Joint Modelling Analysis of Prostate Cancer Incidence: Frequentist and Bayesian Approaches

Piulachs, X.[1], *Serrat, C.*[2], *Rué, M.*[3], *Armero, C.*[4], *Forte, A.*[5], *Perpiñán, H.*[4], *Luján, M.*[6,7] *and Páez, A.*[8,7]

[1] xavierpiulachs@hotmail.com, Dept. of Statistics and OR, Universitat Politècnica de Catalunya
[2] carles.serrat@upc.edu, Dept. of Applied Mathematics I, Universitat Politècnica de Catalunya
[3] montse.rue@cmb.udl.cat, Institut de Recerca Biomèdica de Lleida, Universitat de Lleida
[4] {carmen.armero, hector.perpinan}@uv.es, Dept. of Statistics and OR, Universitat de València
[5] anabel.forte@uv.es, Dept. of Economics, Universitat Jaume I de Castelló
[6] mlujang@salud.madrid.org, Urology Dept., Hospital Universitario Infanta Cristina
[7] Spanish Branch of the European Randomized Study of Screening for Prostate Cancer
[8] alvaro.paez@salud.madrid.org, Urology Dept., Hospital Universitario de Fuenlabrada

Abstract

Prostate specific antigen (PSA) is a biomarker for prostate cancer (PCa) that is widely used for PCa screening. Using a database of 2415 men included in the Spanish screening arm of the ERSPC Study, we will use joint modelling strategies to analyze if longitudinal PSA profiles and time to PCa incidence allow to obtain a better estimate of the individual risk of PCa. Conclusions and limitations of the study will be discussed.

Keywords: Joint Modelling, Longitudinal data, Survival Analysis

AMS: AMS classification. (Optional)

1. Introduction

Screening consists of special exams to diagnose a disease when it is asymptomatic. Diagnosing and treating the disease early, before signs/symptoms appear, may result in more cures and lower mortality. In the last decades, screening for prostate cancer (PCa) using the *Prostate Specific Antigen* (PSA), a biomarker related to PCa, has been widespread in Western countries. PCa screening has remained controversial due to inconsistent results, in terms of benefit, of the two main studies, one in Europe and the other in the USA. In addition, screening healthy population may cause adverse effects, like overdiagnosis and overtreatment of indolent PCa.

In 1994 began the European Randomized Study of Screening for Prostate Cancer (ERSPC), which included eight participating European countries: Belgium, Finland, France, Italy, The Netherlands, Spain, Sweden, and Switzerland. The study recruited more than 250000 men. The aim of the study was to determine if screening reduced PCa specific mortality.

2. Data Description and Aims of the Study

In the Spanish center of the ERSPC ([7], [6]), 4278 male patients were included between 1996 and 1999, aged 45-75 years, and with a residual life expectancy greater than 10 years. Study variables include date of birth, date of randomization, study arm (screening or control), dates of attendance (PSA testing), biopsy results, Gleason score (which measures the aggressiveness of the tumor) and PCa diagnosis. 2416 patients were randomly assigned to the screening arm and 1862 to the control arm. Participants were followed for PCa incidence and mortality (caused by PCa or other causes). In this presentation we will focus on 2415 men in the screening arm, once the exclusion criteria were applied.

Within the ERSPC protocol, all men randomized to the screening arm underwent serum PSA determination, and in some cases also a digital rectal examination (DRE) and a transrectal ultrasound-guide test (TRUS) were performed. When PSA $\geq$ 3 ng/ml (protocol since May 1998), a sextant prostate biopsy was performed. The interval between screening exams was four years, although there were earlier recalls if an out of range PSA level was observed.

Age at randomization ranged from 45.5 to 71.0 years, with a mean and median of 57.7 and 56.9 yr, respectively. Follow-up time range was 0.03-9.31 yr, with mean 6.8 and median 6.6 yr. Among the studied men, 569 (23.6%) were visited only once, 1501 (62.2%) had two visits and 345 (14.3%) three or more. Patients who had more than one, that is with longitudinal information, had a mean and median follow-up times of 5.3 and 4.8 yr, respectively. Overall, the PSA mean was 1.87 ng/ml, and the 25%, 50% and 75% percentiles were 0.60, 1.09, and 2.10 ng/ml, respectively. Values for the DRE or TRUS variables were only available for 509 patients (21.1%).

