

Análisis de la evolución temporal de la mortalidad mediante modelos lineales generalizados

Xavi Puig^{a,b} / Josep Ginebra^b / Rosa Gispert^a

^aServei d'Informació i Estudis. Departament de Salut. Generalitat de Catalunya. Barcelona. España.

^bDepartament d'Estadística. ETSEIB. Universitat Politècnica de Catalunya. Barcelona. España.

(Analysis of mortality time trend using generalized linear models)

Resumen

En este trabajo se muestra cómo los modelos lineales generalizados permiten describir eficientemente diferentes patrones de evolución temporal de datos de mortalidad y, a su vez, llevar a cabo una fácil interpretación. Como aplicación práctica se analiza la evolución de la mortalidad por cáncer de mama en las mujeres de Cataluña entre 1986 y 2000. De los resultados destaca que la mortalidad por cáncer de mama experimenta un aumento y un posterior descenso para todos los grupos de edad. El año en que se inicia el descenso es más reciente en los grupos de edad mayor.

Palabras clave: Tendencias temporales. Regresión de Poisson. Métodos epidemiológicos.

Abstract

In this work it is shown how generalized linear models allow one to describe different patterns of temporary evolution of mortality data, while at the same time allow for an easy interpretation. As a practical application, the evolution of the female breast cancer mortality in Catalonia from 1986 to 2000 is analyzed. Remarkably, the mortality from breast cancer first increases and then decreases for all age groups. Moreover, the year in which the cancer rate starts decreasing is more recent in the older age groups.

Key words: Time trends. Poisson regression. Epidemiologic methods.

Introducción

El análisis de la evolución temporal a partir de una serie de datos sanitarios (p. ej., de mortalidad) puede ser de gran utilidad para la monitorización del estado de salud de una población y para valorar el impacto de algunas intervenciones sanitarias.

La representación gráfica de una serie temporal, a pesar de estar entre los procedimientos de análisis más sencillos, aporta información muy valiosa e ilustrativa. Sin embargo, a menudo los métodos gráficos resultan insuficientes para estudiar la evolución temporal de series numéricas, puesto que no permiten alcanzar una idea clara y concluyente de la magnitud del fenómeno,

bien porque la representación de resultados queda distorsionada por problemas de escala, o bien porque el número de representaciones gráficas es elevado y no permite obtener una interpretación sintética para el conjunto de los datos. Una alternativa es recurrir a modelos estadísticos como herramienta para suavizar los datos eliminando el ruido, y de esta manera representar la realidad de forma parsimoniosa y objetiva.

Los modelos lineales generalizados ofrecen una gran versatilidad, a menudo desaprovechada. Si bien estos modelos se utilizan desde hace tiempo para evaluar tendencias temporales de mortalidad^{1,2}, se suelen asumir tendencias lineales y homogéneas para todos los grupos de edad, cuando dicho supuesto muchas veces no coincide con la realidad. En este trabajo presentamos diversos modelos de estructura polinómica, que tienen una fácil interpretación y que nos permiten abarcar una gran variedad de patrones de evolución temporal, aplicables a series de datos sanitarios.

Para mostrar un ejemplo de su utilización hemos escogido la serie de mortalidad por tumor maligno de mama femenina en Cataluña entre 1986 y 2000. El cáncer de mama es la causa de muerte por cáncer más frecuente entre las mujeres, y la primera causa de muerte entre las edades de 35 a 64 años. Aunque no se co-

Correspondencia: Xavi Puig.

Servei d'Informació i Estudis. Departament de Salut.
Generalitat de Catalunya.
Travessera de les Corts, 131-159. Pavelló Ave Maria.
08028 Barcelona. España.

Correo electrónico: xavier.puig@upc.edu

Recibido: 22 de octubre de 2004.

