartheta)$$

or equivalently solving the score equation

$$\sum_{i=1}^N \psi((y, \mathbf{x})_i, \vartheta) = 0$$

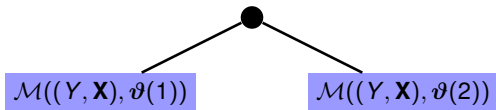
Ψ : objective function

ψ : score function; gradient of Ψ

MOB Basics

Maybe the treatment effect is not the same for all patients, but depends on their characteristics (covariates) .

- ⇒ Find partitions $\{\mathcal{B}_b\}$ ($b = 1, \dots, B$) based on patient characteristics $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_J) \in \mathcal{Z}$
- ⇒ Fit separate models $\mathcal{M}((Y, \mathbf{X}), \vartheta(b))$ in partitions.



$$\vartheta(b) = (\alpha(b), \beta(b), \gamma, \nu)^\top$$

MOB Basics

If partition $\{\mathcal{B}_b\}$ is known, the partitioned model parameters $\vartheta(b)$ could be estimated by minimizing the segmented objective function:

$$(\hat{\vartheta}(b))_{b=1,\dots,B} = \arg \min_{\vartheta(b)} \sum_{i=1}^N \sum_{b=1}^B \mathbb{1}(\mathbf{z}_i \in \mathcal{B}_b) \Psi((y, \mathbf{x})_i, \vartheta(b))$$

Subgroup-specific intercept and treatment parameters can be written as functions of the partitioning variables

$$\alpha(\mathbf{z}) = \sum_{b=1}^B \mathbb{1}(\mathbf{z} \in \mathcal{B}_b) \cdot \alpha(b) \quad \text{and} \quad \beta(\mathbf{z}) = \sum_{b=1}^B \mathbb{1}(\mathbf{z} \in \mathcal{B}_b) \cdot \beta(b).$$

Partitioning

How to find the partitions?

⇒ Test:

$$H_0^{\alpha,j} : \psi_{\alpha}((Y, \mathbf{X}), \hat{\vartheta}) \perp \mathbf{Z}_j$$
$$H_0^{\beta,j} : \psi_{\beta}((Y, \mathbf{X}), \hat{\vartheta}) \perp \mathbf{Z}_j, \quad j = 1, \dots, J$$

$\psi_{\alpha}, \psi_{\beta}$ partial derivatives of Ψ with respect to α/β .

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

Example: Linear model

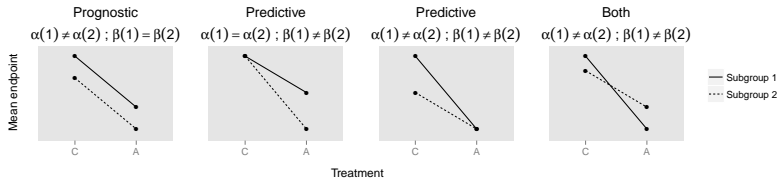
Two-arm trial comparing active (A) to control (C):

$$Y|\mathbf{X} = \mathbf{x} \sim \mathcal{N}(\alpha + \beta x_A + \gamma x_{\text{stratum}}, \sigma^2).$$

$$Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} \sim \mathcal{N}(\alpha(\mathbf{z}) + \beta(\mathbf{z})x_A + \gamma x_{\text{stratum}}, \sigma^2),$$

$$\begin{aligned} \psi((y, \mathbf{x}), \hat{\boldsymbol{\theta}}) &= \left(\begin{array}{c} \frac{\partial \psi((y, \mathbf{x}), \theta)}{\partial \alpha} \\ \frac{\partial \psi((y, \mathbf{x}), \theta)}{\partial \beta} \end{array} \bigg|_{\theta = \hat{\boldsymbol{\theta}}} \right)^{\top} \\ &= \frac{1}{\sigma^2} \left(\begin{array}{c} y - (\hat{\alpha} + \hat{\beta}x_A + \hat{\gamma}x_{\text{stratum}}) \\ (y - (\hat{\alpha} + \hat{\beta}x_A + \hat{\gamma}x_{\text{stratum}})) \cdot x_A \end{array} \right)^{\top} \end{aligned}$$

