Treatment Selection to Improve Survival using the Predicted Individual Treatment Effect

Gerd Rosenkranz

Center for Medical Statistics, Informatics and Intelligent Systems
Medical University of Vienna
A-1090 Vienna, Austria

EFSPI Meeting on Oncology and Survival Analysis
17-Nov-2017, BMS, Brussels

Joint work with Nicolas Ballarini, Vienna.

Prepared under funding provided by the UK Medical Research Council, Project No. MR/M005755/1.

Outline

What this talk is about

Consider the following situation:

- A patient presents to a physician and receives a diagnosis
- The physician can choose between two treatments to remedy the condition of the patient
- How can the physician select the right, i.e., most beneficial treatment for the patient?

This talk is about:

- How can statistics support the physician to get “the right treatment to the right patient?”[1]

[1] The mantra of personalized medicine

Learning dataset

Assume we have the following:

- $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$, the characteristics (covariates) of N individuals from some population
- $\{Z_1, \ldots, Z_N\}$, $Z_n \in \{0, 1\}$ the treatments assigned to the N individuals
- $\{Y_1, \ldots, Y_N\}$ the outcomes under $\{Z_1, \ldots, Z_N\}$
- For survival data, $Y_n = (T_n, \delta_n)$, a survival time and a censoring indicator

$\mathcal{L} = \{(Y_n, \mathbf{X}_n, Z_n), n = 1, \ldots, N\}$ is called a **learning dataset**

Remark:

- Treatment assignments should be independent of other covariates
- All covariates should be standardized, i.e, have zero mean and unit variance

Modeling the learning data

Consider a Cox regression model with treatment by covariate interaction terms:

$$\lambda(t, \mathbf{x}, z) = \lambda_0(t) \exp\left\{ \beta z + \sum_{k=1}^{K} \gamma_k x_k + z \sum_{k=1}^{K} \delta_k x_k \right\} \qquad (1)$$

Define the **predicted individual treatment effect** (PITE) of $z = 1$ relative to $z = 0$ given x in terms of the log-hazard ratio:

$$D(\mathbf{x}) = \log \lambda(t, \mathbf{x}, 1) - \log \lambda(t, \mathbf{x}, 0) = \beta + \sum_{k=1}^{K} \delta_k x_k \qquad (2)$$

We favor treatment 1 over 0 if

$$D(\mathbf{x}) \leq c$$

for some threshold $c \leq 0$ determined from clinical relevance

Predicting the individual treatment (effect)

$D(\mathbf{x})$ can be estimated as a linear combination of the partial maximum likelihood estimates $\hat{\beta}(\mathcal{L})$ and $\hat{\delta}_k(\mathcal{L})$.

The predicted benefit of a future individual with covariate x_0 from the same population is then

$$\hat{D}(\mathbf{x}_0, \mathcal{L}) = \hat{\beta}(\mathcal{L}) + \sum_{k=1}^{K} \hat{\delta}_k(\mathcal{L}) x_{0k} \quad (3)$$

Estimates of the standard deviation of $\hat{D}(x_0, \mathcal{L})$ and confidence intervals for $D(\mathbf{x}_0)$ can be obtained in the usual way (Wald statistics, profile likelihood)

The predicted beneficial treatment is then

$$\hat{\eta}(\mathbf{x}_0, c, \mathcal{L}) = I[\hat{D}(\mathbf{x}_0, \mathcal{L}) \leq c]$$

Questions

1. What if the Cox model (1) is not correct because
   - a different type of model is more appropriate?
   - relevant covariates or interactions are missing?
2. Are all covariates included in the model relevant?
   - Can we select the relevant ones and drop the others?
   - How does selection affect estimation?

Model builder's dilemma:

- Too many non-informative covariates increase noise and decrease prediction precision
- Omission of informative covariates in the Cox model leads to biased estimates of the remaining parameters [1]

Variable selection

Forward selection[2]

1. Start with null model
2. Compute score statistics for every effect not in model
3. If the largest score statistics is significant at a pre-set entry level, the effect is added to the model
4. Repeat steps 2. and 3. until no effect meets the entry criterion.

