



# Kernel and regression spline smoothing techniques to estimate coefficient in rates model and its application in psoriasis

Freshteh Osmani<sup>1</sup>, Ebrahim Hajizadeh\*<sup>2</sup>, Parvin Mansouri<sup>3</sup>

Received: 29 Nov 2015

Published: 31 Aug 2019

## Abstract

**Background:** Data types are recurrent events in studies in which each person may experience an event at different times. One of the most popular approaches to analyze recurrent event data is obtaining an estimate of the means/rate of events at different times. In this context, determining the variability over time can help better understand the effect of factor on the response. In this study, we applied smoothing methods to estimate coefficients in time-dependent rate model, and we also showed its application in data of psoriasis patients.

**Methods:** In the present study, psoriasis patients who experienced relapse that led to hospitalization during 2005 and 2014 in the Dermatology Department of Imam Khomeini hospital in Tehran were examined. To investigate the rate of relapse during a year, time-dependent rate model was used and variability of the effects was assessed using Wald test. Both b-spline and kernel methods were used to estimate time varying coefficients in rates model. Finally, results from methods were compared based on the obtained estimates.

**Results:** Based on the results of the Wald test, the effect of season on the occurrence of psoriasis was significantly different ( $p < 0.01$ ). Also according to the estimated coefficients from both b-spline and kernel methods, there was little difference between them.

**Conclusion:** In situations in which the effect of a variable is different at different times, using time-dependent coefficients rate model can lead to a better estimate of the effect of variable on the response. On the other hand, smoothing methods can smooth the effects of the variables that vary over time.

**Keywords:** Recurrent event, Regression spline, Kernel, Rates model, Psoriasis

**Conflicts of Interest:** None declared

**Funding:** None

\*This work has been published under CC BY-NC-SA 1.0 license.

Copyright© Iran University of Medical Sciences

**Cite this article as:** Osmani F, Hajizadeh E, Mansouri P. Kernel and regression spline smoothing techniques to estimate coefficient in rates model and its application in psoriasis. *Med J Islam Repub Iran.* 2019(31 Aug);33:90. <https://doi.org/10.34171/mjiri.33.90>

## Introduction

Psoriasis is a common chronic inflammatory and hyperproliferative skin disease. It affects approximately 2% of the world's population, with equal distribution among sexes (1). According to the current concept, psoriasis is a

complex genetic disorder caused by the interplay of many different genes (2). Some candidate genes such as p53 are suspected of being involved in the pathogenesis of the disease (3). Unfortunately, accurate statistics on the preva-

**Corresponding author:** Dr Ebrahim Hajizadeh, [hajizadeh@modares.ac.ir](mailto:hajizadeh@modares.ac.ir)

<sup>1</sup> Infectious Diseases Research Center, Birjand University of Medical Sciences, Birjand, Iran

<sup>2</sup> Department of Biostatistics, Faculty of medical sciences, Tarbiat Modares University, Tehran, Iran

<sup>3</sup> Department of Dermatology, Faculty of Medicine, Tehran university of Medical Sciences, Tehran, Iran

### ↑What is "already known" in this topic:

Recent pharyngeal infection has been implicated in the case of psoriasis. In addition, the risk for an aggravation or an abrupt recurrence of this autoimmune disease is very high. In spite of the high incidence and recurrence rate of psoriasis in Iran, the etiological studies in this field are very sparse.

### →What this article adds:

Data on recurrent events have been studied extensively using marginal and random effects models. In these models, it is often assumed that the effect of risk factors is unchanged over the follow-up period; however, time-dependent coefficient rates model is appropriate in situations where the effect of an exposure or covariates changes over time.

