Comparison of Cutoff Determination Methods

A Biomarker Case Study

Anke Schulz, Bayer AG, Laura Schlieker, Staburo GmbH

Introduction

Biomarkers (BM) are often dichotomized using a cutoff to identify patient populations with a better prognosis or benefiting from a specific therapy.

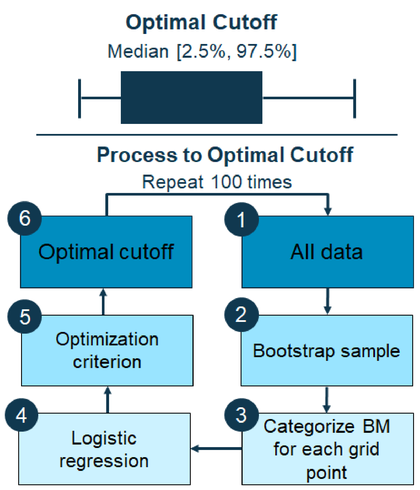
The optimal cutoff needs to be determined considering both context and objective of the study, as different cutoff determination strategies are tailored to different objectives.

We compared different cutoff determination methods for binary endpoints in a prognostic setting using a phase III oncology study as well as simulated data.

Methodology

The optimal cutoff and the corresponding confidence interval (CI) was calculated by employing a repeated bootstrap approach (see Figure 1) based on a logistic regression model:

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 \cdot I(BM > cut)$$



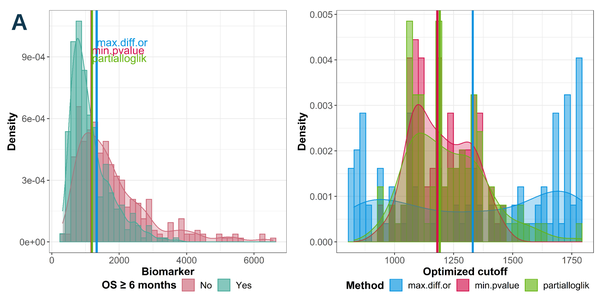
Bootstrap approach

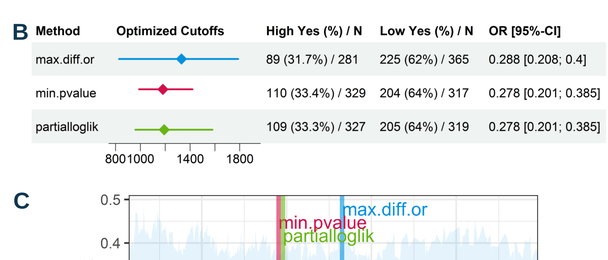
The table shows the cutoff optimization criteria:

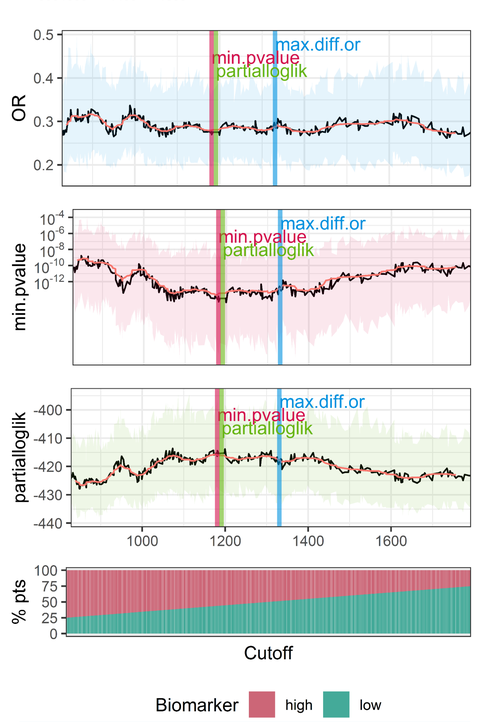
Method	Details
Partialloglik	Maximize log-likelihood: $\max_{cuts} \log\left(\prod_{i=1}^{n} p(x_i)^{y_i}\left(1-p(x_i)\right)^{1-y_i}\right)$
max.diff.or	Maximize absolute difference of OR_{BM} to 1: $\max_{cuts}\left\lvert \text{Odds}_{\text{BM } high}/\text{Odds}_{\text{BM } low} - 1\right\rvert$
min.pvalue	Minimize p-value of OR_{BM}: $\min_{cuts} p_{OR_{BM}}$

Real Data – Results

Clinical and biomarker data from a randomized phase III trial in colorectal cancer was investigated. The endpoint was overall survival (OS) ≥ 6 months (Yes vs. No).



