Computational Statistics and Data Analysis 119 (2018) 39–54

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

A Joint modelling of socio-professional trajectories and cause-specific mortality

M. Karimi [a,b,*], G. Rey [a], A. Latouche [c,d]

[a] Inserm-CépiDc, 80, Rue du Général Leclerc, Le Kremlin-Bicêtre, Cedex 94270, France
[b] MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK
[c] Conservatoire national des arts et métiers (Cnam), EA4629, Paris, France
[d] Inserm U900, Institut Curie, Saint-Cloud, France

ARTICLE INFO

Article history:
Received 6 March 2017
Received in revised form 27 September 2017
Accepted 7 October 2017
Available online 18 October 2017

Keywords:
Generalized linear mixed model
Cause-specific hazards
Joint model
Membership probability
Large-scale data

ABSTRACT

The association between life-course socio-professional trajectories and mortality has already been discussed in the literature. However, these socio-professional trajectories may be subject to informative censoring due to death. This loss to follow-up which is related to an individual's survival, should not be ignored and thus, it is of interest to model jointly these professional trajectories and their survival. The main focus has been made on continuous, binary or ordinal variables while much less attention has been paid to nominal categorical data. Therefore, an extension to the joint modelling of longitudinal nominal data and survival under a likelihood-based approach is proposed. A generalized linear mixed model is considered for modelling the longitudinal nominal data, in addition to two cause-specific proportional hazards model for the survival competing risks data. The association between longitudinal and survival outcomes is captured by assuming a multivariate Gaussian distribution for the joint distribution of the random effects of two sub-models. The proposed joint model provides a robust framework for estimating longitudinal membership probabilities, accounting for informative censoring caused by individual's death. Simulations are carried out to assess the performance of this joint model comparing with the results of the separate longitudinal and competing risks analysis. A disadvantage of joint models is that they are computationally intensive. To overcome this problem, an approach mimicking a meta-analysis strategy of individual participant data is suggested. The relevance of this approach is then illustrated on a large sample of the French salaried population, which contains all employment records between 1976 and 2002.
© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The motivation of this work comes from an epidemiological question that is measuring the association between life-course socio-professional position and the occurrence of mortality by different causes of death. For this purpose, we have the Cosmop-DADS database, which contains individuals' annual occupational episodes between 1976 and 2002 linked to the causes of death. The data on employment episodes are classified based on various social characteristics without an obvious hierarchical order such as income. In earlier studies it was common to consider just a limited number of episodes, either an

* Corresponding author at: MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK.
E-mail addresses: m.karimi@imperial.ac.uk (M. Karimi), gregoire.rey@inserm.fr (G. Rey), aurelien.latouche@cnam.fr (A. Latouche).

https://doi.org/10.1016/j.csda.2017.10.002
0167-9473/© 2017 Elsevier B.V. All rights reserved.

individual's employment at entry into the labour market or his/her position at mid-life age (Melchior et al., 2006; Davey Smith et al., 1997). To capture the evolution of socio-economic position and the transitions between social groups, the whole professional trajectory, which can be modelled as a sequence of consecutive professional positions occupied by an individual, should be taken into account. Model-based, such as Multistate models, and model-free such as Sequence Analysis approaches of life course analysis are traditionally used to study life course data, but these approaches consider completed trajectories only, and thus left-truncated and not randomly selected data (Eerola and Helske, 2012). Considering other characteristics of professional trajectories, such as cumulated time spent in each professional position and the number of transitions between professional episodes, in addition to the professional position itself, will aid in the understanding of professional trajectories process. To simultaneously consider the professional and mortality process, a naïve and common approach is to use the administrative employment positions and different characteristics of professional trajectories as time-dependent covariates in the extended Cox proportional hazards model (Karimi et al., 2015). However, the employment episodes are collected only for subjects under the study, and so are endogenous time-dependent covariates. Thus, analysis based on the extended Cox model (Andersen et al., 1993), which assumes that the covariates path is predictable, is not optimal for endogenous time-dependent covariates and might generate biases.

