Exploring Complexity in Health: An Interdisciplinary Systems Approach
A. Hoerbst et al. (Eds.)

doi:10.3233/978-1-61499-678-1-745

Constructing an ROC Curve to Assess a Treatment-Predictive Continuous Biomarker

Alexander ENGELHARDT[a,1], Yu-Ming SHEN[a], and Ulrich MANSMANN[a,b]
[a] *IBE, University of Munich*
[b] *DKTK, German Consortium for translational cancer therapy*

Abstract. This paper presents the idea of an ROC curve, which quantifies the discriminatory potential of a continuous biomarker for treatment selection when the outcome is continuous. The analysis assumes data from a randomized parallel group design. We use non-parametric density estimators to construct an ROC curve based on the probabilities that a (non-)responder, defined by better (worse) response to treatment as opposed to control, in the treatment group has a biomarker value above a value c. Our non-parametric approach comes close to the true AUC in a simulation study based on a normal distribution. Application to a real data set shows that despite a significant interaction term in a proportional hazards model, a biomarker may not be helpful for treatment decisions. Our proof-of-principle study opens the door to further developments and generalizations.

Keywords. Treatment predictive biomarker, non-parametric density estimation, ROC curve

1. Introduction

Treatment biomarker interaction is an important concept to operationalize personalized medicine. From a statistical point of view, regression models for treatment outcome with an interaction term between treatment and continuous biomarker are widely used to detect treatment biomarker interaction. Royston et al. [1] provide an interesting example modelling the treatment-biomarker interaction term by a fractional polynomial.

But, a significant treatment-biomarker interaction term may be misleading with regard to its interpretation in terms of an individual treatment decision [2]. In this paper an alternative approach is provided which uses the ROC curve as an instrument to quantify how the distribution of a biomarker differs between responders and non-responders in a specific therapeutic setting. Our approach relies on data from a randomized parallel group design.

For this scenario, the paper presents the basic idea for the inference of the ROC curve using non-parametric conditional density estimates [3] and demonstrates its potential value based on a simulation study and an example, the reanalysis of the data presented in [1].

The strategy is developed for a continuous outcome. It can also be generalized to survival data when censoring is replaced by imputation. Huang et al. provide an alternative ROC estimate for a binary outcome in a counterfactual setting [2]. The

[1] Corresponding author: LMU, Marchioninistr. 15, 81377 München; E-mail: engelhdt@ibe.med.uni-muenchen.de.

paper modifies the counterfactual ROC approach in [2] to data from a parallel-group randomized trial.

2. Methods

The continuous outcome in both treatment groups is noted by Y1 and Y2, with higher values indicating better outcome. The continuous biomarker value is represented by X. The treatment effect between both treatments given a biomarker value x is noted by $\Delta x = Y1|x - Y2|x$. Conditioned on the value of the biomarker x response on treatment 1 is defined by by $\Delta x > 0$.

The ROC curve is defined by (F(c), G(c)), c ε R. $F(c) = P(X>c|\Delta<0)$ is the probability that a non-responder under therapy 1 has biomarker values above c and $F(c) = P(X>c|\Delta>0)$ is the probability that a responder under therapy 1 has biomarker values above c.

2.1. Bivariate normal (X, Yi) data model

An ideal randomized trial is simulated by assuming two independent bivariate normal distributed (X, Y1) and (X, Y2) measurements. The randomization is relevant to assure that the marginal distribution of X is equal in both treatment groups. The correlation between biomarker X and outcome Yi is different in both groups (ρ1=0.5, ρ2=-0.5).

The following specifications are made to fix the data generating mechanism in our ideal trial: $\mu_X = 10$, $\sigma_X^2 = 25$, $\mu_{Y1} = 15$, $\sigma_{Y1}^2 = 4$, $\mu_{Y2} = 15$, $\sigma_{Y2}^2 = 4$, $\sigma_{X,Y1} = 0.5 \cdot \sigma_{Y1}\sigma_X$, and $\sigma_{X,Y2} = -0.5 \cdot \sigma_{Y2}\sigma_X$. In order to visualize the distribution of the trial data: (1) a sample of N=10000 realizations for a biomarker X are drawn from a pre-specified normal distribution $N(\mu_x, \sigma_x^2)$; (2) for each realization X_i, a realization for each of $Y_{1|x}$ and $Y_{2|x}$ are sampled from the distribution $N(\mu_j + \frac{\sigma_{jX}}{\sigma_X^2}(x - \mu_x), \sigma_{Yj}^2 - \frac{\sigma_{jX}}{\sigma_x^2})$, where $\sigma_{jX} = \mathrm{Cov}(Y_j, X)$. The data for group 1 (2) is presented in the upper left panel of Figure 1 by red (black) points. With the paired samples for $Y_{1,i}$ and $Y_{2,i}$ conditional to the biomarker value X=x, we create a bivariate sample of (X,Δ) which is presented in Figure 1, upper right panel. The blue (green) points show the sample which is conditioned to $\Delta<0$ ($\Delta>0$).

