Computational Statistics and Data Analysis 112 (2017) 170–185

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Bayesian variable selection for a semi-competing risks model with three hazard functions

Andrew G. Chapple [a,*], Marina Vannucci [a,b], Peter F. Thall [b], Steven Lin [c]

[a] Rice University, Department of Statistics, 6100 Main St., Duncan Hall 2124, Houston, TX 77005, USA

[b] Department of Biomathematics, Box 237, M.D. Anderson Cancer Center, University of Texas, 1515 Holcombe Boulevard, Houston, TX 77030, USA

[c] Department of Radiation Oncology, Unit 0097, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030, USA

ARTICLE INFO

Article history:
Received 19 May 2016
Received in revised form 25 February 2017
Accepted 1 March 2017
Available online 22 March 2017

Keywords:
Semi-competing risks
Variable selection
Metropolis–Hastings

ABSTRACT

A variable selection procedure is developed for a semi-competing risks regression model with three hazard functions that uses *spike-and-slab* priors and stochastic search variable selection algorithms for posterior inference. A rule is devised for choosing the threshold on the marginal posterior probability of variable inclusion based on the Deviance Information Criterion (DIC) that is examined in a simulation study. The method is applied to data from esophageal cancer patients from the MD Anderson Cancer Center, Houston, TX, where the most important covariates are selected in each of the hazards of effusion, death before effusion, and death after effusion. The DIC procedure that is proposed leads to similar selected models regardless of the choices of some of the hyperparameters. The application results show that patients with intensity-modulated radiation therapy have significantly reduced risks of pericardial effusion, pleural effusion, and death before either effusion type.

1. Introduction

Global cancer incidence estimates from 2008 indicate that esophageal cancer is the eighth most common and the sixth most deadly among cancers (Ferlay et al., 2010). Torre et al. (2015) estimated that there were 455,800 new cases and 400,200 deaths in 2012. The two most common types of esophageal cancer are squamous cell carcinoma and adenocarcinoma, the latter of which has been linked to obesity and gastrointestinal problems. Definitive concurrent chemoradiotherapy (CRT) is the standard treatment for esophageal cancers for patients with inoperable tumors. Several different methods for delivering radiation are used, particularly three dimensional conformal radiation therapy (3D-CRT). All of these methods increase patient survival but also have several adverse effects, the most common being pleural effusion (PE) and pericardial effusion (PCE) (Ishikura et al., 2003; Wei et al., 2009). Pleural and pericardial effusion occur when excess fluid is present around the lungs and heart, respectively, and can lead to poor function of these organs and death. These adverse events are associated with higher doses of radiation to the heart and lungs (Wei et al., 2009).

Intensity-modulated radiation therapy (IMRT) has been shown to reduce the volume of a patient's non-cancerous organs exposed to radiation and increase volume of radiation on esophageal tumors compared to 3D-CRT (Fenkella et al., 2008). Chandra et al. (2005) showed that IMRT reduced the volume of lungs that received different radiation doses compared to

* Corresponding author. Fax: +713 348 5476.
E-mail address: agc6@rice.edu (A.G. Chapple).

http://dx.doi.org/10.1016/j.csda.2017.03.002
0167-9473/

3D-CRT. Due to the relationship between increased dosage and effusion rates, IMRT potentially could result in fewer incidences of pleural and pericardial effusion in esophageal cancer patients compared to standard 3D-CRT treatment. Assessing the impact of IMRT on time to effusion is more complicated than assessing the impact of IMRT on overall survival time, even when pleural and pericardial effusion are considered separately. When the survival time of interest is a non-terminal event such as effusion, death is commonly assumed to be a non-informative independent censoring event (Shirai et al., 2011). This assumption is invalid because both pleural and pericardial effusion could lead to death, which means that death may indicate that a patient experienced effusion, which is informative censoring. Another complication with this data structure is that patients can experience effusion followed by death, but not death first and effusion afterwards. Administrative right censoring could occur before a patient experiences either event type or after a patient has effusion. Due to these complications, this data structure must be analyzed with a semi-competing risks model. This model has different hazards for three events: a given non-terminal event, death before the non-terminal event and death after the non-terminal event.