Main issue:

- Standard errors of estimators and confidence intervals for parameters after selection
- Ignored in statistical practice and/or software until recently

[2] as implemented in PROC PHREG of SAS

Bootstrap smoothing [2]

1. Fit Cox model (1) with variable selection to learning data $\mathcal{L}$ to obtain an estimate $\hat{D}(\mathbf{x}_0, \mathcal{L})$ for $D(\mathbf{x}_0)$
2. For $b = 1, \ldots, B$, fit Cox model (1) with variable selection to bootstrap sample $\mathcal{L}^*_b$ from $\mathcal{L}$ to obtain $\hat{D}^*_b(\mathbf{x}_0) = \hat{D}(\mathbf{x}_0, \mathcal{L}^*_b)$
3. Obtain a $100(1-2\alpha)$% confidence interval for $D(\mathbf{x}_0)$ from the 100αth upper and lower percentile of the distribution of $\{\hat{D}^*_1(\mathbf{x}_0), \ldots, \hat{D}^*_B(\mathbf{x}_0)\}$

Remarks:

- The estimates $\hat{D}^*_b(\mathbf{x}_0)$ in step 2 above may be obtained from models different from the one based on $\mathcal{L}$ reflecting the uncertainty of the selection process
- In step 3 the infinitesimal jackknife variance estimator can be used to obtain a confidence interval [2]

Prediction error

- The probability to select treatment $z = 1$ for an individual with covariates $\mathbf{x}_0$ can be estimated from

$$\hat{Q}^*_B(\mathbf{x}_0, c, \mathcal{L}) = \frac{1}{B} \sum_{b=1}^{B} \hat{\eta}(\mathbf{x}_0, c, \mathcal{L}^*_b)$$

- An estimator of the probability to take the right treatment decision for an individual with $\mathbf{x}_0$ is

$$\begin{aligned}\hat{P}^*_B(\mathbf{x}_0, c, \mathcal{L}) &= \hat{Q}^*_B(\mathbf{x}_0, c, \mathcal{L})\hat{\eta}(\mathbf{x}_0, c, \mathcal{L}) \\ &\quad + [1 - \hat{Q}^*_B(\mathbf{x}_0, c, \mathcal{L})][1 - \hat{\eta}(\mathbf{x}_0, c, \mathcal{L})] \\ &= \frac{1}{B} \sum_{b=1}^{B} I[\hat{\eta}(\mathbf{x}_0, c, \mathcal{L}^*_b) = \hat{\eta}(\mathbf{x}_0, c, \mathcal{L})]\end{aligned}$$

The prostate cancer data set [3]

- 475 subjects randomized to placebo and 3 doses of diethyl stilbestrol
- Placebo and lowest dose combined to control, two highest doses combined to test treatment
- Primary endpoint: survival time

Patient characteristics (covariates):

- Bone metastases (BM) (0 = absent, 1 = present)
- Disease stage (DS) (3 or 4)
- Performance status (PF) (0 or 1)
- History of CV events (HX) (0 = absent, 1 = present)
- Age (AGE) (years)
- Weight (WT) (kg)

Kaplan-Meier plot (all data)

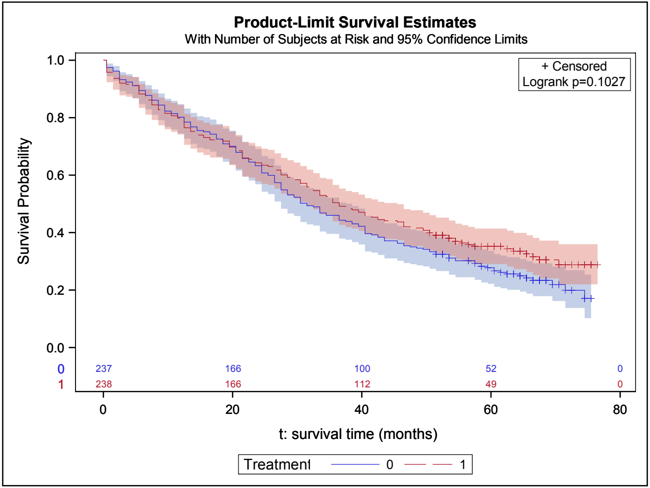

Figure 1: Survivor plot of the prostate cancer data by treatment.

