Journal of Multivariate Analysis 131 (2014) 65–78

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Varying coefficient subdistribution regression for left-truncated semi-competing risks data

Ruosha Li [a], Limin Peng [b,*]

[a] Department of Biostatistics, University of Pittsburgh, 130 DeSoto Street, Pittsburgh, PA 15261, USA

[b] Department of Biostatistics and Bioinformatics, Emory University, 1518 Clifton Road NE., Atlanta, GA 30322, USA

ARTICLE INFO

Article history:
Received 8 April 2013
Available online 17 June 2014

AMS subject classifications:
62N01
62N02
62P10

Keywords:
Cumulative incidence
Left truncation
Hypothesis testing
Observational studies
Registry data analysis
Time-varying coefficient

ABSTRACT

Semi-competing risks data frequently arise in biomedical studies when time to a disease landmark event is subject to dependent censoring by death, the observation of which however is not precluded by the occurrence of the landmark event. In observational studies, the analysis of such data can be further complicated by left truncation. In this work, we study a varying coefficient subdistribution regression model for left-truncated semi-competing risks data. Our method appropriately accounts for the specifical truncation and censoring features of the data, and moreover has the flexibility to accommodate potentially varying covariate effects. The proposed method can be easily implemented and the resulting estimators are shown to have nice asymptotic properties. We also present inference, such as Kolmogorov–Smirnov type and Cramér–Von-Mises type hypothesis testing procedures for the covariate effects. Simulation studies and an application to the Denmark diabetes registry demonstrate good finite-sample performance and practical utility of the proposed method.

1. Introduction

In biomedical studies, semi-competing risks data [7] arise, for example, when time to a nonterminating landmark event of disease may be dependently censored by time to death but not vice versa. The data structure has attracted much research interest, due to its frequent occurrence in practical situations ([11,20,4]; among others). One example of the semi-competing risks data is the Denmark diabetes registry study [1,14], a prospective cohort study that included 2727 type I diabetic patients referred to the Steno Memorial Hospital in Greater Copenhagen between 1931 and 1988. An intermediate endpoint of interest is the development of diabetic nephropathy (DN), a syndrome indicating kidney failure. Since death may preclude the observation of DN but remains observable after DN onset, time to DN, say T_1, and time to death, say T_2, with diabetes diagnosis set as time origin, form a semi-competing risks structure. An important complication in this registry study, like in many other prevalence studies, is left truncation, which occurred because patients can be observed only if they were alive at the study enrollment.

With semi-competing risks data, the cumulative incidence function (i.e., subdistribution function) for T_1, $F_1(t) = \Pr(T_1 \le t, T_1 < T_2)$, has been advocated to characterize the distribution of the nonterminating event time, T_1. This crude quantity

* Corresponding author.
E-mail address: lpeng@sph.emory.edu (L. Peng).

http://dx.doi.org/10.1016/j.jmva.2014.06.005
0047-259X/

depicts the progression of disease while accounting for the presence of death, and moreover has the virtue of being nonparametrically identifiable [24]. In the absence of left truncation, analysis of $F_1(t)$ with semi-competing risks data can follow the approaches developed for competing risks data. For example, in the one-sample case, one may estimate $F_1(t)$ by adopting the methods in [1]. Under regression settings, several approaches have been developed to study the covariate effects on the cumulative incidence function. For example, a popular method is the proportional subdistribution hazards model proposed by Fine and Gray [6], which essentially amounts to a transformation model for $F_1(t|\boldsymbol{Z}) = \Pr(T_1 \le t, T_1 < T_2|\boldsymbol{Z})$ in the form of

$$F_1(t|\boldsymbol{Z}) = 1 - \exp\{-\Lambda_{10}(t)\exp(\boldsymbol{Z}^T\widetilde{\boldsymbol{\gamma}})\}. \tag{1}$$

Here $\Lambda_{10}(t)$ is the baseline cumulative subdistribution hazard, $\boldsymbol{Z}^T = (Z_1, Z_2, \ldots, Z_p)^T$ is the $p \times 1$ covariate vector, and $\widetilde{\boldsymbol{\gamma}}$ is the $p \times 1$ regression coefficient. Under model (1), covariates are assumed to have linear effects on $F_1(t|\boldsymbol{Z})$ after a complementary log–log transformation. Several other types of regression modeling of $F_1(t|\boldsymbol{Z})$ were studied by Fine [5] and Klein and Andersen [13], among others.