In this work, the focus is on the association between professional trajectories and death due to different causes of death. Death of each individual withdraws him/her from the study and results in informative censoring. Since the missing data conducted from this censoring is directly related to the individual's survival, the informative censoring is similar to the Missing Not At Random (MNAR) mechanism in longitudinal studies and should not be ignored. Consequently, it is of interest to model jointly the professional trajectories and the missingness process, in other words, the competing risks.

Joint analysis of longitudinal outcomes and survival data can be categorized into *pattern-mixture models*, *selection models* and *random effects models* (Diggle, 1998). Although mathematically all these models describe the joint distribution of longitudinal outcomes and survival data, they have different statistical interpretations. The pattern-mixture models focus on estimating the longitudinal trajectory conditional on the missingness process, in opposition to the selection models where the interest lies in estimating the missingness process given the longitudinal outcomes. The third category is based on the assumption that both missingness and longitudinal processes have arisen from an underlying health condition. This health condition is defined by a random effect and hence, longitudinal and missingness processes are independent, conditional on this random effect. Most of the proposed joint models in this category, known as *shared-parameter models*, use a shared random effect between longitudinal process and missingness mechanism. An extension of this model was proposed by Henderson et al. (2000) which postulates a latent bivariate Gaussian process. They assume that the longitudinal and missingness processes are independent given the Gaussian process and covariates. As a result, the association between longitudinal and missingness processes can be described by the correlation between the two components of the bivariate Gaussian process. Since both processes are not sharing the same random effects, this extension allows random effects to influence the missingness process independently of the longitudinal process. We focus on this category, and we refer to this class of models as *joint models for longitudinal and time-to-event data*.

The first forms of the joint models were introduced by Faucett and Thomas (1996), and Wulfsohn and Tsiatis (1997), and since several extensions have been proposed in the literature. The fundamental idea of the so-called joint models is based on linking the survival model with a suitable model for the longitudinal measurements, usually a random effect model (Laird and Ware, 1982), in which the correlations between the repeated measures are not ignored, via a common unobserved structure, to capture the correlations between the two longitudinal and survival processes. Different approaches have been developed in the literature for the association structure in joint models. One may include the mixed model defined for the longitudinal outcomes as a covariate in the survival sub-model (Faucett and Thomas, 1996; Rizopoulos, 2012). An alternative approach would be including the random effects directly in both longitudinal and survival sub-models with an assumed joint distribution for the random effects (Henderson et al., 2000; Elashoff et al., 2007; Li et al., 2010). Excellent overviews of these models can be found in the literature (Tsiatis and Davidian, 2004; Rizopoulos, 2013). Most of these works have focused on continuous measurements or the quality of life measurements, on binary measurements or ordinal responses and there has been less attention to non-ordinal categorical longitudinal outcomes, as in Cosmop-DADS database. The recent extension of the joint modelling of categorical longitudinal data and time-to-event data, developed by Murawska and Rizopoulos (2013), is based on a Bayesian approach. However, the main issue of this method is the high calculation time, even for a small dataset. Given the structure of our large dataset, the Cosmop-DADS database, in this work we extend the work of Li et al. (2010), by proposing a joint model for nominal longitudinal data and competing risks data in a likelihood-based framework.

To model the nominal longitudinal data, we focus on mixed-effects regression models. This choice is based on the characteristics of these classes of models that include subject-specific random effects into the regression models, that take into account for the influence of subjects on their repeated observations. Therefore, in the joint model, we adopt a Generalized Linear Mixed Model (GLMM) for nominal responses to model the longitudinal trajectories and two cause-specific proportional hazards models for competing risks survival data. The association between longitudinal and survival outcomes is captured by assuming a multivariate Gaussian distribution for the joint distribution of the random effects of two sub-models. The Expectation–Maximization (EM) algorithm is used to estimate the parameters in both longitudinal and survival sub-models, as well as the inverse of the empirical Fisher information from the profile likelihood for approximating the variance–covariance matrix of the parameter estimates. The proposed joint model also provides a robust framework to estimate longitudinal membership probabilities, accounting for the informative censoring.

So far, the existing joint models have been applied to sample sizes up to 2000 individuals, which is much smaller than the motivating database of this work. The second objective of this study is to propose an approach mimicking a meta-analysis,

Download English Version:

https://daneshyari.com/en/article/6868867

Download Persian Version:

https://daneshyari.com/article/6868867

Daneshyari.com