Cancer Epidemiology 37 (2013) 449–456

Contents lists available at SciVerse ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Description of an approach based on maximum likelihood to adjust an excess hazard model with a random effect

Cyrielle Dupont [a,b,c,d], Nadine Bossard [a,b,c,d], Laurent Remontet [a,b,c,d], Aurélien Belot [a,b,c,d,e,*]

[a] Hospices Civils de Lyon, Service de Biostatistique, F-69003 Lyon, France
[b] Université de Lyon, F-69000 Lyon, France
[c] Université Lyon 1, F-69100 Villeurbanne, France
[d] CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Biostatistique-Santé, F-69100 Villeurbanne, France
[e] Institut de Veille Sanitaire, Département des maladies chroniques et traumatismes, F-94410 Saint-Maurice, France

ARTICLE INFO

Article history:
Received 20 July 2012
Received in revised form 26 March 2013
Accepted 2 April 2013
Available online 28 April 2013

Keywords:
Cancer
Excess mortality hazard
Survival
Net survival
Mixed-effect model

ABSTRACT

Objective: To adjust an excess hazard regression model with a random effect associated with a geographical level, the Département in France, and compare its parameter estimates with those obtained using a "fixed-effect" excess hazard regression model. *Methods*: An excess hazard regression model with a piecewise constant baseline hazard was used and a normal distribution was assumed for the random effect. Likelihood maximization was performed using a numerical integration technique, the Quadrature of Gauss–Hermite. Results were obtained with colon-rectum and thyroid cancer data from the French network of cancer registries. *Result*: The results were in agreement with what was theoretically expected. We showed a greater heterogeneity of the excess hazard in thyroid cancers than in colon-rectum cancers. The hazard ratios for the covariates as estimated with the mixed-effect model were close to those obtained with the fixed-effect model. However, unlike the fixed-effect model, the mixed-effect model allowed the analysis of data with a large number of clusters. The shrinkage estimator associated with Département is an optimal measure of Département-specific excess risk of death and the variance of the random effect gave information on the within-cluster correlation. *Conclusion*: An excess hazard regression model with random effect can be used for estimating variation in the risk of death due to cancer between many clusters of small sizes.

1. Introduction

The observed or overall survival probability is an important epidemiological indicator that allows quantifying the overall impact of a disease, such as cancer, on a given population. However, most cancers affect elderly persons who are also exposed to non-cancer mortality. Thus, a more useful survival indicator to assess and compare the performance of health systems regarding cancer management is net survival. Net survival is the survival that would occur if cancer were the only cause of death [1 pp 34–47, 2 pp 247–266]. It was recently shown by Pohar-Perme et al. that the "classical" relative survival methods (Ederer 1, Ederer 2, and Hakulinen) do not correctly estimate net survival because of their inability to take into account the informative censoring mechanism. An unbiased estimate of net survival can be obtained using either the non-parametric estimator of Pohar-Perme or a correctly specified multivariable regression model [3]. A recent simulation study showed the substantial biases associated with the "classical" relative survival methods [4] and a recent empirical study computed, on real data, the magnitude of the errors made with the classical relative survival methods used by cancer registries [5]. These unbiased approaches are very useful in analyzing population-based registry data because they do not require cause of death data – in cancer registries, these are often unavailable or unreliable. The general principle of these approaches is to use the mortality hazard of the general population (life tables) as reference mortality. Being non-parametric, the Pohar-Perme estimator does not allow estimating the impact of prognostic variables. For that purpose, the excess hazard regression model is more helpful.

Various parametric or semi-parametric excess hazard regression models were proposed during the last two decades. The main differences between them stem from variations in the modeling of the baseline hazard as well as from considering non-proportional and/or non-linear effects of the covariates [6–13]. All these models assume that survival times are independent between patients.

* Corresponding author at: Service de Biostatistique, Hospices Civils de Lyon, Centre Hospitalier Lyon Sud, F-69495 Pierre-Bénite Cedex, France.
Tel.: +33 4 788 657 73; fax: +33 4 788 657 74.
E-mail address: aurelien.belot@chu-lyon.fr (A. Belot).

1877-7821/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.canep.2013.04.001

However, the data collected by cancer registries have a hierarchical structure because each registry covers a pre-specified area, the Département in France. Due to shared characteristics (socioeconomic status, medical practices, environmental factors...), patients living in the same Département tend to have correlated survival times whereas survival times observed in two different Départements tend to differ systematically, generating heterogeneity between Départements.

Models with shared frailty (i.e., random effects applied to survival data) are able to take into account this correlation in the statistical inference [14]. This consists in adding a parameter common, called "shared frailty", to the individuals of each cluster of cases (i.e., the Département in this work). In overall mortality hazard models with shared frailty, a parametric distribution for the baseline hazard (e.g., Weibull) and a gamma distribution for the shared frailty are usually assumed. This is mainly due to a mathematical convenience because the marginal likelihood for a cluster has a simple analytic form [15 pp 43–61]. So, the maximization of the complete likelihood is straightforward with a standard optimization algorithm. However, in the excess hazard setting, these choices of distribution for the baseline hazard and the shared frailty do not lead to the same simplification and so, until now, there is no procedure in standard software packages to implement this type of model.

