An Ordinal Joint Model for Breast Cancer

Carmen Armero, Carles Forné, Montserrat Rué, Anabel Forte, Hector Perpiñán, Guadalupe Gómez and Marisa Baré

Abstract We propose a Bayesian joint model to analyze the association between longitudinal measurements of an ordinal marker and time to a relevant event. The longitudinal process is defined in terms of a proportional-odds cumulative logit model and the time-to-event process through a left-truncated Cox proportional hazards model with information of the longitudinal marker and baseline covariates. Both longitudinal and survival processes are connected by a common vector of random effects.

C. Armero (✉) · A. Forte · H. Perpiñán
Department of Statistics and Operational Research, Universitat de València, València, Spain
e-mail: carmen.armero@uv.es

A. Forte
e-mail: anabel.forte@uv.es

H. Perpiñán
e-mail: hector.perpinan@uv.es

C. Forné
Department of Basic Medical Sciences, Universitat de Lleida, IRB-Lleida and Oblikue Consulting, Lleida, Spain
e-mail: carles.forne@gmail.com

M. Rué
Department of Basic Medical Sciences, Universitat de Lleida, IRB-Lleida, Lleida, Spain
e-mail: montse.rue@cmb.udl.cat

H. Perpiñán
Fundación para el Fomento de la Investigación Sanitaria y Biomédica (FISABIO), Generalitat Valenciana, València, Spain

G. Gómez
Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain
e-mail: lupe.gomez@upc.edu

M. Baré
Clinical Epidemiology and Cancer Screening, Corporació Sanitària Parc Taulí-UAB, Sabadell, Spain
e-mail: mbare@tauli.cat

© Springer International Publishing AG 2017

E.A. Ainsbury et al. (eds.), *Extended Abstracts Fall 2015*,
Trends in Mathematics 7, DOI 10.1007/978-3-319-55639-0_2

1 Introduction

Joint modeling of longitudinal and time-to-event data is an increasing area of statistical research devoted to jointly analyze longitudinal and survival processes. It enhances longitudinal modeling by allowing for the inclusion of non-ignorable dropout mechanisms, and survival modeling by the inclusion of internal time-dependent covariates. Shared-parameter models are joint models connecting the longitudinal and time-to-event processes by means of common subject-specific random effects which, in the presence of covariates and parameters, endow both processes with conditional independence; see [4]. They can quantify both the population and individual effects of the longitudinal outcomes on the risk of an event, and obtain individualized dynamic predictions.

When longitudinal outcomes are ordinal, the non-linear nature of the data produce a complex likelihood function which is difficult to maximize under the frequentist paradigm. This paper discusses a Bayesian joint model for the association between longitudinal measures of an ordinal marker and a time-to-event outcome; see [1] for more details about the model. We propose a proportional-odds cumulative logit model [3] for the ordinal measurements based on the idea of a continuous latent variable, and a Cox proportional hazards model with left truncation for the time-to-event of interest with information of the longitudinal process. The model is applied to estimate the risk of breast cancer in women attending a population-based screening program with regard to longitudinal measurements of mammographic breast density.

2 A Bayesian Joint Model for Ordinal Longitudinal and Left Truncated Survival Data

Let $\{D_1, \ldots, D_K\}$ be the set of ordinal categories and y_{ij} the category of individual i, $i = 1, \ldots, n$, at time t_{ij}, $j = 1, \ldots, n_i$. We assume an underlying continuous latent variable y_{ij}^* that determines the ordinal category of individual i at time t_{ij}. This latent variable has no interest *per se* but it is useful for motivating and interpreting the longitudinal model. The relationship between y_{ij} and y_{ij}^* is the following

$$y_{ij} = D_k \;\Leftrightarrow y_{ij}^* \in (\gamma_{k-1},\, \gamma_k],\;\; k = 1, \ldots, K,$$

where $-\infty = \gamma_0 < \gamma_1 < \cdots < \gamma_K = \infty$ are unknown cutpoints. We choose a logistic distribution for y_{ij}^*, $\text{Lo}(m_{ij}, s = 1)$, with mean m_{ij} and scale parameter $s = 1$. The choice of that distribution implies a logit link for the cumulative probabilities as follows

$$q_{ijk} = P(y_{ij} > D_k) = P(y_{ij}^* > \gamma_k) = \frac{1}{1 + \exp(\gamma_k - m_{ij})}. \tag{1}$$

Despite $s = 1$ in the logistic distribution, the model is overparameterized. To obtain an identifiable model, we arbitrarily introduced a reference point on the latent scale, in particular $\gamma_{K/2} = 0$ if K is even and $\gamma_{(K-1)/2} = 0$ or $\gamma_{(K+1)/2} = 0$ if K is odd.