Lee et al. (2015) developed a novel Bayesian semi-parametric semi-competing risks regression model for a non-terminal event and death. Their motivating non-terminal event of interest was hospital readmission for patients diagnosed with advanced pancreatic cancer. Since pancreatic cancer has high mortality rates, they were concerned with end of life care and keeping patients comfortable at home during their final days. They considered three different hazard functions: the hazard of a non-terminal event, the hazard of death without a non-terminal event, and the hazard of death after a non-terminal event. Each of these three hazard functions resembled a Cox-type regression including a baseline hazard function which was assumed to be piecewise exponential, individual patient frailty parameters, and a linear combination of patient covariates. They used the posterior sample of the beta coefficients in the three hazards for inference on what types of homecare affected the hazard that a patient would return to the hospital, the hazard of death before returning to the hospital, and the hazard of death after patients were readmitted to the hospital. They implemented their algorithm in the package *SemiCompRisks* (Lee and Haneuse, 2015).

We initially aimed at implementing the semi-competing risks model of Lee et al. (2015) to analyze the effects of IMRT on effusion and overall survival for an observational study consisting of 470 patients at The University of Texas M.D. Anderson Cancer Center in Houston, TX, treated between January 1998 and April 2012 (He et al., 2016). However, it was unclear which baseline covariates should be included in the model for analyzing the IMRT effect, particularly because of the correlation between treatment group assignment and the baseline covariates, which could affect clinical conclusions. Consequently, in this paper we develop a variable selection procedure for the semi-competing risks model of Lee et al. that uses *spike-and-slab* priors and stochastic search variable selection (SSVS) algorithms for posterior inference. The proposed procedure performs variable selection for each of the linear terms in the three hazard functions. Furthermore, we devise a protocol to choose the threshold on the marginal posterior probability of variable inclusion based on the Deviance Information Criterion (DIC). The code for the described methodology can be found in the R package SCRSELECT (Chapple, 2016). In the application to the data from the esophageal cancer patients we do not perform variable selection on the IMRT status. To correct for some of the bias introduced in a nonrandomized observational study, we estimate the probability of receiving IMRT for each patient as a function of other covariates and include this propensity score in each hazard function. This allows us to compare the effects of IMRT on effusion and death while correcting for bias for non-randomization in every potential model. We present results from analyses done separately for pleural and pericardial effusion, where we show how the DIC procedure we propose leads to similar selected models, regardless of the choice of some of the hyperparameters. We find that patients with IMRT radiation have significantly reduced risks of pericardial effusion, pleural effusion and death before either effusion type. The rest of the paper is organized as follows: In Section 2, we describe the Bayesian semi-parametric semi-competing risks model, the variable selection priors and the Markov Chain Monte Carlo procedure for posterior inference. We also present the DIC-based procedure that we propose for the final covariate selection. In Section 3, we perform a simulation study to assess our proposed DIC-based procedure. In Section 4, we describe the case study data and discuss results and sensitivity to hyperparameter choices. Section 5 concludes the paper with a discussion.

2. Methods

2.1. Semi-parametric semi-competing risks model

Let T_{1i} denote the time to a non-terminal event and T_{2i} be the time to death for patient i. Lee et al. (2015) model covariate effects in the three hazard functions in the following manner. They denote h_1, the hazard of a non-terminal event, h_2 the hazard of a terminal event when the non-terminal event has not occurred and h_3, the hazard of a terminal event after the non-terminal event has occurred. Let x_i denote the vector of patient covariates and $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, $\boldsymbol{\beta_3}$ denote the three coefficient vectors associated with x_i in hazards 1, 2, and 3, respectively. They list the functional forms of the three hazards for the semi-Markov model as

$$h_1(T_{1i}|\gamma_i, \boldsymbol{\beta_1}, x_i) = \gamma_i h_{01}(T_{1i}) \exp(x_i^t \boldsymbol{\beta_1}), \tag{1}$$

$$h_2(T_{2i}|\gamma_i, \boldsymbol{\beta_2}, x_i) = \gamma_i h_{02}(T_{2i}) \exp(x_i^t \boldsymbol{\beta_2}) \tag{2}$$

and

$$h_3(T_{2i}|T_{1i}, \gamma_i, \boldsymbol{\beta_3}, x_i) = \gamma_i h_{03}(T_{2i} - T_{1i}) \exp(x_i^t \boldsymbol{\beta_3}). \tag{3}$$