Bull. Aust. Math. Soc. **99** (2019), 349–352
doi:10.1017/S0004972718001399

JOINT MODELS FOR NONLINEAR LONGITUDINAL AND TIME-TO-EVENT DATA USING PENALISED SPLINES

HUONG THI THU PHAM

(Received 7 November 2018; first published online 7 January 2019)

2010 *Mathematics subject classification*: primary 62-07; secondary 62N05, 62F15, 65D07.

Keywords and phrases: survival data, longitudinal data, two-stage approach, shared random effects approach, joint models.

In follow-up studies, different types of response variables are collected for each individual. They are longitudinal outcomes which are measured on each subject repeatedly and the time when the subject meets an event of particular interest. There are many research questions focusing on the association between longitudinal data and survival time in clinical, epidemiological and educational studies. In many clinical studies, the researchers want to evaluate the impact of biomarkers for their prognostic capabilities on survival time outcomes. Tsiatis *et al.* [15] investigated the association between the number of CD4 lymphocytes and the time to death in an acquired immune deficiency syndrome (AIDS) study. The link between serum bilirubin level and survival time was investigated in liver cirrhosis studies [6, 10]. In addition, there has been interest in the interrelation between these two types of data in other fields. For instance, environmental factors or seasonal patterns may be associated with the occurrence of some types of diseases such as asthma or depression [9, 11].

Joint models aim to measure the association between survival time and longitudinal responses. These models can be used to better estimate the survival and longitudinal processes as well as evaluating their association. There are different types of longitudinal covariates for modelling survival time and trajectory for each individual. Flexible joint models are introduced to suit each type of longitudinal covariate and parametrise individual curves [1–3, 11, 14]. In addition, different approaches and techniques need to be considered to estimate parameters for joint models [4, 7, 8, 10, 13].

Cox [2, 3] introduced joint models using proportional hazard models. The Cox model has been, and remains, a very popular joint model to deal with time-independent

---

Thesis submitted to Flinders University in July 2017; degree approved on 6 April 2018; principal supervisor Darfiana Nur; co-supervisors Alan Branford and Murk Bottema.

covariates using a partial likelihood approach. However, the Cox model contains many disadvantages for handling time-dependent covariates [2]. Time-dependent covariates are also divided into two types, which are external and internal covariates. Cox [3] extended his method to handle the external longitudinal covariates. These models are known as the extended Cox models and also use the partial likelihood approach for estimation [1, 3–5].

Another category of time-dependent covariates is internal longitudinal outcomes, found in many clinical studies. The extended Cox model using a partial likelihood approach can cause large biases and poor coverage properties for handling internal covariates [12, 14]. Rizopoulos [11] proposed standard joint models postulating from the proportional hazard model. He used the full likelihood approach to estimate the parameters in the joint models. This approach performs acceptably for handling internal covariates and better than the Cox model and the extended Cox model [7, 11].

In the full likelihood approach, the whole history of biomarkers influences the survival function. Thus, it is important to obtain good models for longitudinal data in order to estimate the survival time accurately. Moreover, in practice, subject-specific trajectories may show nonlinear curves for a long period of measurement. Estimating parameters for standard joint models is often quick and easy. However, they may not fit nonlinear longitudinal data and cannot handle smoothing. This potential problem can be addressed by proposing an appropriate longitudinal submodel to handle nonlinear longitudinal data [7, 14]. In this thesis, we mainly focus on modelling the association between the internal nonlinear longitudinal outcomes and event-time outcomes as well as parameter estimation using different approaches.

This thesis introduces penalised spline joint models to handle nonlinear longitudinal outcomes in Chapter 3. These models are not only a good fit for nonlinear longitudinal data, but can also control the roughness of fit for the individual curves. To estimate the parameters in these models, the full likelihood approach is applied. In particular, parameter estimation is obtained by using the expectation conditional maximisation (ECM) algorithm. These models can improve the biases and the goodness of fit compared to the standard linear joint models. However, the penalised spline joint models can become complicated quickly when the number of knots in the longitudinal submodel increases. The full likelihood approach can lead to a computational problem for which the algorithm takes a long time to converge.