We consider a mixed-effects model to describe the subject-specific time trajectories of the latent variable

$$y_{ij}^* = m_{ij} + \epsilon_{ij} = \boldsymbol{x}_{ij}^{(l)\prime}\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{b}_i + \epsilon_{ij}, \tag{2}$$

where $\boldsymbol{x}_{ij}^{(l)}$ is a vector of covariates associated to individual i at time t_{ij} with regression coefficients vector $\boldsymbol{\beta}$; $\boldsymbol{z}_i$ a vector of explanatory variables attached to the random effects $\boldsymbol{b}_i$ for the i-th individual; and ϵ_{ij} an error term for the i-th individual at time t_{ij}, modeled in terms of the logistic distribution Lo(0, 1). The random effects $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)^T$ are conditionally i.i.d. $(\boldsymbol{b}_i \mid \boldsymbol{\phi}) \sim f(\boldsymbol{b}_i \mid \boldsymbol{\phi})$, where $f(\boldsymbol{b}_i \mid \boldsymbol{\phi})$ is usually taken to be a Multivariate Normal distribution with mean 0 and unknown covariance matrix.

Let T_i, $i = 1, \ldots, n$, be the observed event time for the i-th subject, obtained as the minimum between the true failure time, T_i^*, and the right-censoring time, C_i, $T_i = \min(T_i^*, C_i)$. The event indicator $\delta_i = I(T_i^* \leq C_i)$ takes the value 1 if the observed time is a true event time, and 0 otherwise. Event times corresponding to individuals who enter the study at delayed entry times introduce left-truncation. We define the hazard function of T_i^* in terms of the left-truncated Cox proportional hazard model [5]

$$h_i(t) = h_0(t \mid \boldsymbol{\lambda}) \exp\{\boldsymbol{x}_i^{(s)\prime}\boldsymbol{\eta} + \alpha\, m_{it}\}, \;\; t > a_i, \tag{3}$$

and zero otherwise, where $h_0(t \mid \boldsymbol{\lambda})$ is the baseline risk function with parameters $\boldsymbol{\lambda}$; $\boldsymbol{x}_i^{(s)}$ is the vector of baseline covariates with coefficients $\boldsymbol{\eta}$; α assesses the effect of the longitudinal marker of subject i on the event of interest in terms of the latent variable mean; and a_i is the delayed entry time of individual i.

We complete the Bayesian modeling eliciting a prior distribution, $\pi(\boldsymbol{\theta})$, for all the unknown parameters and hyperparameters of the model. From a Bayesian perspective, $\pi(\boldsymbol{\theta}, \boldsymbol{b} \mid \mathcal{D})$, where $\mathcal{D}$ represents all the data collected from the longitudinal and the survival processes, is the joint posterior distribution of the parameters, hyperparameters, and random effects, which can be obtained by hierarchical modeling.

3 Breast Cancer and Mammographic Breast Density

The joint model is applied to the assessment of breast cancer risk in women attending a population-based screening program including 13760 women aged 50–69 years who participated in the breast cancer early-detection program in the Vallès Occidental Est area in Catalonia (Spain), between October 1995 and June 1998; see [2].

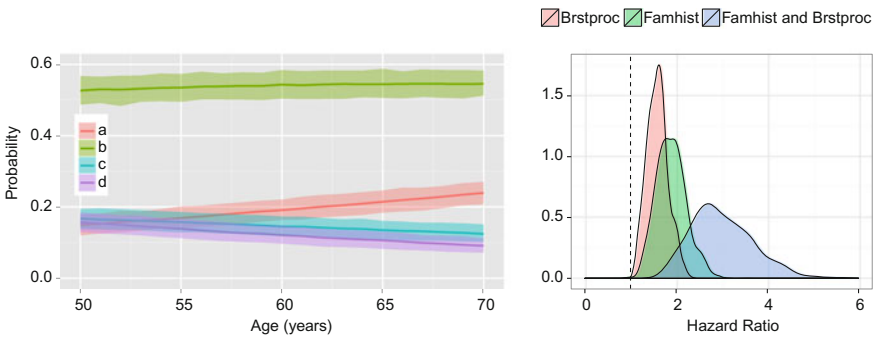


Fig. 1 Posterior mean and 95% credible band of the probability associated to each BI-RADS category with respect to age (*left*) and posterior distribution of the hazard ratios associated to family history of breast cancer, prior breast procedures, and both together (*right*)

The longitudinal ordinal marker is mammographic breast density in the scale BI-RADS, with a total of 81621 screening exams. The BI-RADS scale is ordinal with categories $\{a, b, c, d\}$, which represent low, medium, high, and very high breast density. The survival process focuses on time to a breast cancer diagnosis and incorporates *family history of breast cancer (Famhist)* and *prior breast procedures (Brstproc)* as dychotomous baseline covariates.

The posterior distribution is computed using Markov Chain Monte Carlo (MCMC) methods through the JAGS software. In particular, we run three MCMC chains with 100000 iterations, 10000 of which were used for the burn-in period. The chains were thinned by only storing every 270-th iteration to reduce autocorrelation in the saved sample. Convergence was assessed through the potential scale reduction factor and the effective number of independent simulation draws.