The data scenario focused in this paper is the semi-competing risks data subject to left truncation to the terminating event, as occurred in the Denmark diabetes registry example. While some efforts have been made to address the left truncation issue for subdistribution regression with competing risks data [9,32,29], naively adapting these methods to the semi-competing risks setting in the presence of left truncation would incur information loss from coercing semi-competing risks data into competing risks data, even in the crude analysis. For example, in the Denmark diabetes registry study, such "coercion" would require exclusion of subjects who had developed DN before the registry entry. As shown by Peng and Fine [21], such artificial truncation can incur considerable efficiency loss. In addition, Jiang et al. [12] have studied semi-competing risks data with the complication of left-truncation. They adopted the copula assumption to study the marginal quantity $P(T_1 \le t|\boldsymbol{Z})$, which differs from the $F_1(t|\boldsymbol{Z})$ of interest here.

It is also worth noting that most of the existing subdistribution regression methods constrain all covariate effects to be constant. This assumption may be too restrictive in many practical situations. For example, the efficacy of a treatment may fade over time due to drug resistance, or the gender difference may be unnoticeable at the beginning but gradually manifest itself over time. For cohort studies with long follow-up periods, such as the Denmark diabetes registry study, the validity of the constant effect assumption may be especially questionable. By these considerations, it is often of practical appeal to consider varying-coefficient regression models that can allow for changing effect patterns, thereby enabling a more comprehensive view of covariate effects, and perhaps, more accurate predictions.

Two modeling approaches have been studied for accommodating varying covariate effects on the subdistribution function of interest. One is the direct binomial regression modeling of cumulative incidence probability [28], which directly links covariates with the cumulative incidence function while allowing for regression coefficients to vary over time. This modeling strategy shares similar spirit of the additive complementary log–log survival model studied for randomly right censored data [23] but more flexibly accommodates the presence of competing risks. Of note, the direct binomial regression model of subdistribution offers a strict extension of Fine and Gray's [6]proportional subdistribution hazard model. More recently, Peng and Fine [22] proposed competing risks quantile regression which formulates covariate effects on a range of conditional quantiles of the cumulative incidence function. The regression coefficients are permitted to change with different quantile levels, and thus render a less restricted model compared to a competing risks accelerated failure time model. The method of Peng and Fine [22] was extended by Li and Peng [16] to handle left truncation to semi-competing risks data.

A general goal of this paper is to develop a new regression approach for left-truncated semi-competing risks data, which appropriately accounts for the specifical censoring and truncation features of this data type, and also has the flexibility to accommodate non-constant effects of covariates. In this work, we opt for the direct binomial regression modeling of subdistribution, which may lead to the following varying-coefficient subdistribution regression model,

$$F_1(t|\boldsymbol{Z}) = g\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(t)\}, \tag{2}$$

where $g(\cdot)$ is a monotone link function, $\boldsymbol{Z} = (1, \widetilde{\boldsymbol{Z}}^T)^T$, and $\boldsymbol{\beta}_0(t)$ is a $(p+1) \times 1$ vector of time-varying coefficients. Compared to competing risks quantile regression modeling in [22,16], model (2) provides a more direct approach to assessing the covariate effects on and predicting cumulative incidence probabilities, and thus may be preferred in scenarios where the scientific interest centers on the cumulative incidence function itself. The method developed by Scheike et al. [28] for model (2) purely focuses on standard competing risks data. How to extend their method to handle left truncation, particularly in the semi-competing risks setting, does not seem straightforward. In our simulation studies, we observe that naively applying Scheike et al.'s [28] method to left-truncated semi-competing risks data can incur large estimation bias.

In the rest of the paper, we propose estimation and inference procedures for the varying-coefficient subdistribution regression model (2), tailored to left-truncated semi-competing risks data. We construct monotone estimating equations that can be readily solved by existing statistical software. We establish the asymptotic properties of the proposed estimator including uniform consistency and weak convergence. The influence functions of the proposed estimators are also derived, and greatly facilitate the inference, such as Kolmogorov–Smirnov type and Cramér–Von-Mises type hypothesis testing procedures for the covariate effects. Via extensive simulations, we show that the proposed method performs well with various truncation rates and link functions. An application to the Denmark diabetes registry data demonstrates the practical utility of the proposed methods.