Time to Event Prediction Based on Risk Calculation from Longitudinal Biomarkers using Bayesian Hierarchical Changepoint Mixture Models

Lynette M. Smith[1], Morshed Alam[1], Sukhwinder Kaur[2]
[1]Department of Biostatistics, [2]Department of Biochemistry and Molecular Biology
University of Nebraska Medical Center, Omaha, NE, USA

Purpose/Aims

- ❑ Develop a Bayesian changepoint mixture model on longitudinally measured biomarkers for prediction of overall survival in early pancreatic cancer.
- ❑ The newly developed method will utilize serial marker measurements in combination with patient characteristics to improve biomarker performance in predicting pre- or post-treatment survival outcome.

Background

Pancreatic cancer (PC) is an extremely aggressive malignancy with a 5-year overall survival (OS) rate of <8%. Efforts have been made to develop methodologies for diagnostic tests, but there is also a need for biostatistical methods that utilize serially measured biomarkers to enable improved prediction of OS pre- and post-treatment following resection. Longitudinally measured biomarkers can show a variety of trends over time. The rate of change in biomarker levels in combination with the absolute expression level could better predict presence/progression of disease and OS. Currently, few methods incorporate a joint model for OS that could capture these changes in cancer patients. The growing interest in personalized medicine as well as tailored decision making on patient conditions has popularized dynamic prediction in statistical joint modeling with longitudinal and survival outcomes. Longitudinal biomarkers could be utilized in pancreatic cancer to predict overall survival for individuals to aid in decision making. We hypothesize that the use of longitudinal biomarkers in a changepoint mixture model will improve survival prediction in PC.

Analysis Methods

Mixture Model:
We will model the longitudinal biomarker data and OS in a Bayesian framework. $y_{ij}(t)$ is the biomarker data for patient i at time j.

Change point model: For cases that produce the biomarker, we will model the biomarkers as a piecewise linear function, with

$$E(Y_{ij}|t_{ij}, I_i = 1) = \theta_i + \gamma_i(t_{ij} - \tau_i),$$

where t_{ij} is the age of patient i when they enter the study at time $j = 0, \dots, k_i$. I_i is an unobserved indicator included to distinguish cases that produce extra levels of the biomarker to those that do not. θ_i is the mean level of the biomarker for patient i and γ_i is the slope of the biomarker, which begins at the unobserved change point τ_i.

Constant model: We assume the biomarker is relatively constant for cases that do not produce extra levels of the biomarker, shown in this model.

$$E(Y_{ij}|t_{ij}, I_i = 0) = \theta_i$$

We will investigate multiple priors for these parameters before making the final decision. Normal priors will be used for the biomarker distributions, as a starting point.

For event time data, the Cox proportional hazards model is assumed:

$$\lambda_0(t_i)\exp[\beta' x(t_i, \theta_i, \gamma_i, \tau_i)]$$

$\lambda_0(t_i)$ is the baseline hazard function, β is a vector of regression coefficients and x contains the biomarker information. We will apply Bayesian model fitting and testing using Markov chain Monte Carlo (MCMC). The method will be tested in the PC dataset as well as in a simulated dataset. Goodness of fit will be assessed by Q-Q plots as in Skates. We will also examine the sensitivity to prior distribution selection by choosing multiple priors and comparing the posterior summaries for the chosen priors.

Strategy

- Determine patient biomarker profiles (see figure 1).
  1. Flat or
  2. Increasing or decreasing slope with a changepoint.
- Combine patients into groups based on biomarker profile
  - These groups of profiles are the “mixture” part of the model.
- Fit a joint model of survival and the longitudinal biomarker profiles.

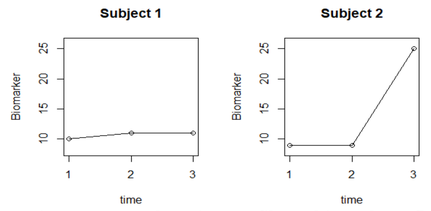


Figure 1. Example patient profiles, subject 1 has a flat profile, subject 2 has a changepoint at time 2.

Data source

- 43 early stage pancreatic cancer patients treated at UNMC from 2002-2011[1].
  - Demographic data: age, sex, race, alcohol smoking history.
  - Pancreatic cancer disease state: stage, grade, pre- and post-chemotherapy, pre- and post-surgery.
  - Overall survival information.
  - Longitudinally measured biomarkers, MUC5AC and CA19-9.

Preliminary Results

- Table 1 shows the patient characteristics. Most are stage II, male, show moderate differentiation, and have a family history of PC.
- Patient profiles are shown in Figure 2 for MUC5AC and CA19-9 on the log scale. CA19-9 shows mostly flat profiles whereas MUC5AC shows extreme changes in biomarker levels. They produce a challenge in that time points when biomarkers are collected are not uniform.
- Overall survival distribution of the patients. The median OS is 1.8 years in these early stage PC patients (figure 2).
- Figure 3 shows the correlation between the biomarkers and histograms of their distributions. MUC5AC and CA19-9 show a positive correlation.