Kaplan-Meier plot (subgroup)

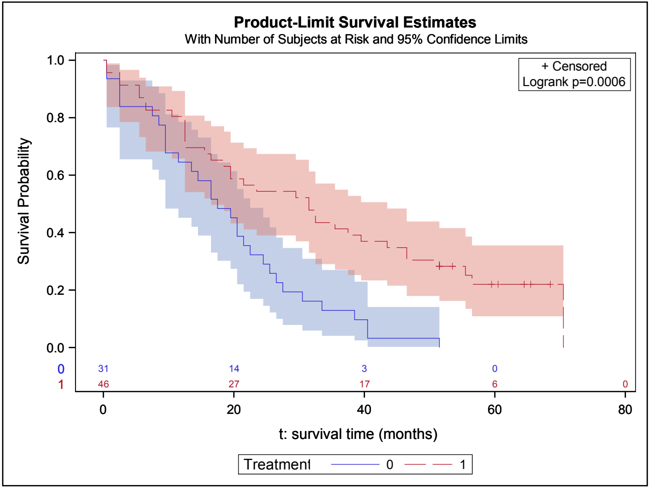

Figure 2: Survivor plot of the prostate cancer data for metastasized tumors by treatment.

Parameter estimates

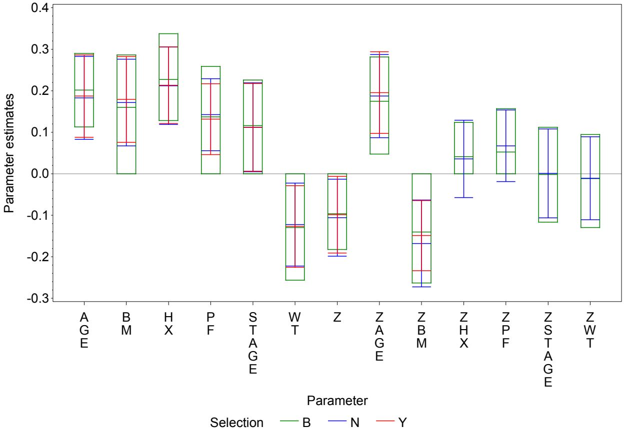

Figure 3: Parameter estimates and 90% confidence intervals. N stands for full model without variable selection, Y for forward selection without adjustment of the confidence intervals and B for forward selection with bootstrap derived intervals.

Variable selections

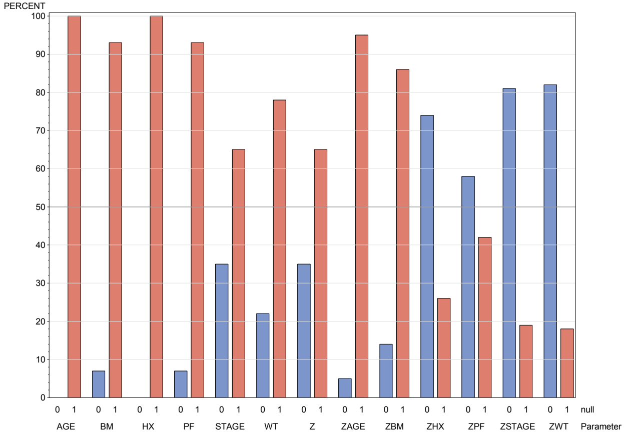

Figure 4: Percentage of bootstrap samples with selected (red bars) and not selected covariates (blue bars) ("Poor man's posteriors").