lence of psoriasis is not available in Iran. Although a genetic basis exists, environmental triggers exacerbate the symptoms. Various lifestyle factors (eg, smoking, alcohol use, obesity, and emotional stress) have been associated with its morbidity (4-7). In the case of psoriasis, recent pharyngeal infection has been implicated. In addition, the risk for an aggravation or an abrupt recurrence of this autoimmune disease is very high. In spite of the high incidence and recurrence rate of psoriasis in Iran, the etiological studies in this field are very sparse. Cox's proportional hazards model is the most commonly used model in survival analysis (8), which provides estimates of the relative risk associated with time-to-event occurrence. In longitudinal studies, events of interest are often observed for multiple times, for example, hospital readmissions after stroke or episodes of wheezing among young children. These types of recurrent events data have been studied extensively using marginal (rate or mean) models (9) and random effects models (10). In these models, it is often assumed that the risk factor's effect is unchanged over the follow-up period. This convenient assumption may not be realistic and may lead to bias in the coefficient estimates. Time-dependent coefficient rates model is a statistical analysis method that would be appropriate in situations in which the effect of an exposure or covariates changes over time in recurrent event data. Non-parametric estimates of the rate function derived from methods above generate a step-function estimator. Here, we aimed to analyze recurrent event data using regression splines and kernel smoothing technique to model the rate function. Regression splines are a natural extension of the piece-wise constant models described above and have been used recently in survival analysis using the Cox model. For example, Sleeper and Harrington model covariate effects using fixed-knot regression splines (11). Rosenberg applied cubic regression splines to estimate the baseline hazard function (12). Kooperberg used linear splines and their tensor products to estimate one or more covariate (13). The main contribution of this study was using a flexible, robust, and efficient combination of smoothing techniques for coefficient estimation in rates model. The model proposed herein is similar to the proportional intensity poison parametric model except that the coefficient was estimated using regression splines and kernel, and the variable checked in this study was the different seasons of the year; moreover, the impact associated with recurrence of psoriasis was identified.

## Methods

This was a longitudinal study conducted at the Dermatology Department of Imam Khomeini hospital in Tehran during March 2006 and April 2014, Iran. Only patients who had relapses that led to hospitalization and regularly referred to the hospital were included. After a preliminary assessment, a total of 160 patients were identified. Interviews were performed using a structured questionnaire. The important variable in this study was season of relapse and survey effects of each season on recurrence of psoriasis. The current study was approved by the Ethical Committee of Medical Science Faculty of Tarbiat Modares

University. Rates models have been used to analyze multiple time-to-event data, where the rate of recurrence is modeled as a function of observed covariates and the effect of the covariates is assumed to be constant (14).

In this study, we focused on a time-to-event approach for recurrent events that allowed us to estimate the effects that may change over time. By this, we properly modeled the functional form of covariates using a rates model that incorporates a smoothing technique called regression splines.

The rate function for the its individual is modeled as:

$$d\mu_i(t) = \exp\{\beta^T Z_i(t) + \theta(t) W_i(t)\} d\mu_0(t)$$

where  $\beta$  is a  $(p-1) \times 1$  vector of fixed regression parameters  $\theta(t)$  is the time-varying regression parameter and  $d\mu_0(t)$  is the baseline rate function. The covariates  $Z(t)$  and  $W(t)$  could be time independent or time-dependent. For instance, when  $W$  is a time-independent binary exposure or covariate such as a treatment group, the RR of the 2 groups at time  $t$  is given by exponentiating  $\theta(t)$  (ie,  $RR(t) = \exp(\theta(t))$ ) (13). The estimation of  $\theta(t)$  might be done by approximating  $\theta(t)$  through standard cubic B-spline basis functions  $\tilde{B}_k(t)$ , ( $k = 1, \dots, m+3$ ), such as

$$\hat{\theta}(t) = \tilde{\gamma}_0 + \sum_{k=1}^{m+3} \tilde{\gamma}_k \tilde{B}_k(t)$$

where  $m$  defines the number of interior knots. Splines are piecewise polynomials satisfying continuity constraints at the knots joining the pieces. B-splines, originally introduced by de Boor (15), are a popular type of regression splines in statistical applications, mainly due to their flexibility and numerical properties. The proposed model uses products of a covariate and B-spline functions of time to yield models that allow effects to change over time in a flexible way. Thus, by replacing  $\theta(t)$  by its B-spline approximation in the above time-varying coefficient model, we have

$$d\mu_i(t) = \exp\{\beta^T Z_i(t) + (\gamma_0 + \sum_{k=1}^{m+3} \gamma_k \tilde{B}_k(t)) W_i(t)\} d\mu_0(t)$$

By multiplying  $W_i(t)$  we have

$$d\mu_i(t) = \exp\left\{\beta^T Z_i(t) + \gamma_0 W_i(t) + \sum_{k=1}^{m+3} \gamma_k \tilde{B}_k(t) W_i(t)\right\} d\mu_0(t)$$