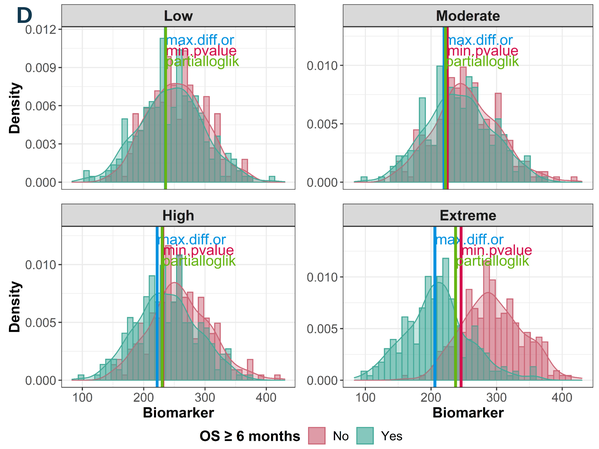


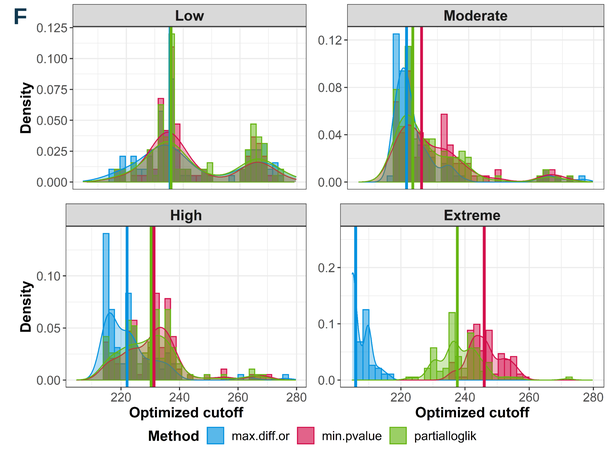


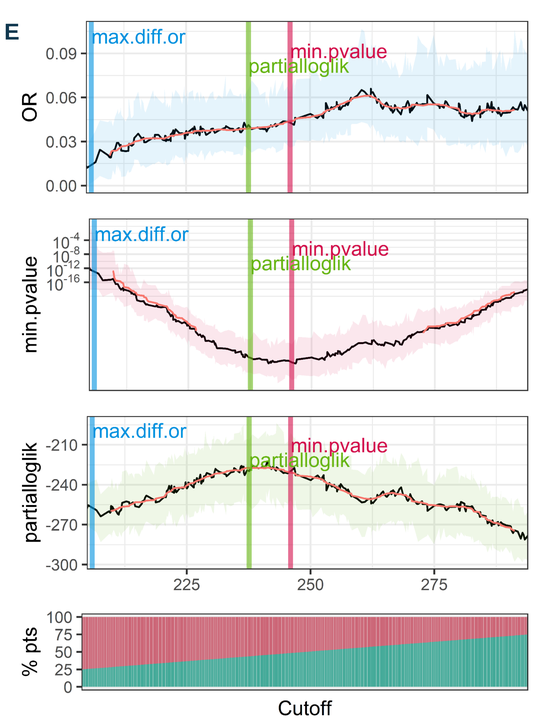
A: Histogram biomarker by endpoint
B: Optimized Cutoffs
C: Bootstrap results

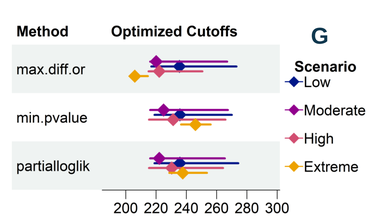
Simulated Data – Results

Artificial data was simulated for different prognostic scenarios. The biomarker was assumed to be normally distributed with standard deviation (SD) 50 and mean values of 250 ± 2 (Low), 250 ± 7 (Moderate), 250 ± 12 (High) and 250 ± 40 (Extreme). The simulated endpoint was OS ≥ 6 months.









D: Histogram by scenario
E: Histogram of optimized cutoffs from bootstrap
F: Bootstrap results for Extreme scenario
G: Forest plot of optimized cutoffs by scenario

Summary

In the case study we could show a strong prognostic effect for each method, i.e., Odds Ratio (OR) ~ 0.280. However, the CIs for the optimal cutoff are still covering a wide range of potential cutoffs.

In general, the cutoff depends on the optimization method and prognostic effect.

The results of the simulation study revealed similar optimized cutoffs and CIs for min.pvalue and partialloglik.

In case of an extreme prognostic effect all methods achieve small CIs for the optimal cutoff, but the point estimates differ (G).

Acknowledgements

We would like to thank Maria Blanco and Yiyao Chen from Staburo GmbH for the data simulation and for the validation of the analyses.