Example: Linear model



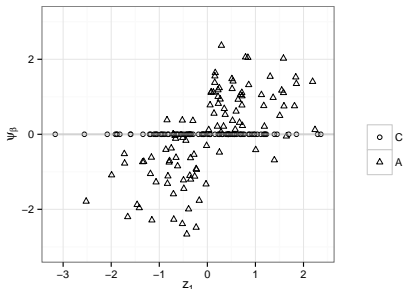
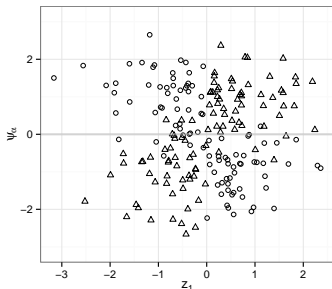
Example: Linear model

Linear model:

$$Y|\mathbf{X} = \mathbf{x} \sim \mathcal{N}(\alpha + \beta x_B, \sigma^2)$$

Data generating process (Loh, He, Man, 2014, 4):

$$Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} \sim \mathcal{N}(1.9 + 0.2 \cdot x_A + 1.8 \cdot \mathbf{1}(z_1 < 0) + 3.6 \cdot \mathbf{1}(z_1 > 0) \cdot x_A, 0.7)$$



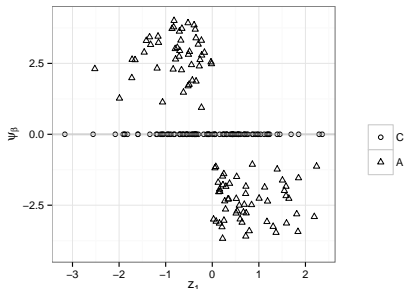
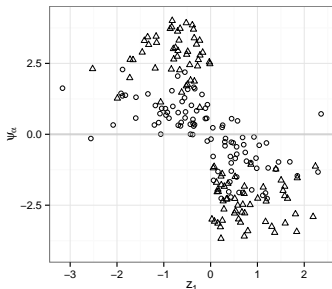
Example: Linear model

Linear model:

$$Y|\mathbf{X} = \mathbf{x} \sim \mathcal{N}(\alpha + \beta x_B, \sigma^2)$$

Data generating process:

$$Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} \sim \mathcal{N}(1.9 + 0.2 \cdot x_A + 1.8 \cdot \mathbf{1}(z_1 < 0) + 3.6 \cdot \mathbf{1}(z_1 < 0) \cdot x_A, 0.7)$$



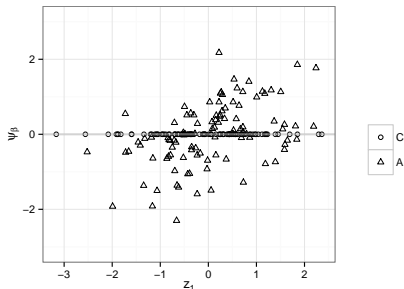
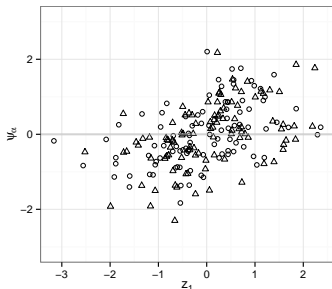
Example: Linear model

Linear model:

$$Y|\mathbf{X} = \mathbf{x} \sim \mathcal{N}(\alpha + \beta x_B, \sigma^2)$$

Data generating process:

$$Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} \sim \mathcal{N}(2 \cdot x_A + \mathbb{1}(z_1 > 0), 0.7)$$



Partitioning effects of Riluzole on ALS patients

PRO-ACT database (2)

- Amyotrophic lateral sclerosis (ALS) patients
- Data of several clinical trials
- Treatment of interest: Riluzole
- Primary endpoints of interest:
 - ALS Functional Rating Scale (ALSFRS)
 - ALSFRS items
 - Survival time

Question

Riluzole modestly prolongs life expectancy

But: Are there any groups of patients for whom it is better or worse?

⇒ Subgroup analysis

ALS functional rating scale: Measure of functional status of ALS patients

Sum-score of ten items ($0 < 1 < 2 < 3 < 4$):

- speech
- salivation
- swallowing
- handwriting
- cutting food and handling utensils,
- dressing and hygiene
- turning in bed and adjusting bed clothes
- walking
- climbing stairs
- breathing

$$\mathbb{E} \left(\frac{\text{ALSFRS}_6}{\text{ALSFRS}_0} \middle| X = x \right) = \frac{\mathbb{E}(\text{ALSFRS}_6 | X = x)}{\text{ALSFRS}_0} = \exp\{\alpha + \beta x_R\}$$