Suppose we define matrix  $\tilde{W}_i(t)$  and vector  $\gamma$  as follows:

$$\tilde{W}_i(t) = (W_i(t), \tilde{B}_1(t) W_i(t), \tilde{B}_2(t) W_i(t), \dots, \tilde{B}_{m+3}(t) W_i(t))'$$

$$\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{m+3})'$$

Thus, we can write the rate model as follows:

$$d\mu(t) = \exp\{\beta^T Z(t) + \gamma' \tilde{W}(t)\} d\mu_0(t)$$

AIC criterion was used to determine the number of nodes. Therefore, first by using different nodes ( $m=1, 2, \dots, 6$ ), different models were fitted, and AIC index was calculated for each of them. Then, the number of nodes corresponding to the model with the lowest value of AIC was selected. Considering that the number of knots is held

and fixed as the sample size  $n \rightarrow \infty$ , we defined a Wald-type statistic to test whether  $\theta(t)$  is constant over time.

$\gamma^* = (\gamma_1, \dots, \gamma_{m+3})'$ ; thus, the hypothesis of interest is that  $H_0 : \gamma^* = 0$  by analogy with the usual parametric likelihood procedures; this statistic can be defined by:  $Q_w = (\hat{\gamma}^*)' (Cov(\hat{\gamma}^*))^{-1} \hat{\gamma}^* \sim \chi^2_{1-\alpha, m+3}$

$Cov(\hat{\gamma}^*)$  is the  $(m+4) \times (m+4)$  matrix on the right bottom side of  $\hat{\Gamma}$ . The test rejects for large values of the statistic. Under the fixed knot framework, it is further assumed that the usual conditions are satisfied so that the standard asymptotic results hold for this model. Hence, under the null hypothesis, the statistic QW asymptotically follows a chi-square distribution with  $(m+3)$  degrees of freedom (15). An advantage of such an approach is that the estimates can be obtained by software R Version 3.0.3 with statistical package {stats} for kernel and package bs for B-splines smoothing.

**Results**

A total of 160 patients who had been diagnosed with psoriasis were studied; and 294 (148) relapses occurred that led to hospitalization of the patients. Separation of recurrence season is demonstrated in Table 1 and Fig. 1. As displayed, 45 (15.3%) relapse cases occurred in the spring, 85 (28.91%) in the summer, 72 (24.48%) in the autumn, and 92 (31.29%) in the winter. Table 1 demonstrates the frequency of relapses leading to hospitalization according to sex in different seasons.

To fit the model whose specific coefficient was time-dependent, b-spline functions with third degree as well as kernel functions were used. Thus, first, we had to specify the number of nodes used to build B-spline; to do so, the model was fitted using 2, 3, and 6 knots. For each case, AIC criteria were calculated and the model with a minimum AIC value was selected as our final model. According to the results of AIC value, we used  $m = 3$  knots to make B-spline.

At this stage, we used kernel function to fit the specific model in which the coefficient was time dependent. After selecting different kernels with 2 band width (10 and 50), we obtained similar results, and finally, we used kernel with bandwidth 50.

In general, according to the results of Tables 2 and 3, the highest risk of relapse occurred in the winter compared to the spring (about 7 times higher risk of recurrence in the winter). To test whether the recurrence of psoriasis is different in different seasons of the year, we used the following statistic:

with substituting the following:

$$Q_w = 43.06, \chi^2_{0/95,6} = 14.07$$

$$p\text{-value} < 0.01$$

Therefore, it can be concluded that the effect of season on the recurrence of psoriasis is variable over time. To test whether the model coefficients vector was zero or not, Wald type statistic was used.

$$\chi^2_w = 253.2, \chi^2_{0/95,12} = 21.02, p < 0.001$$

Therefore, the null hypothesis was rejected.