To deal with this computational problem, a modified two-stage approach is proposed in Chapter 4. We introduce an algorithm to estimate the parameters for the penalised spline joint models. This approach allows the allocation of as many knots as possible to the penalised spline joint models. In addition, this approach not only reduces the time for convergence, but also has biases comparable to the full likelihood approach.

Finally, to avoid the approximation from calculating multiple integrals in the frequentist approach and to quantify uncertainty using a probability density function for the penalised spline joint models, a fully Bayesian approach is applied to the penalised spline joint models in Chapter 5. In this approach, based on the likelihood function, we formulate the joint posterior distribution. The main algorithm using the Metropolis–Hastings (MH) and Gibbs sampler (GS) algorithms is proposed to sample the parameters for the penalised spline joint models. In addition, prior sensitivity analysis is performed to confirm the results of the inferences based on different prior distributions of some important parameters in joint models.

The thesis is organised into six chapters as follows: Chapter 1 is this introductory chapter. The background for longitudinal analysis, survival analysis and joint modelling is introduced in Chapter 2. The frequentist and Bayesian approaches for joint models are also reviewed in this chapter. Penalised spline models are proposed in Chapter 3. In this chapter, we also introduce the ECM algorithm and a set of R code written to estimate the parameters in the proposed joint models. The modified two-stage approach is introduced in Chapter 4. In this chapter, a proposed two-stage algorithm is also presented and a set of R code is provided. Intensive simulation studies are conducted to compare with the full likelihood approach. Chapter 5 uses a fully Bayesian approach to estimate parameters in the penalised joint models. The Markov chain Monte Carlo (MCMC) method is applied to sample parameters. Finally, conclusions about the main results obtained in this thesis, remaining problems and future research for joint models are discussed in Chapter 6.

References

[1] P. K. Andersen, O. Borgan, R. D. Gill and N. Keiding, *Statistical Models Based on Counting Processes* (Springer, New York, 1993).

[2] D. R. Cox, 'Regression models and life-tables', *J. R. Stat. Soc. Ser. B Stat. Methodol.* **34**(2) (1972), 187–220.

[3] D. R. Cox, 'Partial likelihood', *Biometrika* **62**(2) (1975), 269–276.

[4] D. R. Cox and D. V. Hinkley, *Theoretical Statistics* (Chapman and Hall/CRC Press, New York, 1979).

[5] D. Cox and D. Oakes, *Analysis of Survival Data* (Chapman and Hall, London, 1984).

[6] J. Ding and J. Wang, 'Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data', *Biometrics* **64**(2) (2008), 546–556.

[7] A. Gould, M. E. Boye, M. J. Crowther, J. G. Ibrahim, G. Quartey, S. Micallef and F. Y. Bois, 'Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group', *Stat. Med.* **34**(14) (2014), 2181–2195.

[8] J. G. Ibrahim, M. Chen and D. Sinha, *Bayesian Survival Analysis*, Wiley Online Library (John Wiley, Hoboken, NJ, 2005).

[9] J. Kalbfleisch and R. Prentice, *The Statistical Analysis of Failure Time Data*, 2nd edn (Wiley, New York, 2002).

[10] D. Rizopoulos, 'Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule', *Comput. Statist. Data Anal.* **56** (2011), 491–501.

[11] D. Rizopoulos, *Joint Models for Longitudinal and Time-to-Event Data with Applications in R* (Chapman and Hall/CRC Press, New York, 2012).

[12] M. J. Sweeting and S. G. Thompson, 'Joint modelling of longitudinal and time to event data with application to predicting abdominal aortic aneurysm growth and rupture', *Biom. J.* **53**(5) (2011), 750–763.

[13] A. A. Tsiatis and M. Davidian, 'A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error', *Biometrika* **88**(2) (2001), 447–458.

[14] A. A. Tsiatis and M. Davidian, 'Joint modeling of longitudinal and time-to-event data: an overview', *Statist. Sinica* **14** (2004), 809–834.
[15] A. A. Tsiatis, V. Degruttola and M. S. Wulfsohn, 'Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS', *J. Amer. Statist. Assoc.* **90**(429) (1995), 27–37.

HUONG THI THU PHAM, Department of Mathematics,
An Giang University, An Giang Province, Vietnam
e-mail: ptthuong@agu.edu.vn