Aceptado: 25 de julio de 2005.

nocen de forma clara sus factores de riesgo más importantes, el abordaje terapéutico del que está siendo objeto durante estos últimos años ha tenido una repercusión muy importante sobre la supervivencia, motivo por el cual el análisis de su evolución temporal es de gran interés sanitario.

Material y métodos

Se han utilizado las defunciones por cáncer de mama de las mujeres residentes en Cataluña ocurridas entre los años 1986 y 2000, procedentes del Registro de Mortalidad de Cataluña. Se incluyen, pues, todas las defunciones en las que la causa básica de muerte (la que origina el proceso letal) es el tumor maligno de mama (Clasificación Internacional de Enfermedades, novena revisión [CIE-9]: 174 entre el año 1986 y el 1998 y C50 de la CIE-10 para los años 1999 y 2000). La población proviene de estimaciones intercensales y poscensales elaboradas por el Institut d'Estadística de Catalunya a partir de censos y padrones de los años 1986, 1991 y 1996.

En el análisis gráfico se han representado las tasas anuales de mortalidad por tumor de mama ajustadas por edad según el método directo, utilizando como población de referencia el censo de Cataluña de 1991. Con el objetivo de observar si el patrón global se repite para cada grupo de edad y si hay diferencias en la magnitud de las tasas, se han representado las tasas específicas por grupos de edad.

En el ajuste de los modelos se han truncado los datos para las mujeres mayores de 34 años y se han agregado los datos en grupos decenales de edad; el último abierto es el de ≥ 85 años.

Para el análisis de la evolución temporal de la mortalidad se han utilizado modelos log-lineales que asumen una distribución de Poisson³, es decir, modelos lineales generalizados con distribución de Poisson que utilizan la función de nexo canónica⁴. Los datos se trabajan agregados, estratificados por grupos de edad y año de defunción⁵.

Así, para describir la evolución se ha ajustado el modelo:

$$M_{e,t} \sim \text{Poisson}(E[M_{e,t}])$$

$$\ln(E[M_{e,t}]) = \alpha_0 + \alpha_e + \beta_1 t + \beta_2 t^2 + \gamma_e t + \eta_e t^2 + \ln(P_{e,t})_{e,t}$$

donde $M_{e,t}$ representa el número de muertos en la edad e y período o año de defunción t ; $E[M_{e,t}]$ es el número de muertos esperados en la edad e y período t ; $P_{e,t}$ la población en riesgo en la edad e y período t ; α_0 es la constante del modelo; α_e son los coeficientes de la edad (variable indicadora); β_1 es el coeficiente del período; β_2 es el coeficiente del término cuadrático del período;

γ_e es la interacción entre la edad y el período, y η_e es la interacción entre la edad y el término cuadrático del período. El término $\ln(P_{e,t})$ se llama *offset*. Al introducirlo indirectamente estamos modelando el valor esperado de la tasa de mortalidad, $m_{e,t} = E[M_{e,t}]/P_{e,t}$.

El modelo resultante de ajustar una serie de tasas de mortalidad por una causa determinada permitirá describir y entender mejor su evolución. A continuación se describe la interpretación de los posibles submodelos:

– Modelo (a): $\ln(m_{e,t}) = \alpha_0$. Cuando se selecciona este modelo, significa que el valor esperado de la tasa de mortalidad se ha mantenido constante con el tiempo y ésta ha sido la misma para todos los grupos de edad.

– Modelo (b): $\ln(m_{e,t}) = \alpha_0 + \alpha_e$. Cuando se selecciona este modelo, significa que el valor esperado de la tasa de mortalidad se ha mantenido constante con el tiempo, pero que ésta es distinta para cada grupo de edad.

– Modelo (c): $\ln(m_{e,t}) = \alpha_0 + \alpha_e + \beta_1 t$. Cuando se selecciona este modelo, significa que ha habido la misma evolución lineal para el logaritmo del valor esperado de las tasas de mortalidad para todos los grupos de edad. El porcentaje de cambio medio anual de la tasa de mortalidad ajustada por edad (*PCA*) se estima a través de: $PCA = (e^{\beta_1} - 1) \times 100$.