Table 1. Patient characteristics (n=43)

		N(%)
Age, years	Median (range)	67 (46-83)
Gender	Male Female	28 (65%) 15 (35%)
Stage	I II	6 (14%) 37 (86%)
Grade	Moderate Poor Well Unknown	28 (65%) 9 (21%) 3 (7%) 3 (7%)
Smoking status	Current Former Never	15 (35%) 14 (33%) 14 (33%)
Alcohol use	Yes	24 (57%)
Family history of PC	Yes	33 (79%)
Diabetes	Yes	15 (35%)

Figure 2. Patient profiles of biomarkers over time, a) MUC5AC shows extreme changes in profiles and b) CA19-9 shows generally flatter profiles, c) overall survival distribution

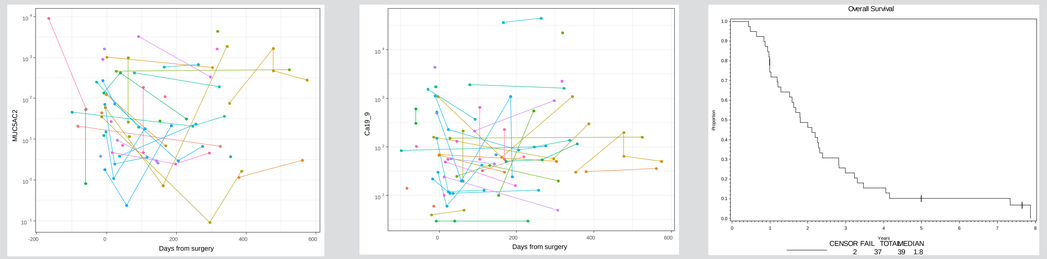
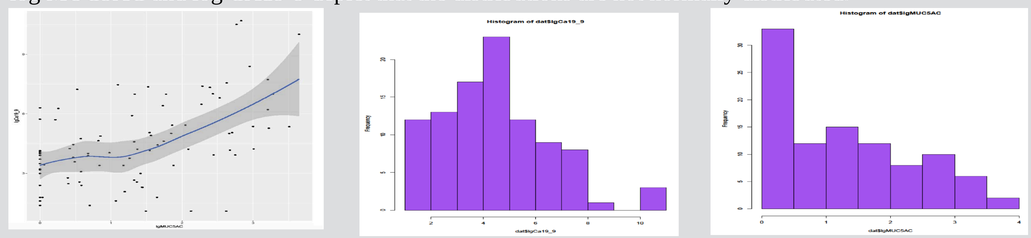


Figure 3. Log biomarkers (MUC5AC and CA19-9) are positively correlated (r=0.52). Histogram for both log MUC5AC and log CA19-9 depict that the distributions are not normally distributed.



Results cont.

Additionally, we fit a frailty model of OS looking at the longitudinally measured biomarkers, accounting for correlation within person. Here we used post surgical measurements only.

Table 2. Overall survival with two biomarker predictors

	HR	95% CI	P-value
ln(CA19-9)	1.27	1.06-1.51	0.0098
ln(MUC5AC)	0.95	0.90-1.0	0.061

Higher levels of CA19-9 following surgery are significantly associated increased risk of death. Lower levels of MUC5AC are marginally associated with increased risk of death.

Next steps

- Classify biomarker profiles into patterns.
- Fit the mixture model, determining how to implement the changepoints.
- Determine the fit of the model.
- Compare to a simpler model without the mixture component.

Acknowledgments

The project described was supported by the National Institute Of General Medical Sciences, 1U54GM115458. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Contact information

Lynette M. Smith, PhD
Assistant Professor, Department of Biostatistics
College of Public Health, University of Nebraska Medical Center
Phone: (402) 559-8114
email: lmsmith@unmc.edu

References

1. Kaur S, Smith LM, Patel A, Menning M, Watley DC, Malik SS, Krishn SR, Mallya K, Aithal A, Sasson AR, Johansson SL, Jain M, Singh S, Guha S, Are C, Raimondo M, Hollingsworth MA, Brand RE, Batra SK. A Combination of MUC5AC and CA19-9 Improves the Diagnosis of Pancreatic Cancer: A Multicenter Study. Am J Gastroenterol. 2017 Jan;112(1):172-183. doi: 10.1038/ajg.2016.482. PMID: 27845339
2. Skates SJ, Pauler DK, Jacobs IJ. Screening Based on the Risk of Cancer Calculation from Bayesian Hierarchical Changepoint and Mixture Models of Longitudinal Markers. Journal of the American Statistical Association. 2001 Jun;96:429-439
3. Pauler DK, Finkelstein DM. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. Statistics in Medicine. 2002 21:3897-3991.