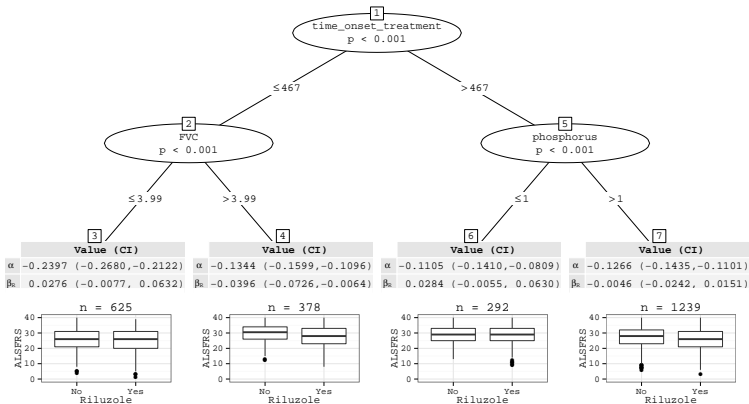
or equivalently

$$\mathbb{E}(\text{ALSFRS}_6 | X = x) = \exp\{\alpha + \beta x_R\} \cdot \exp\{\log(\text{ALSFRS}_0)\}$$

GLM with log-link and offset

ALSFRS

Results



ALSFRS items

Unidimensionality of ALSFRS is questionable.

Two components to consider:

1. Baseline adjustment

Adjust for item score at beginning of treatment

⇒ compute separate models

2. Multivariate primary endpoint

Look at 10 items simultaneously

⇒ compute 10 item-models in every node

Score matrix of dimension $n \times p \cdot 10$

ALSFRS items

Each item ordinal with values 0, ..., 4

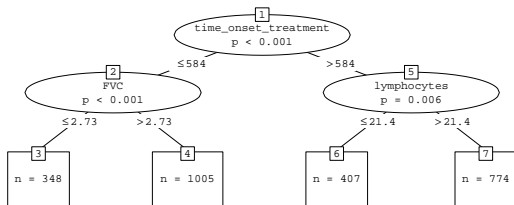
⇒ proportional odds model adjusted for baseline:

$$\mathbb{P}(Y_6 \leq r | Y_0 = k, X = x) = \frac{\exp(\alpha_{rk} - \beta_k x_R)}{1 + \exp(\alpha_{rk} - \beta_k x_R)} \quad \text{for } k = 0, \dots, 4$$

Compute stratified permutation tests treating baseline item values as blocks.

ALSFRS items

Results



ALSFRS items

Results

Item	No.	Start	Node 3	Node 4	Node 6	Node 7	
Speech	1	0					
		1					
		2			-0.27		
		3		0.33	0.04	-0.27	-0.06
		4		0.84	0.11		
Salivation	2	0					
		1					
		2		0.22	0.43	1.36	-0.20
		3		0.15	-0.26	-0.24	-0.05
		4		0.49	-0.03		
Swallowing	3	0					
		1					
		2		0.35	-0.89	1.51	-0.75
		3		0.57	-0.36	-0.40	0.35
		4		0.62		0.15	0.28
Handwriting	4	0				-1.45	
		1			-1.15	-0.36	-0.08
		2			-0.54		0.04
		3		-0.13	0.14	-0.08	-0.28
		4		-0.10	0.04		-0.14
Cutting	5	0					
		1			-0.01	-0.79	
		2			0.15		0.48
		3		-0.03	-0.07	0.10	0.52
		4		0.13	-0.09	-0.14	-0.21
		⋮		⋮		⋮	

ALSFRS items

Results

Item	No.	Start	Node 3	Node 4	Node 6	Node 7	
Hygiene	6	0					
		1					
		2			-0.11		-0.37
		3		-0.22	-0.37	0.14	0.27
		4		0.26	0.01	0.14	0.30
Bed	7	0					
		1					
		2				-0.03	0.29
		3		0.15	-0.32	-0.12	-0.05
		4		-0.21	-0.10	-0.11	-0.35
Walking	8	0					
		1					
		2				0.48	-0.04
		3			0.11		0.46
		4		0.51	0.13		
Stairs	9	0					
		1			-0.02	-0.01	-0.39
		2				-0.80	0.07
		3			0.26	-0.65	-0.16
		4		1.01	0.06	0.72	0.29
Respiratory	10	0					
		1					
		2					
		3					
		4		0.58	-0.08		

Survival time

We present two ways of modeling:

- Weibull model
(parametric survival model)
- Cox model
(semiparametric survival model)

Weibull model

$$\mathbb{P}(Y \leq y | X = x) = F \left(\frac{\log(y) - \alpha_1 - \beta x_R}{\alpha_2} \right)$$