– Modelo (d): $\ln(m_{e,t}) = \alpha_0 + \alpha_e + \beta_1 t + \beta_2 t^2$. Cuando se selecciona este modelo, significa que el logaritmo del valor esperado de las tasas para cada grupo de edad sigue parábolas paralelas.

– Modelo (e): $\ln(m_{e,t}) = \alpha_0 + \alpha_e + \beta_1 t + \gamma_e t$. Cuando se selecciona este modelo, significa que ha habido una evolución lineal del logaritmo del valor esperado de las tasas de mortalidad, pero que ésta ha sido diferente para cada grupo de edad. El *PCA* para cada grupo de edad se estima a través de $PCA = (e^{\beta_1 + \gamma_e} - 1) \times 100$.

– Modelo (f): $\ln(m_{e,t}) = \alpha_0 + \alpha_e + \beta_1 t + \beta_2 t^2 + \gamma_e t$. Cuando se selecciona este modelo, el logaritmo de las tasas sigue una evolución cuadrática que depende del grupo de edad, pero las parábolas de los diferentes grupos de edad comparten la misma curvatura, ya que la segunda derivada que define la curvatura es constante en la edad.

– Modelo (g): $\ln(m_{e,t}) = \alpha_0 + \alpha_e + \beta_1 t + \beta_2 t^2 + \gamma_e t + \eta_e t^2$. Cuando se selecciona este modelo, el logaritmo de la tasa de mortalidad sigue una evolución parabólica diferente para cada grupo de edad.

Para los modelos (d), (f) y (g), el año en que se alcanza la tasa máxima o mínima en cada grupo de edad, se puede calcular, derivando respecto a t e igualando a cero, a través de:

$$t_{e,(máx,min)} = - \frac{\beta_1 + \gamma_e}{2(\beta_2 + \eta_e)}$$

donde la tasa será máxima o mínima, dependiendo del signo del coeficiente de β_2 , y la tasa de mortalidad en

este punto para cada grupo de edad, se calcula a través de:

$$m_{e,(max,min)} = e^{(\alpha_0 + \alpha_1 t - \frac{(\beta_1 + \gamma_1 t^2)}{4(\beta_2 + \eta_2)})} \times 10^5$$

donde para el modelo (d), $\gamma_e = \eta_e = 0$, mientras que para el modelo (f), $\eta_e = 0$.

Para valorar la bondad del ajuste y seleccionar el modelo, se ha realizado el test de la razón de verosimilitudes para los diferentes modelos anidados, se han analizado los residuos de *deviance*, y se ha estimado el parámetro de escala, ϕ , como el cociente entre la *deviance* y los grados de libertad, descartando todos los modelos para los que se observara el fenómeno de la sobredispersión⁶.

Resultados

La primera impresión visual del gráfico de la evolución de la mortalidad por cáncer de mama (fig. 1) ya sugiere que el efecto del año de defunción no ha sido lineal, sino más bien cuadrático. La pregunta es si la evolución es similar por grupos de edad, o no. Al observar el gráfico de las tasas específicas por edad (fig. 2), está claro que hay un gradiente de las tasas con la edad; a mayor edad, tasas de mortalidad más altas. Lo que no se observa tan claramente, es si hay un efecto cuadrático o no, dado que esto es difícil de valorar gráficamente.

Para decidir cuál de los posibles submodelos descritos en el apartado de métodos es el mejor, los contrastamos todos. Los resultados se presentan en la tabla 1. Observamos que al realizar el test de la razón de verosimilitudes entre el modelo completo (g) y el modelo (f) obtenemos un valor de $p = 0,46018$; por tanto, se puede asumir que los modelos no son estadísticamente diferentes y, por consiguiente, seleccionamos el modelo (f) por ser más parsimonioso. Con este mismo criterio, descartamos los modelos restantes, al obtener valores de p muy pequeños en los respectivos contrastes. El parámetro de escala más próximo a 1 se obtiene con el modelo (f) ($\phi = 1,08$), lo que refuerza este modelo como el más adecuado.