2.2. Non-parametric density estimation and sampling approach

Given a data set for a randomized two group RCT, non-parametric kernel conditional density estimation [3] helps to estimate densities for Yj|X, regardless of the distribution of biomarkers or outcomes. Together with univariate kernel density estimates for X, we can simulate bivariate nonparametric samples from (X,Yj), and subsequently compute samples from (X, Δ). The rejection sampling approach is used to obtain realizations of Yj for each given X [4].

The ROC curve is constructed by defining the samples conditioned on ($\Delta>0$) and ($\Delta<0$). The corresponding densities are shown in the low left panel of Figure 1. Now it is straightforward to get estimates of F(c) and G(c), to determine the ROC curve (shown in the lower right panel of Figure 1), and to calculate its AUC. We used the R software [5] and its package ROCR [6] to compute the ROC curves and their

corresponding AUCs. The calculation of a 95% confidence interval for the AUC can be derived from a wild bootstrap on the data set under analysis. Similarly we can draw a 95% confidence region for a specific (FP,TP) point on the ROC curve.

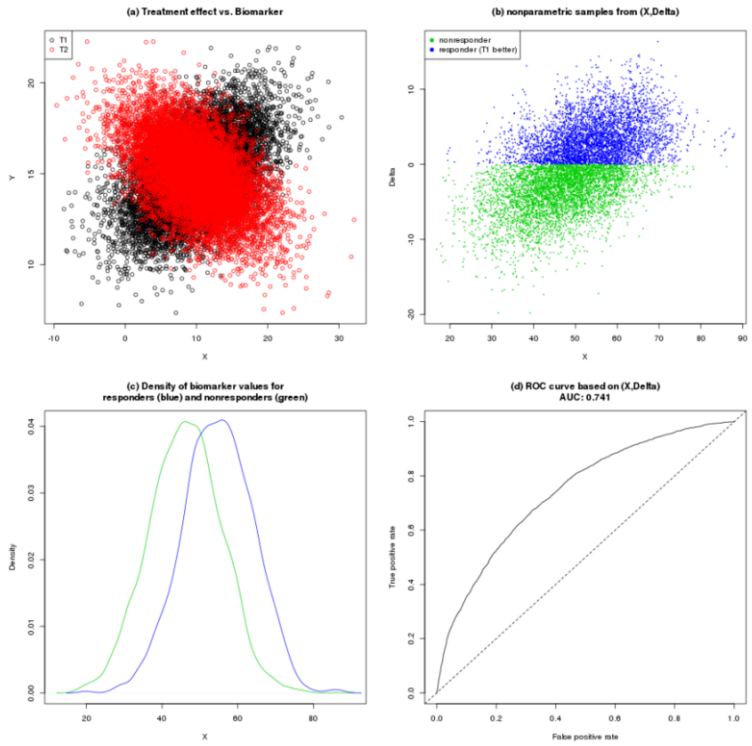

Figure 1. The simulation procedure for 10000 normal distributed samples of section 2.1.

2.3. Performing a simulation study to compare the non-parametric AUC estimate with the true AUC value derived from the data model

The binormal data model presented in section (2.1) can be used to calculate the true AUC value, this value can be compared to AUC estimates derived from trial data of different sizes.

3. Results

3.1. Simulation Study

The true AUC derived from the data mechanism described in section 2.1 is 0.7956. For each sample size in Table 1 a total of 1000 bootstrap simulations have been performed based on the scenario of section 2.1. The mean bias (true AUC minus bootstrap-AUC) seems to decrease with increasing sample size, while the empirical 95% region gets narrower. For small sample size, the bias appears to be substantial.

Table 1. Results of the simulation study

Sample size per group	100	200	500	1000
Mean bias	0.0709	0.0550	0.0500	0.0390
95% region for bias	[0.0004, 0.1482]	[0.0013, 0.1127]	[0.0156, 0.0991]	[-0.0010, 0.0813]