Some predictions

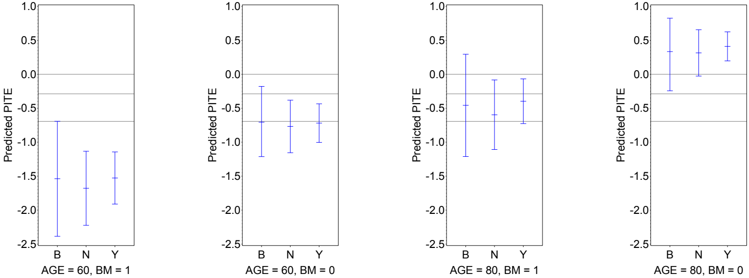

Figure 5: Predicted PITEs and 90% prediction intervals for different subject characteristics. N stands for no, Y for forward selection without adjustment and B for forward selection with bootstrap derived intervals. The 3 horizontal lines correspond to no, a 25% or a 50% hazard reduction of test versus control.

Younger subjects with bone metastases benefit most from test treatment while test may be harmful to older patients without bone metastases.The estimated rate of correct treatment predictions averaged over the population is 92% for a 50% hazard reduction.

Simulations

- Data generated from Cox model

$$\lambda(t, z, x) = \exp\{\beta z + \gamma x + \delta x z\}$$

with $\beta = \gamma = -0.5$, $\delta = -1$, $X \sim N(0, \sigma^2)$ with $\sigma = 0.5$ and $\Pr[Z = -0.5] = \Pr[Z = 0.5] = 0.5$.

- Censoring occurs after a total study duration of 5 units with uniform enrollment over 3 units (censoring rate of 6.6%)
- Data analyzed with Cox model

$$\mu(t, z, (x, \mathbf{x})) = \lambda(t, z, x) \exp\left\{\sum_{k=1}^{9} \gamma_k x_k + z \sum_{k=1}^{9} \delta_k x_k\right\}$$

i.e., with 9 additional non-informative covariates $X_k \sim N(0, \sigma^2)$

Results

Selection	β	γ	δ
None	-0.62 (0.75)	-0.64 (0.75)	-1.29 (0.78)
Forward	-0.55 (0.62)	-0.55 (0.70)	-1.06 (0.65)
Smoothed forward	-0.68 (0.84)	-0.67 (0.88)	-1.14 (0.87)
None	-0.50 (0.87)	-0.53 (0.85)	-1.05 (0.87)
Forward	-0.49 (0.88)	-0.53 (0.88)	-1.03 (0.86)
Smoothed forward	-0.51 (0.86)	-0.55 (0.84)	-1.07 (0.84)

Table 1: Mean parameter estimates and parameter coverage for total sample size of 100 (upper block) and 500 subjects (lower block) from 100 simulations with 100 bootstrap samples for smoothing.

Selective inference

- Inference conditional on the selected model
- Inference valid for all situations that result in the same model selection
- Equivalent to restricting the observations to a subset of the sample space and consequently to truncated distributions
- Exact for normal model regression [4], asymptotically valid in generalized linear models and survival analysis [5]
- Implemented for some variable selection procedures like forward selection, LAR and lasso

Closing remarks

- Models selecting treatment based on a PITE should ideally contain all relevant covariates but not more since non-informative variables increase noise without a signal
- Small number of predictors reduce costs and ease interpretation
- The PITE has a straightforward interpretation, however the induced split of the covariate space is less well interpretable as compared to subgroups defined in terms of individual covariates
- The non-linearity of the Cox model does not allow for exact solutions for selective inference
- In the end, demonstration of replicability of subgroup findings may be necessary to substantiate findings

References

[1] C Schmoor and M Schumacher.
Effects of covariate omission and categorization when analysing randomized trials with the Cox model.
Statistics in Medicine, 16:225–237, 1997.

[2] B Efron.
Estimation and accuracy after model selection.
Journal of the American Statistical Association, 109:991–1007, 2014.

[3] D Byar and S Green.
The choice of treatment for cancer patients based on covariate information: application to prostate cancer.
Bulletin du Cancer, 67:477–490, 1980.

[4] R J Tibshirani, J Taylor, R Lockhart, and R Tibshirani.
Exact post-selection inference for sequential regression procedures.
Journal of the American Statistical Association, 111:600–620, 2016.

[5] J Taylor and R Tibshirani.
Post-selection inference for ℓ_1 penalized likelihood models.
The Canadian Journal of Statistics, pages 1–21, 2017.