Computational Statistics and Data Analysis 104 (2016) 197–208

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Cause-specific hazard regression for competing risks data under interval censoring and left truncation

Chenxi Li

Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Article history:
Received 24 September 2015
Received in revised form 21 April 2016
Accepted 7 July 2016
Available online 14 July 2016

Keywords:
Competing risks
Cause-specific hazard
Interval censoring
Left truncation
Penalized likelihood
Smoothing parameter selection

ABSTRACT

Inference for cause-specific hazards from competing risks data under interval censoring and possible left truncation has been understudied. Aiming at this target, a penalized likelihood approach for a Cox-type proportional cause-specific hazards model is developed, and the associated asymptotic theory is discussed. Monte Carlo simulations show that the approach performs very well for moderate sample sizes. An application to a longitudinal study of dementia illustrates the practical utility of the method. In the application, the age-specific hazards of AD, other dementia and death without dementia are estimated, and risk factors of all competing risks are studied.

1. Introduction

Interval censored failure time data arise widely from longitudinal studies where the occurrence of the failure-defining event can only be detected at periodic study visits. In many such kind of longitudinal studies, multiple types of events could occur to a subject. If those types of events are dependent but preclude each other or only time to the first event is of interest, competing risks issue comes up for the time-to-event analysis. Furthermore, if a subject's follow-up starts later than the time origin of the time-to-event analysis, there is also left truncation that needs to be accounted for. Clinical studies of elder people often give rise to left truncated and interval censored competing risks data. For instance, when studying age to onset of a chronic disease like diabetes, osteoporosis and Alzheimer's disease (AD), the event time is interval censored between two consecutive assessments, death is a competing risk that precludes those clinical endpoints if occurring prior to them, and enrolled study participants are usually required to be free of the disease and of course alive at the entry to follow-up. A real example of such studies is the Memory and Aging Project (MAP) (Bennett et al., 2012). Since 1997, the MAP has recruited more than 1400 older individuals from about 40 retirement communities and senior housing facilities in the Chicago metropolitan area to study how dementia evolves in the elderly. The participants were all dementia-free when they entered the study and had yearly evaluation for dementia during the follow-up, which leads to left-truncated and interval censored age-to-dementia data. Dementia could be categorized into two major types, AD and other dementia, which are competing risks because time to the first incidence of dementia is of scientific interest. Besides, a considerable proportion of these elder participants passed away before they were diagnosed to be demented, making death another competing risk.

Competing risks data under interval censoring and left truncation was first studied by Hudgens et al. (2001), who developed the nonparametric maximum likelihood estimator (NPMLE) for the cumulative incidence function of a failure type. Since then, a great deal of attention has been drawn to the inference of the cumulative incidence function with interval

E-mail address: cli@epi.msu.edu.

http://dx.doi.org/10.1016/j.csda.2016.07.003

censored competing risks data. Relevant works include Jewell et al. (2003), Groeneboom et al. (2008a,b), Li and Fine (2013), Hudgens et al. (2014) and Li (2016) among others. But the important problem of inference of the cause-specific hazard with interval censored competing risks data has not been studied extensively. Along this line, Li and Fine (2013) proposed four cause-specific hazard estimators for current status competing risks data based on smoothing the nonparametric cumulative incidence estimators, and derived their asymptotic distributions. The estimation methods of Li and Fine (2013) can be naturally extended to the setting of mixed case interval censoring. With a slightly different target, Frydman and Liu (2013) developed the NPMLE of the cumulative cause-specific hazard function in an interval censored competing risks model. However, works on inference for cause-specific hazards with interval censored competing risks data, especially regression modeling, are still much needed.