Table 1. Frequency of relapses leading to hospitalization in different seasons by gender

Season	Sex				Total	
	Female		Male		n	%
	n	%	n	%		
Spring	27	18.24	18	12.32	45	15.3
Summer	55	37.16	30	20.6	85	28.91
Autumn	29	19.6	43	29.45	72	24.48
Winter	37	25.1	55	37.67	92	31.29
Total	148	100.0	146	100.0	294	100.0

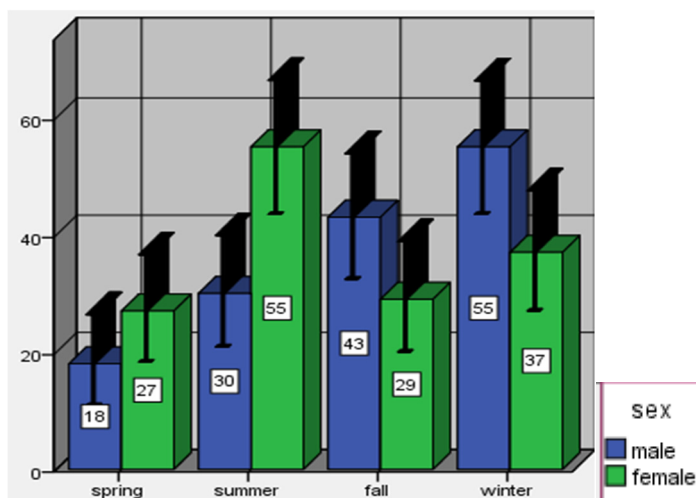


Fig. 1. Distribution of relapses in different seasons by gender

Table 2. Estimated coefficients of "relapse season" in the time-dependent rate model with b-spline smoothing

Relapse season	$\beta$	Standard Error	$Exp(\beta)$	0.95 Confidence interval for the Rate ratios	p
Summer	1.79	0.51	5.007	(3.75,10.05)	0.004
Autumn	1.37	6.99	4.02	(2.22,7.33)	0.041
Winter	2.03	1.35	7.21	(2.91,15.56)	0.003

Table 3. Estimated coefficients of "relapse season" in the time-dependent rate model with kernel smoothing

Relapse season	$\beta$	Standard Error	$Exp(\beta)$	Confidence interval for the Rate ratios .95	p
Summer	1.72	0.62	4.98	(3.25,9.75)	0.004
Autumn	1.39	6.99	4.21	(2.02,7.56)	0.038
Winter	2.11	1.35	7.25	(2.67,16.06)	0.002

## Discussion

This study was designed to find the effect of season variability on psoriasis recurrence. In this study, 2 smoothing methods were used to estimate time-dependent effects in the rates model under the independent censoring model on the recurrent event process. In fact, we used the smoothing method to obtain better estimates of the effects. Several investigators (16, 17) have used spline functions to model the relative risk in the proportional hazards model. By introducing regression splines, which are splines with a small number of knots, in the marginal rates model and extending the known methods for recurrent time-to-event data, we developed a method that allows the investigator to describe with details the behavior of the effects of interest over time on the rate of event occurrence. The splines are well known for their usefulness in providing a smooth approximation to a covariate function. A spline is a piecewise polynomial and its shape depends on the degree of the spline function and on the number and location of the breakpoints or knots. The proposed method was applied to evaluate the relationship between recurrence of psoriasis and seasons and according to the obtained results the impact of season on recurrence of psoriasis was different. Much effort has also been devoted to the development of methods to estimate the means/rates of recurrent events in the recent years (2, 8, 18). Lin and Wei used this model to examine the factors influencing the rate of relapse in infectious patients. Amorim et al. used rates model with time-dependent coefficients for the first time and showed that the effect of vitamin A on reducing diarrhea has been dropping over time (19). However, other studies did not find significant reductions in either the incidence or the mean daily prevalence of diarrhea, so using the proportional rates model in conditions when the effect of the interest variable under study on the occurrence of events vary at different time points, can be ignored or can reduce the visibility of the variable's effect. In summary, a time-dependent coefficient rates model with small number of knots was proposed to estimate the effects in the rates model that may vary over time. The inclusion of splines and kernel technique in the estimation procedure provided flexibility to capture the time-varying effect and allowed the inferences using standard techniques. Compared with the existing models in the literature, the proposed smoothed rate model with time-varying coefficients is more versatile and flexible as it can summarize the effects more clearly and is more effective as a diagnostic tool for time-varying coefficients in the marginal means and rates models. The choice of the bandwidth will affect the re-

sults. We examined some different choices of bandwidths for kernel method and obtained almost identical results. For the 2 versions of smoothing the model, the estimates of the effects of seasons on relapse of psoriasis appear to be rather stable to different choices of smoothing methods.

