Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Semiparametric transformation joint models for longitudinal covariates and interval-censored failure time

Chyong-Mei Chen [a,*], Pao-sheng Shen [b], Yi-Kuan Tseng [c]

[a] Institute of Public Health, School of Medicine, National Yang-Ming University, Taiwan
[b] Department of Statistics, Tunghai University, Taiwan
[c] Graduate Institute of Statistics, National Central University, Taiwan

ARTICLE INFO

Article history:
Received 29 August 2017
Received in revised form 21 March 2018
Accepted 5 July 2018
Available online xxxx

Keywords:
Interval-censored data
Joint modelling
Longitudinal data
Self-consistency equation
Transformation models

ABSTRACT

In many clinical trials and epidemiology research, subjects are followed-up repeatedly, and repeated measurements on longitudinal covariates as well as an observation on a possibly censored time-to-event are collected on each subject. The longitudinal covariates are often measured intermittently with measurement errors, and the measurement process is terminated by a correlated event process, leading to informative missing. Methods for joint modelling of longitudinal and time to-event data have received much attention in the statistics literature in recent years. Most research has focused on right-censoring mechanism for the event time. In practice, the event time is often examined at the pre-scheduled times, at which the longitudinal covariates are also measured, resulting in interval-censored survival data. To take the interval censoring into account and to provide a more general framework for studying the effects of covariates on survival time, a new class of joint models is proposed. The joint model comprises a linear mixed-effects model for the longitudinal biomarkers and a class of semiparametric transformation models for the failure time, which incorporates the underlying longitudinal biomarkers as time-dependent covariates. The likelihood approach and an EM algorithm for obtaining the semiparametric maximum likelihood estimator (SPMLE) are developed. In M-step, a hybrid algorithm combining the Newton–Raphson and self-consistency algorithms is used to compute the finite-dimensional and infinite-dimensional parameters. The existence and consistency of the SPMLE are established. The proposed method is investigated through simulation studies and illustrated using a real dataset from a Taiwanese HIV/AIDS cohort study.

1. Introduction

In many survival data analyses, subjects are often followed-up at discrete and intermittent time points, and longitudinal responses measured at these observation times are usually treated as important indicators of the event time of interest. The interest of longitudinal studies mainly lies on inferring the event time and assessing the association between the event time and longitudinal covariates. However, the longitudinal covariates are often measured with error since many measures cannot be accurately ascertained, e.g., the measurement values of serum bilirubin for patients with primary biliary cirrhosis (Ding and Wong, 2008) and CD4 counts for HIV-infected patients (Wulfsohn and Tsiatis, 1997). Moreover, the measurement process is often terminated by a correlated event process, leading to informative missing/dropout. It is well known that

* Corresponding author.
E-mail address: cmchen2@ym.edu.tw (C.-M. Chen).

https://doi.org/10.1016/j.csda.2018.07.001
0167-9473/

ignoring the measurement error would dilute the effect of longitudinal covariate on the event time (Prentice, 1982; Wulfsohn and Tsiatis, 1997). Separate analysis for event time and longitudinal covariates has been studied extensively. The Cox proportional hazard model is commonly used in time to event data analysis while mixed-effects models and the GEE method are widely used for longitudinal measurements. However, modelling the event time and longitudinal covariates separately can lead to biased effect estimates if the two processes are correlated, which is often the case (Hsieh et al., 2006). Specifically, the procedure of estimating the effects of covariates on the longitudinal data alone can be complicated by the disease-related dropout process or death, resulting in informative dropout. Such non-ignorable missing data can lead to biased inferences if a separate analysis is performed on the longitudinal data using the mixed-effects model or the GEE method. The approaches that employ the observed covariates as time-varying covariates in event-time model are also infeasible since longitudinal covariates are often observed intermittently such that we have no knowledge of the entire history of longitudinal process and the observed longitudinal covariates usually contain errors of measurement. Hence methods for joint modelling of longitudinal biomarkers and event time have received much attention in the statistics literature in recent years.