We are aware that several works on inference for multistate models from interval censored data have been published in the past 17 years. Two major papers are Joly and Commenges (1999), which studied a progressive three-state model, and Joly et al. (2002), which studied an illness-death model. The censoring mechanisms for the event history data considered therein are different from the one we consider, though. Specifically, Joly and Commenges (1999) assume that the second transition is only subject to right censoring; Joly et al. (2002) assume that the death time is exactly known if a subject dies and that a subject could be ill at death even if s/he is healthy at the last seen time; we focus on the situation where there are several interval censored competing events but any two of them cannot occur within the same interval. Ideally, the penalized likelihood approach in Joly and Commenges (1999) and Joly et al. (2002) can be adopted to analyze competing risks data under interval censoring and left truncation, yet some gaps in practice and theory remain to be filled. In terms of practice, how to minimize the cross-validation criterion to select smoothing parameters was not discussed by either paper, neither was how to determine the number of knots as well as the boundary endpoints for spline smoothing. In terms of theory, there was no theoretical justification for using cubic splines to approximate the intensity functions in the particular penalized likelihoods proposed by the two papers, and neither paper investigated the finite sample performance of the proposed variance estimator and confidence interval for transition intensities. In this article, we present a slightly different penalized likelihood approach to analyze competing risks data subject to interval censoring and left truncation and address the aforementioned issues. Additionally, we provide some heuristic arguments about the asymptotic theory associated with the methods.

The rest of the paper is organized as follows. Section 2 presents the estimation and inferential procedures for cause-specific hazards with competing risks data under interval censoring and possible left truncation when covariates are available. The associated asymptotic theory is also discussed therein. In Section 3, we conduct extensive simulations to investigate the finite sample performance of the proposed methods. An analysis of the MAP data using the proposed methods is given in Section 4, followed by some concluding discussions in Section 5 that point out several future research directions. The computational details are collected in the Appendix.

2. Cause-specific hazard regression

2.1. Observations

We describe the competing risks data under interval censoring and possible left truncation as follows. Let $(V_1^{(i)}, V_2^{(i)}, \ldots, V_{M_i}^{(i)})$ be the sequence of inspection times for subject $i (i = 1, \ldots, n)$ and $V_0^{(i)}$ be the left truncation time, if any, and 0 otherwise. Define $\vec{V}_i = (V_0^{(i)}, V_1^{(i)}, \ldots, V_{M_i}^{(i)})$. T_i denotes the failure time, K_i denotes the failure cause, and J denotes the number of failure causes. Define $\Delta_{jk}^{(i)} = I(V_{j-1}^{(i)} < T_i \leq V_j^{(i)}, K_i = k)$, $\Delta_k^{(i)} = (\Delta_{1k}^{(i)}, \ldots, \Delta_{M_i k}^{(i)})$ and $\vec{\Delta}_i = (\Delta_1^{(i)}, \ldots, \Delta_J^{(i)}) (i = 1, \ldots, n,\ j = 1, \ldots, M_i,\ k = 1, \ldots, J)$. $\mathbf{Z}_i$ denotes a d-dimensional vector of time-independent covariates whose effects on the distribution of (T_i, K_i) are of interest. The observable competing risks data under interval censoring and possible left truncation consist of n i.i.d. vectors of $(M_i, \vec{V}_i, \vec{\Delta}_i, \mathbf{Z}_i)$.

2.2. Model and likelihood

We are interested in estimating the conditional cause-specific hazard functions given the covariates, $\lambda_k(t|\mathbf{Z})$ $(k = 1, \ldots, J)$. For this purpose, we assume a proportional cause-specific hazards model:

$$\lambda_k(t|\mathbf{Z}) = \lambda_{0k}(t) \exp(\mathbf{Z}^T \boldsymbol{\beta}_k), \quad k = 1, \ldots, J, \tag{1}$$

where $\lambda_{0k}(t)$'s are baseline cause-specific hazards and $\boldsymbol{\beta}_k$'s are cause-specific regression parameters.

The development of the likelihood considers two kinds of interval censoring schemes. The first one is a generalization of mixed case interval censoring (Schick and Yu, 2000) to the setting with covariates, which assumes that the inspection process is independent of the failure time and cause given the covariates, that is, for any subject i,

$$(M_i, \vec{V}_i) \perp (T_i, K_i)|\mathbf{Z}_i. \tag{2}$$

The second censoring scheme is a generalization of the independent inspection process (IIP) model (Lawless, 2003, Section 2.3.1), for which the inspection process stops if any type of failure is detected, to the setting with covariates. Under this