## Conclusion

It seems that these differences are due to biological and climatic conditions as well as different nutritional conditions. Therefore, we recommend examining all patients with psoriasis, regardless of their disease severity. This methodology may be potentially useful for describing the behavior of many other exposures or covariates associated with research questions in epidemiology and public health. The smoothing results also suggest that proposed estimates are robust to choices of bandwidths in the scenarios considered. For being unknown dependent coefficients in the medical field, there was allowing comparison of the method used in this study with no similar method.

## Acknowledgments

Authors wish to acknowledge their great appreciation for the valuable contribution of Tarbiat Modares University and Dermatology Department of Imam Khomeini hospital to this study.

## Conflict of Interests

The authors declare that they have no competing interests.

## References

1. Raychaudhuri SP, Farber E. The prevalence of psoriasis in the world. *J. Eur. Acad. Dermatol. Venereol.* 2001;15(1):16-7.
2. Guilhou JJ, Molès JP. New hypotheses in the genetics of psoriasis and other 'complex' diseases. *Dermatology.* 2008;216(2):87-92.
3. Yazici AC KA, Ozen O, EkSiolu M, Ustün H. Expression of p53 in lesions and unaffected skin of patients with plaque-type and guttate psoriasis quantitative comparative study. *J Dermatol.* 2007;34:367-74.
4. Higgins EM PT, du Vivier AW. Smoking, drinking and psoriasis. *Br J Dermatol.* 1993;28:730-2.
5. Kormeili T, Lowe NJ, Yamauchi PS. Psoriasis: immunopathogenesis and evolving immunomodulators and systemic therapies; U.S. experiences. *Br J Dermatol.* 2004;151(1):3-15.
6. Naldi L, Chatenoud L, Linder D, Fortina AB, Peserico A, Virgili AR, et al. Cigarette smoking, body mass index, and stressful life events as risk factors for psoriasis: result from an Italian case-control study. *J Invest Dermatol.* 2005;127:61-7.
7. Naldi L, Parazzini F, Brevi A, Peserico A, Fornasa CV, Grosso G, et al. Family history, smoking habits, alcohol consumption and risk of psoriasis. *Br J Dermatol.* 1992;127:212-7.
8. Lin D, Wei L, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *J Royal Stat Soc: Series B*

- (Statistical Methodology). 2000;62(4):711-30.
9. Duchateau L, Janssen P. The frailty model: Springer Science & Business Media; 2007.
  10. Sleeper LA, Harrington DP. Regression splines in the Cox model with application to covariate effects in liver disease. *J Am Stat Assoc.* 1990;85(412):941-9.
  11. Rosenberg PS. Hazard function estimation using B-splines. *Biometrics.* 1995;874-87.
  12. Dauxois JY, Sencey S. Non-parametric Tests for Recurrent Events under Competing Risks. *Scand J Stat.* 2009;36(4):649-70.
  13. Osmani F, Hajizadeh E, Rasekhi A, Akbari ME. Analyzing Relationship Between Local and Metastasis Relapses with Survival of Patients with Breast Cancer: A Study Using Joint Frailty Model. *IJCM*, 2018 Dec 1 (In Press).
  14. Osmani F, Hajizadeh E, Mansoori P. Use of Smoothing Methods for Estimating the Coefficients of Time Dependent Rate Models in Survival Analysis and its Application in Psoriasis Disease. *Iran J Epidemiol.* 2016;12(3):36-46.
  15. Nan B, Lin X, Lisabeth LD, Harlow SD. A Varying-Coefficient Cox Model for the Effect of Age at a Marker Event on Age at Menopause. *Biometrics.* 2005;61(2):576-83.
  16. Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Esteve J, Gouvernet J, et al. A relative survival regression model using B-spline functions to model non-proportional hazards. *Stat Med.* 2003;22(17):2767-84.
  17. Osmani F, Hajizadeh E, Rasekhi A A. Joint Frailty Modeling for Multiple Recurrent Events and Its Application in Patients with Breast Cancer. *JOGCR.* 2018;3(2).
  18. Osmani F, Hajizadeh E, Rasekhi A, Akbari M E. Multivariate Joint Frailty Models for Modeling Multiple Recurrent Events and Its Application in Breast Cancer. *Razi J Med Sci.* 2019;25(10):71-7919.
  19. Amorim LD, Cai J, Zeng D, Barreto ML. Regression splines in the time-dependent coefficient rates model for recurrent event data. *Stat Med.* 2008;27(28):5890.