The main objective of this research work is to present an approach to adjust an excess hazard regression model with a random effect on Département (the geographical area of interest in our data). The second objective is to compare the estimated hazard ratios for the covariates obtained with an excess hazard regression model with a random Département effect versus those obtained with a model with a fixed Département effect.

In the first section, we describe the data used for illustration. Then we present the excess hazard regression model with fixed effects then with random effects. The following section describes and compares the results obtained with the two models on survival data on colon-rectum cancer (a low heterogeneity between Départements is expected) and on survival data on thyroid cancer (a higher heterogeneity between Départements is expected because of differences in cancer diagnostic techniques and treatments). Finally, we discuss various aspects of theses models and provide adequate advice concerning their use.

2. Materials and methods

2.1. The data

In France, cancer surveillance data are collected by Francim, the French network of cancer registries. Each register covers a whole Département. The survival data used in this work concern patients diagnosed between 1989 and 2004 and followed-up until 01/01/2008 (patients still alive at this date were censored). Colon-rectum cancer data were collected by eleven registries and thyroid cancer data by nine registries. The covariates used were: sex, age at diagnosis (in five age classes), and year of diagnosis (in five year classes). A more detailed description of these data is shown in Table 1.

For colon-rectum cancer, 64,171 patients were recorded of whom 39,251 (61.17%) died before the end of the follow-up. Sex, age at diagnosis, year of diagnosis, and Département are available for all patients (Table 1). The number of colon-rectum cancer cases was higher in men than in women and the mean age at diagnosis was 70.53 years. The number of cases between the eleven Départements was not even: the maximum was 9410 patients in Département 8 and the minimum 3013 cases in Département 7.

For thyroid cancer, 6199 patients were recorded of whom 780 (12.58%) died before the end of follow-up. The number of cases of thyroid cancer was greater in women than in men and the mean age at diagnosis was 50.74 years.

Table 1
Description of the data on colon-rectum and thyroid cancers according to the covariates and the Département.

Covariates	Colon-rectum cancer		Thyroid cancer	
	Number of cases	Frequency (%)	Number of cases	Frequency (%)
Sex				
Men	35,082	54.7	1332	21.5
Women	29,089	45.3	4867	78.5
Age at diagnosis				
[15; 45[	1907	3.0	2238	36.1
[45; 55[	5045	7.9	1521	24.5
[55; 65[	11,161	17.4	1173	18.9
[65; 75[	20,025	31.2	798	12.9
[75; +[	26,033	40.6	469	7.6
Year of diagnosis				
[1989; 1991]	8374	13.1	501	8.1
[1992; 1994]	10,286	16.0	591	9.5
[1995; 1997]	12,758	19.9	975	15.7
[1998; 2000]	13,643	21.3	1541	24.9
[2001; 2004]	19,110	29.8	2591	41.8
Département [a]				
1	5079	7.9	760	12.3
2	4549	7.1	534	8.6
3	4026	6.3	565	9.1
4	5669	8.8	1437	23.2
5	9253	14.4	747	12.1
6	8031	12.5	342	5.5
7	3013	4.7	715	11.5
8	9410	14.7	477	7.7
9	6515	10.2	622	10.0
10	4546	7.1	–	–
11	4080	6.4	–	–

[a] The Départements are anonymized; i.e., a given number does not correspond to the same Département (geographical area) for colon-rectum and thyroid cancer.

The net survival calculated with the non-parametric Pohar-Perme estimator [3] differed widely between Départements. For colorectal cancer, the net survival at 5 years ranged from 51.4% to 60.4% and at 10 years from 43.7% to 56.2%. For thyroid cancer, the net survival at 5 years ranged from 86.8% to 96.4% and at 10 years from 81.3% to 96.1%. This heterogeneity of net survival between Départements justifies the use of an excess hazard model with a random effect at the Département level.

2.2. The excess hazard regression model with fixed effects

Let t be the time since diagnosis, a the age at diagnosis, and $\mathbf{x}$ a vector of covariates for each patient. The expected mortality hazard in the general population, $\lambda_P(a+t,\mathbf{z})$, is defined according to the age at time t after diagnosis, $a+t$, and a vector $\mathbf{z}$ of demographic characteristics (usually, the sex, the year, and the Département). The expected mortality hazard is considered known; it is provided by the Institut National de la Statistique et des Etudes Economiques (INSEE). The excess mortality hazard is written $\lambda_E(t,a,\mathbf{x},\mathbf{z})$. The observed mortality hazard, $\lambda_{Obs}(t,a,\mathbf{x},\mathbf{z})$ can be considered as the sum of the expected mortality hazard, λ_P, and the excess mortality hazard, λ_E [6]:

$$\lambda_{Obs}(t,a,\mathbf{x},\mathbf{z}) = \lambda_E(t,a,\mathbf{x},\mathbf{z}) + \lambda_P(a+t,\mathbf{z})$$

An excess hazard regression model with piecewise constant baseline hazard constant per one-year intervals was used with covariates age at diagnosis, the sex, the year of diagnosis, and the Département. If we consider covariate "Département" effect as a