Figure 1 on the left shows the posterior mean and 95% credible interval of the posterior distribution associated to each BI-RADS category for a generic woman in the study. Probabilities associated to category b are always higher than 0.5, and grow slightly with age. Probabilities for categories a, c, and d are initially very similar, but categories c and d decrease with age following a similar pattern while category a increases. The credible intervals indicate high precision in the estimated means. Relevant hazard ratios (HRs) arise from the combination of covariate categories. Figure 1 on the right shows the posterior distribution of the HRs of a breast cancer diagnosis for *Famhist*, *Brstproc*, and both risk factors, with posterior means 1.864, 1.574, and 2.934, respectively. The marginal effects of each covariate are relevant, with posterior probabilities 0.998 and 1.000 to HR values greater than 1 for *Famhist* and *Brstproc*, respectively.

Figure 2 shows the posterior mean and 95% credible band for breast cancer-free survival for four women without cancer at the end of follow-up. Women 942, with stable very high breast density, is cancer-free at 68 years old and her predicted disease-free survival is higher than for women 9672, who has experienced a decrease in breast density and reaches as well 68 year old being cancer-free. The different density

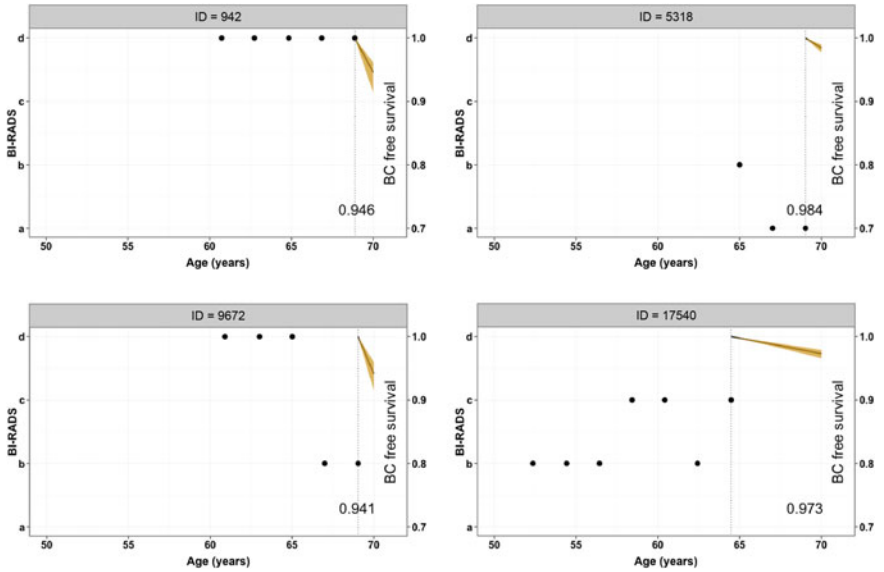


Fig. 2 Posterior mean and 95% credible band of the probability of a breast cancer-free diagnosis for women IDs 942, 5318, 9672 and 17540 without breast cancer at the end of the follow-up

behaviour might be attributed to the presence of prior breast procedures in woman 9672 and absence of them in woman 942. In general, breast cancer-free survival stays with high values, above 0.9, though they decrease with age. Furthermore, women with higher breast density values tend to have lower cancer-free survival and these probabilities depend in part of the corresponding baseline risk factors.

Acknowledgements This paper was partially supported by the research grants MTM2013-42323-P, MTM2012-38067-C02-1, PI14/00113 from the Spanish Ministry of Economy and Competitiveness, ACOMP/2015/202 from the Generalitat Valenciana, and GRBIO-2014-SGR464 and GRAES-2014-SGR978 from the Generalitat de Catalunya.

References

1. C. Armero, C. Forné, M. Rué, A. Forte, H. Perpiñán, G. Gómez, and M. Baré, "Bayesian joint ordinal and survival modeling for breast cancer risk assessment", *Stat. Med.* **35**(28) (2016), 5267–5282.
2. M. Baré, M. Sentís, J. Galcerán, A. Ameijide, X. Andreu, S. Ganau, L. Tortajada, and J. Planas, "Interval breast cancers in a community screening programme: frequency, radiological classification and prognostic factors", *Eur. J. Cancer Prev.* **17**(5) (2008), 414–421.
3. D.J. Lunn, J. Wakefield, and A. Racine-Poon, "Cumulative logit models for ordinal data: a case study involving allergic rhinitis severity scores", *Statistics in Medicine* **20** (2001), 2261–2285.
4. D. Rizopoulos, "Joint models for longitudinal and time-to-event data with applications in R", CRC Press, Biostatistics Series: Boca Raton, FL, 2012.
5. U. Uzunoḡullari and J.L. Wang, "Comparison of hazard rate estimators for left truncated and right censored data", *Biometrika* **79**(2) (1992), 297–310.

http://www.springer.com/978-3-319-55638-3

Extended Abstracts Fall 2015
Biomedical Big Data; Statistics for Low Dose Radiation Research
Ainsbury, E.A.; Calle, M.L.; Cardis, E.; Einbeck, J.; Gómez, G.; Puig, P. (Eds.)
2017, VII, 131 p. 24 illus., 17 illus. in color., Softcover
ISBN: 978-3-319-55638-3
A product of Birkhäuser Basel