Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 129 (2005) 145–157
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Estimating equation approach for regression analysis of failure time data in the presence of interval-censoring

Xingqiu Zhao[a], Hee-jeong Lim[b], Jianguo Sun[c,*]

[a] *Department of Mathematical and Statistical Sciences, 632 Central Academic Building, University of Alberta, Edmonton, Alberta, T6G 2G1 Canada*

[b] *Department of Mathematics and Computer Science, Northern Kentucky University, Highland Heights, KY 41099, USA*

[c] *Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211, USA*

Available online 25 August 2004

Abstract

This article discusses statistical inference for the proportional hazards model when there exists interval-censoring on both survival time of interest and covariates (J. Roy. Statist. Soc. B 34 (1972) 187; Encyclopedia of Biostatistics. Wiley, New York, 1998, pp. 2090–2095). In particular, we consider situations where observations on the survival time are doubly censored and observations on covariates are interval-censored. For inference about regression parameters, a general estimating equation approach is proposed. The proposed estimate of the parameter is a generalization of the maximum partial-likelihood estimate for right-censored failure time data with known or exactly observed covariates (The Statistical Analysis of Failure Time Data. Wiley, New York, 1980). The asymptotic properties of the proposed estimate are established and its finite sample properties are investigated through a simulation study.

Keywords: Estimating equations; Interval-censoring; Proportional hazards models; Regression analysis

* Corresponding author. Tel.: +1-573-882-6376; fax: +1-573-884-5524.
E-mail address: sunj@missouri.edu (J. Sun).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.06.044

1. Introduction

The proportional hazards model is the most commonly used regression model in survival analysis and is defined as

$$\lambda(t) = Y(t)\lambda_0(t)\ \exp(\beta' Z) \tag{1}$$

given a vector of covariates Z (Cox, 1972, Andersen and Gill, 1982). In the above, $\lambda_0(t)$ is an unknown baseline hazard function, β denotes the vector of regression coefficients, and $Y(t)$ is a predictable process taking value 0 or 1 indicating (by the value 1) if a subject is under observation at time t. When right-censored failure time data are available, many authors have studied the inference problem about β (Kalbfleisch and Prentice, 1980). In this paper, we consider a more general situation where the survival time of interest is doubly censored (De Gruttola and Lagakos, 1989) and covariates are interval-censored (Sun, 1998).

By doubly censored survival time, we mean that the survival time of interest is defined as the elapsed time between two related events, originating and end events. Furthermore, observations on the occurrences of the two events are interval- and right-censored, respectively. By interval-censoring, we mean that the occurrence time of the originating event is observed only to belong to an interval. Note that if the occurrence of the originating event is observed exactly, we would have right-censored observations on the survival time. By interval-censored covariates, we mean that the covariates are scalar variables or times to certain events and their values are known or observed only to belong to some intervals instead of being exactly known. Our goal is to make inference about regression parameters β.

One field in which doubly censored failure time data often occur is epidemiological studies, where the originating and end events may represent infection and onset of certain diseases, respectively. In particular, many authors have discussed such data in the context of AIDS studies (De Gruttola and Lagakos, 1989, Kim et al., 1993, Sun et al., 1999). In this case, the two events correspond, respectively, to HIV infection and AIDS diagnosis. The survival time of interest, the time from HIV infection to the diagnosis of AIDS, is often referred to as AIDS incubation time and plays an important role in the study of AIDS epidemic. HIV infection time is usually interval-censored in these studies because HIV status can only be checked periodically. In the meantime, AIDS diagnosis times are commonly right-censored due to patient drop-out from the study or the end of the study.

Goggins et al. (1999) discussed an example about interval-censored covariates arising from an AIDS clinical trial. In the example, the problem of interest is to predict the onset of active cytomegalovirus (CMV) using CMV shedding assuming that they are related by the proportional hazards model. However, the exact time to CMV shedding is usually unobservable since its determination is through clinical screen of blood or urine, which can only be performed periodically. In other words, only interval-censored CMV shedding times are available. Gómez et al (2000) described a similar example also from an AIDS clinical trial.

Several methods have been proposed for inference about β when interval-censoring occurs only on either survival time of interest (Kim et al., 1993, Sun et al., 1999) or covariates (Goggins et al., 1999). One shortcoming of these methods is that their asymptotic properties are unknown. There seems no existing method in the literature for the situation considered