As a result of the study protocol, 687 prostate biopsies on 407 patients (16.9% from the total) were obtained, of which 166 (24.2%) were performed during the first visit, while the remaining 521 (75.8%) were taken at further visits. On the other hand, 332 males (81.6%) had one biopsy performed, and other 75 (18.4%) underwent two or more of them. The screening process resulted in 587 (85.4%) negative biopsies and 100 (14.6%) positive results. In addition to the screen-detected cases, there were 15 interval cancer cases, resulting in 115 patients (4.8% from the total) diagnosed with PCa at the end of follow-up. This result indicates that the time to PCa diagnosis is subject to a 95.2% of right censoring.

The main scientific and statistical objectives in this study are: a) To estimate the time to PCa diagnosis taking into account longitudinal covariates like PSA, DRE, TRUS, among others, and b) To establish a relationship, based on the longitudinal profiles, between the type of tumor (screen-detected or interval case) and the Gleason score (as a measure of the tumor aggressiveness).

3. Joint Models

The methodological approach of this study will be joint modelling of longitudinal and time to event data. Joint models have attracted the interest of many researchers in Public Health with valuable contributions in the development of statistical methodology, both frequentist and Bayesian ([9], [1], [4]).

The implementation of this type of models is not easy and requires a deep exploratory and marginal statistical analysis which accounts for the different components of the problem separately ([5],[8], [3]) in order to improve our knowledge about the problem, which is particularly complex and polyhedral. In this stage we will examine the longitudinal process for the PSA, DRE and TRUS measurements within the framework of the Generalized (Hierarchical) Mixed Linear Models exploring models with different trajectory patterns (linear, quadratic, splines, etc) together with subject specific random effects.

If we denote by $y_i(t)$ the profile over time for a longitudinal outcome (*e.g.* PSA), let $m_i(t)$ be the true (and unobserved) value of the longitudinal outcome at time t. It is important to notice that $m_i(t) \neq y_i(t)$. And let $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$ be the longitudinal process up to time t (*e.g.* the PSA level history for the i-th patient up to time t). The main efforts of the joint modelling strategy will focus on relating this true longitudinal profile with the hazard function of the survival submodel through an association parameter, let say α.

By using the linear mixed effects paradigm $y_i(t)$ is modeled like

$$\begin{cases} y_i(t) &= m_i(t) + \epsilon_i(t) \\ m_i(t) &= x_i'(t)\beta + z_i'(t)b_i \\ b_i &\sim \mathcal{N}(0, D) \\ \epsilon_i(t) &\sim \mathcal{N}(0, \sigma^2) \end{cases}$$

where $x_i(t)$ and $z_i(t)$ are time-dependent design vectors, $\epsilon_i(t)$ is also time-dependent and the error terms are mutually independent and independent of the random effects b_i.

Concerning the survival submodel, the relative risk model is formulated in terms of the Cox proportional hazards model

$$h_i(t|\mathcal{M}_i(t), w_i) = h_0(t)\exp(\gamma' w_i + \alpha m_i(t))$$

where w_i is a vector of covariates, γ a vector of associated parameters, $m_i(t)$ is the true trajectory function of the longitudinal process, α is a parameter which connects the longitudinal processes to the hazard function and $h_0(.)$ and $h_i(.|.)$ are the baseline and conditional hazard functions, respectively. It is important to note that the parameter α (and its structure) has a main role in the model because it provides information about the association between the longitudinal and the survival process. Important issues in the specification of the survival model include the selection of the relevant covariates.

From a Bayesian point of view the model needs to be completed with a prior distribution for all the relevant uncertainties ([2]).

A first approach to the PSA trajectories shows an exponential pattern. We applied the logarithm transformation to accommodate for normality after adding the value 1 to all the original values (some PSA values were null). Figure 1 illustrates the variety of $\log(1 + PSA)$ profiles over time.