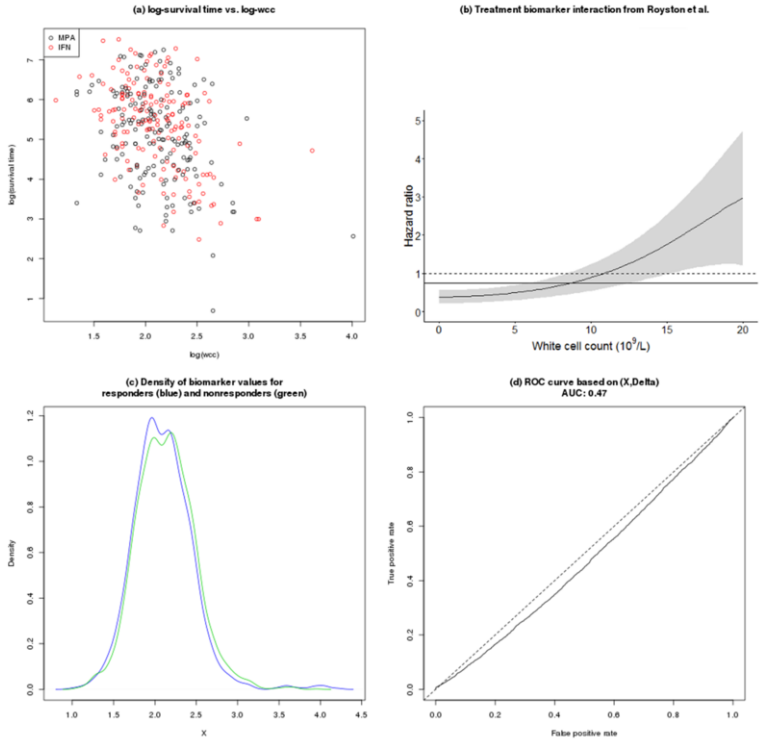

Figure 2. Simulation procedure and results for the data from Royston et al. [1]. Upper right panel: Treatment effect by white cell count with 95% pointwise confidence interval; solid line, overall treatment effect; dash line: no treatment effect level (HR=1).

3.2. Example - non-parametric ROC estimation

Data of 322 patients published by Royston et al. [1] is analyzed following section 2.2: estimation of the ROC curve as well as its AUC. The data is from a randomized breast cancer trial. The estimate of the AUC may be conservative regarding the results of section 3.1. The biomarker treatment interaction for white cell counts (wcc) is analysed. We log-transformed the survival times and wcc. (See Figure 2, left upper panel). Bootstrap sampling was applied to estimate a 95% CI for the AUC.

The (X,Δ) distribution was simulated by generating 10000 samples as described in section 2.2. The densities of the respective samples are shown in the lower left panel of Figure 2. The estimated ROC curve for the discrimination of treatment response by log-white cell count is presented in the right lower panel of Figure 2. The estimate of the AUC is 0.47 (95% bootstrap CI [0.4561, 0.4787]).

3.3. Example – regression based interaction estimate

Royston et al. [1] use fractional polynomials within a proportional hazards model to estimate the interaction between the white cell counts and the treatment. They found a significant interaction (p=0.0001) formalized by the fractional polynomial presented in the upper right panel of Figure 2. This result creates the expression that the biomarker has the potential to be helpful for treatment decisions. In section 3.2, the results seem not to be so promising and point out that the biomarker does not discriminate between responders and non-responders.

4. Discussion

We employed nonparametric conditional kernel density estimation to obtain samples from the distribution of treatment effect conditioned on a biomarker value and constructed the bivariate distributions of biomarker values and therapeutic response. From this we derive a ROC curve to quantify the discrimination of biomarker values between responder groups and the corresponding AUC.

We present a proof of principle which does not answer relevant technical questions related to the inference procedure. They will be discussed elsewhere. The simulation study shows that the inference technique needs improvement to be applied in small samples.

Our ROC differs from the classical ROC approach by having a random variable (Δ) which determines the groups to be compared. There is not a classical gold standard.

Generalization of the response definition introducing a third category (non-responder: $\Delta<-r$; indifference: $-r\leq\Delta\leq r$; responder: $\Delta>r$)) may also be of interest.

A similar approach for predictive values (i.e. $P(\Delta>0|X>c)$) will be developed.

References

[1] Royston, P., W. Sauerbrei, and A. Ritchie. "Is treatment with interferon-α effective in all patients with metastatic renal carcinoma? A new approach to the investigation of interactions." British journal of cancer 90.4: 794-799 (2004).

[2] Huang Y, Gilbert PB, Janes H. “Assessing Treatment-Selection Markers using a Potential Outcomes Framework” Biometrics 68: 687–696 (2012).

[3] Hall, Peter, Jeff Racine, and Qi Li. "Cross-validation and the estimation of conditional probability densities." Journal of the American Statistical Association 99(468): 1015-1026 (2011).

[4] Devroye, Luc. "Sample-based non-uniform random variate generation." Proceedings of the 18th conference on Winter simulation, 260-265. ACM (1986).

[5] R Core Team. “R: A language and environment for statistical computing”. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (2015).

[6] Sing T, Sander O, Beerenwinkel N and Lengauer T. “ROCR: visualizing classifier performance in R.” Bioinformatics, 21(20), pp. 78-81. URL: http://rocr.bioinf.mpi-sb.mpg.de (2005).