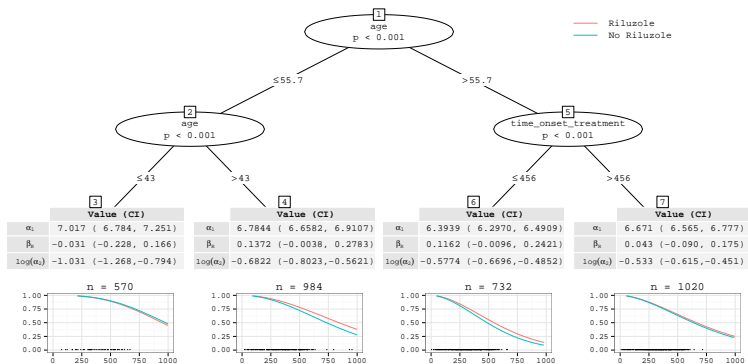
with F cumulative distribution function of Gompertz distribution

and $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$ intercept
scale parameter

α defines the shape of the baseline hazard \Rightarrow use as
"intercept"

Weibull model

Results



Cox model

$$\lambda(y|\mathbf{x}) = \lambda_0(y) \exp(\beta x_R)$$

Objective function: negative partial likelihood
(without baseline hazard)

⇒ No classical score function

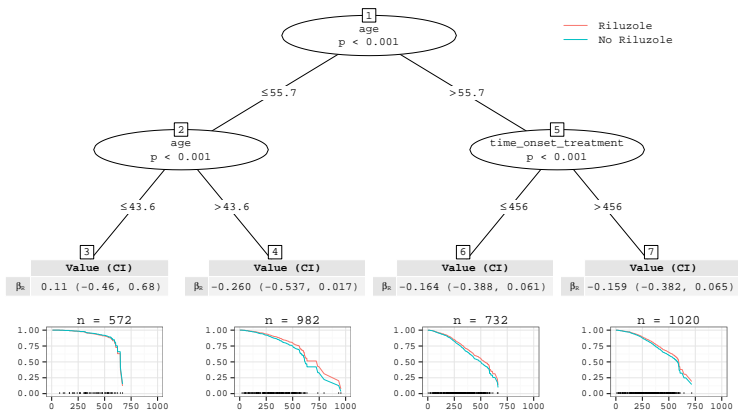
⇒ Use surrogate score function

Martingale residuals (as score with respect to the
baseline hazard/"intercept")

Score residuals (as score with respect to β)

Cox model

Results



Computational details

PRO-ACT data are available at <https://nctu.partners.org/ProACT/>.

The source code for reading and cleaning the database is provided in the **TH.data** package.

All computations were conducted using **partykit** (version 0.8-2) in the R system for statistical computing (version 3.1.2).

Simple implementation

Weibull model

```
library("partykit")





## Function to compute Weibull model and return score matrix
mywb <- function(data, weights, parm) {
  mod <- survreg(Surv(survival.time, cens) ~ Riluzole,
                 data = data, subset = weights > 0,
                 dist = "weibull")
  ef <- as.matrix(estfun(mod)[,parm])
  ret <- matrix(0, nrow = nrow(data), ncol = ncol(ef))
  ret[weights > 0,] <- ef
  ret
}

## Compute tree
tree <- ctree(fm, data = data, ytrafo = my.wb,
              control = ctree_control(maxdepth = 2,
                                     testtype = "Bonferroni"))
```


Summary

- MOB partitions a large class of models suitable for the treatment effect on the primary endpoint of interest.
- Score functions capture instabilities and thus help to identify predictive and prognostic variables.
- It is hard to differentiate between predictive and prognostic variables.
- Permutation tests suitable for all models (distribution free) with good small sample properties and error control.
- Multiplicity adjustment for subgroup-specific treatment effects unclear.

Literature

-  **European Medicines Agency** *EMA Guideline on the investigation of subgroups in confirmatory clinical trials (draft)*
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/02/WC500160523.pdf, 2014.
-  **Massachusetts General Hospital, Neurological Clinical Research Institute** *Pooled Resource Open-Access ALS Clinical Trials Database*
<https://nctu.partners.org/ProACT/>, 2013.
-  **A. Zeileis, T. Hothorn, K. Hornik** *Model-Based Recursive Partitioning*
Journal of Computational and Graphical Statistics, 2008.
-  **W. Loh, X. He, M. Man** *A regression tree approach to identifying subgroups with differential treatment effects*
<http://arxiv.org/abs/1410.1932>, 2014.