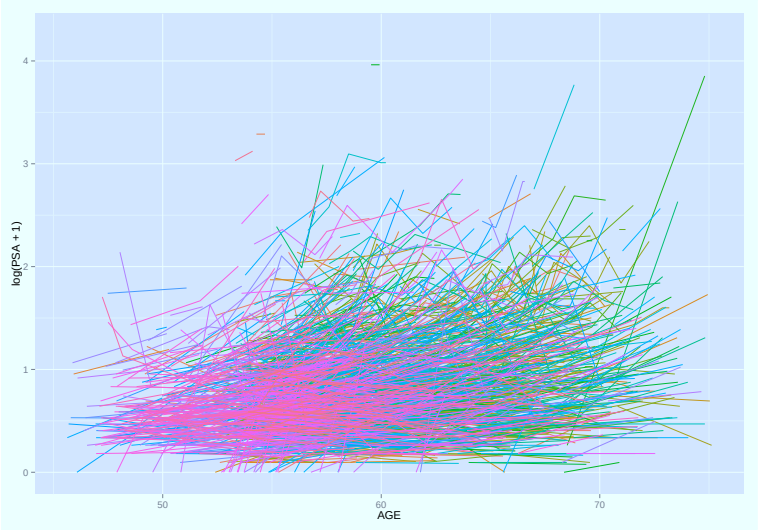

Figure 1: $\log(1 + PSA)$ profiles over time for patients in the Spanish branch of the ERSPC study

4. Discussion

The discussion of the results will focus in two main points. On one hand, in the similarities and differences between the results obtained from the two approaches, the frequentist and the Bayesian, and, on the other hand, in the limitations coming from the *short* longitudinal information in the dataset and the fact that less than 15% of the sample have more than three visits. Some recommendations will be given and discussed.

5. Acknowledgments

This work has been partially supported by the grant MTM2012–38067–C02–01 from the Ministerio de Economía y Competitividad of Spain. Authors are grateful to the TRUEJM group (Andrinopoulou, E.-R., Armero, C., Forné, C., Forte, A., Gómez, G., Lesaffre, E., Murawska, M., Perpiñán, H., Rizopoulos, D., Rué, M. and Serrat, C.) the fruitful discussions on joint modelling issues and to the Spanish Branch of the ERSPC study for providing the data.

6. Bibliography

[1] Brown E.R. and Ibrahim J.G. (2003). A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data. *Biometrics*, **59**, 221-228.

[2] Chi, Y.-Y. and Ibrahim, J.G. (2006). Joint Models for Multivariate Longitudinal and Multivariate Survival Data. *Biometrics*, **62**, 432–445.

[3] Huertas, J., Serrat, C., and Gómez, G (2012) Computational tools for joint modelling time to event and longitudinal data. *Technical Report, DR2012/04*, Dept Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain.

[4] Ibrahim, J.G., Chen M.H. and Sinha D. (2004). Bayesian Methods for Joint Modelling of Longitudinal and Survival data with application to Cancer Vaccine trials. *Statistica Sinica*, **14**, 863-883.

[5] Guo X. and and Carlin B.P. (2004). Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages. *The American Statistician*, **58**, 16-24.

[6] Luján, M. Páez, A. Berenguer, A. and Rodríguez, J.A. (2012) Mortalidad por cáncer de próstata en la rama española del *European Randomized Study of Screening for Prostate Cancer* (ERSPC). Resultados tras 15 años de seguimiento. *Actas Urológicas Españolas*, **36**, 403-409.

[7] Lujan, M., Paez, A., Miravalles, E., Fernandez, I., Llanes, L. and Berenguer, A. (2004). Prostate Cancer Detection is Also Relevant in Low Prostate Specific Antigen Ranges. *European Urology*, **45**, 155–159.

[8] Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Application in R.* Chapman and Hall/CRC Biostatistics Series.

[9] Wulfsohn, M.S., and Tsiatis, A.A. (1997) A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330–339.