The motivation of research in this field has mainly stemmed from three objectives: (1) studying the relationship between longitudinal biomarkers and event time and assessing influences of covariates on both outcomes (Henderson et al., 2000; Zeng and Cai, 2005a); (2) improving inference for the longitudinal biomarkers subject to an informative dropout mechanism (Wu and Carroll, 1988; Schluchter, 1992; Little, 1995; Dupuy and Mesbah, 2002) (3) performing inference for the event time while taking the longitudinal biomarkers into account (Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997; Wang and Taylor, 2001; Brown and Ibrahim, 2003). One commonly used joint modelling is to apply a linear or non-linear mixed-effects model for longitudinal measurements and a semiparametric or parametric survival model for event time data, where a set of random effects is assumed to induce their interdependence. Further details of joint modelling of longitudinal and survival data can be found in Henderson et al. (2000) and Tsiatis and Davidian (2004). For the third objective, one commonly used joint modelling is to apply mixed-effects models to account for the heterogeneity and measurement errors of longitudinal measurements among individuals and then apply the survival model with longitudinal data as covariates to evaluate the association between longitudinal biomarker and event time (Wulfsohn and Tsiatis, 1997; Tseng et al., 2005; Ding and Wong, 2008; Tseng et al., 2015). To this end, frequentist approaches have been developed for statistical inference, including two-stage approach (Tsiatis et al., 1995; Dafni and Tsiatis, 1998) and likelihood approach (Wulfsohn and Tsiatis, 1997; Tseng et al., 2005), with the latter being generally more robust and producing less bias than the former (Hsieh et al., 2006). In particular, Wulfsohn and Tsiatis (1997) assumed a random components model with normal errors for the CD4 counts and the Cox (1972) proportional hazards (PH) model for AIDS-free survival time. Tseng et al. (2005) considered the joint modelling approach under the accelerated failure time assumption when covariates follow a linear mixed-effects model with measurement errors. Both articles develop likelihood approach and apply the expectation–maximization (EM) algorithm (Dempster et al., 1977) by treating the random effects as missing data to obtain the semiparametric maximum likelihood estimator (SPMLE). Brown and Ibrahim (2003) proposed a Bayesian approach that relaxes the distributional assumptions for the longitudinal model using Dirichlet process priors on the model parameters.

Most research has concentrated on right-censoring mechanism for the event time. In practice, the status of event of interest may not be followed continuously. Instead, the event status is examined at the pre-scheduled times, at which the longitudinal covariates are also measured, such that the event time of interest is only known to be bracketed between two adjacent examination times, resulting in interval-censored data. Furthermore, a longitudinal study is usually stopped after the occurrence of the failure is ascertained at a certain examination time. For example, in Taiwanese HIV/AIDS cohort study, some HIV-infected patients received highly active antiretroviral therapy, i.e., cocktail therapy, and some received standard treatment, which consisted of one or two anti-HIV drugs. In this longitudinal study, each patient returned periodically to hospital for the measurement of the CD4 T-cell counts and examination of the development of AIDS. Since the status of AIDS was examined at the pre-scheduled visit times, the AIDS onset time was subject to interval censoring. Since changes in patients' treatments are needed after AIDS onset, the longitudinal study of CD4 counts during the stage of HIV infection stopped once AIDS onset was ascertained. This leads to informative censoring of longitudinal CD4 counts.

Compared to the right-censored data, research is much scarcer on joint modelling of longitudinal covariates and interval-censored data. To assess the effect of dropouts for different reasons on inferences, Gueorguieva et al. (2012) constructed a parametric joint model for longitudinal measurements and cause-specific dropouts that allows for interval-censored dropout times. Rouanet et al. (2016) proposed a joint latent class model combining a mixed-model and an illness-death model that can handle both interval censoring and semi-competing risks data.

In this article, we consider joint modelling of longitudinal covariates and the interval-censored event time. For the longitudinal covariates, we consider a linear mixed-effects model with measurement errors, in which random effects are used to characterize the heterogeneity in longitudinal measurements among individuals. For the event time, we consider a class of semiparametric transformation models, which incorporates the underlying longitudinal covariates as time-dependent covariates. Since semiparametric transformation model includes the Cox PH model and proportional odds (PO) model as special cases, our proposed model is an extension of the joint model of Wulfsohn and Tsiatis (1997), where the Cox model is used for right censored survival time. It is well known that the interval-censoring mechanism induces challenging problems in estimation and large sample studies and then requires sophisticated optimization and theoretical justification techniques. Due to the existence of random effects, we adopt an EM approach to obtain the SPMLE. To tackle the estimation challenge due to the unknown function, we propose a hybrid algorithm consisting of the Newton–Raphson and self-consistency algorithms, which respectively computes the finite-dimensional and infinite-dimensional parameters

Please cite this article in press as: Chen C.-M., et al., Semiparametric transformation joint models for longitudinal covariates and interval-censored failure time. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.07.001.