El modelo ajustado nos indica que una parábola con la misma curvatura, pero con un valor extremo distinto para cada grupo de edad, explica la evolución de la mortalidad por cáncer de mama.

Como validación del modelo ajustado, se evaluaron los residuos *deviance* estandarizados en función del valor previsto para las muertes, y no se observó ningún patrón ni ningún valor anómalo, dado que la mayoría de sus valores estaban entre -2 y 2 . También se exploró la posibilidad de que los residuos estuvieran autocorrelacionados, pero ésta se descartó, lo que indica que el modelo (f) seleccionado es adecuado a la hora

Figura 1. Tasas ajustadas, según la población de Cataluña del censo de 1991 como referencia, de mortalidad por cáncer de mama (Cataluña, 1986-2000).

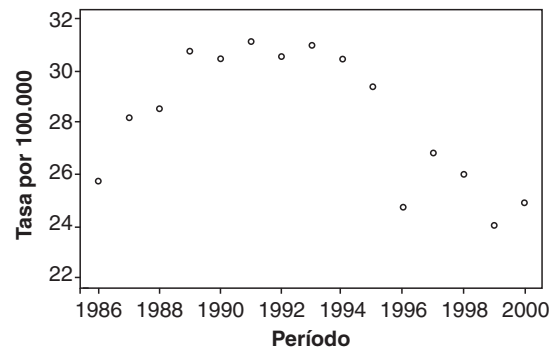


Figura 2. Tasas de mortalidad específicas por edad, observadas y estimadas por el modelo ajustado (f). Cáncer de mama (Cataluña, 1986-2000).

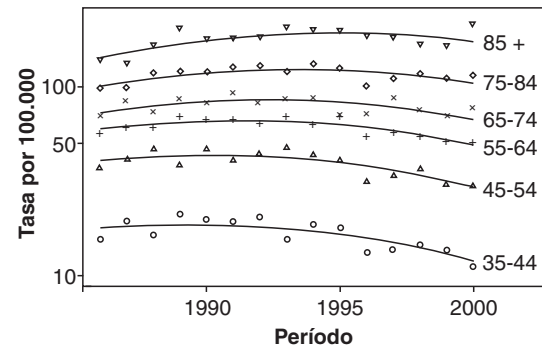


Tabla 1. Test de la razón de verosimilitudes para los diferentes modelos anidados. Mortalidad por cáncer de mama en Cataluña (1986-2000)

Modelo	p						ϕ^*
	(b)*	(c)*	(d)*	(e)*	(f)*	(g)*	
(a)	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	69,05
(b)		0,00001	0,00000	0,00000	0,00000	0,00000	2,31
(c)			0,00000	0,00000	0,00000	0,00000	2,10
(d)					0,00000	0,00000	1,50
(e)					0,00000	0,00000	1,80
(f)						0,46018	1,08
(g)							1,09

ϕ^* : parámetro de escala; (g)*: contraste con el modelo (g); (f)*: contraste con el modelo (f); (e)*: contraste con el modelo (e); (d)*: contraste con el modelo (d); (c)*: contraste con el modelo (c); (b)*: contraste con el modelo (b)

de establecer la evolución temporal de las tasas de mortalidad por cáncer de mama.

En la tabla 2 se presenta la tasa de mortalidad prevista en función del tiempo, para los distintos grupos de edad. A partir de los coeficientes del modelo ajustado estimamos el año en que se alcanza la tasa máxima para cada grupo de edad, así como el valor de la tasa estimada por el modelo en este punto, utilizando las expresiones descritas en el apartado de material y métodos. De dichos resultados destaca que el año en que se logra la tasa máxima varía gradualmente con la edad, de manera que el inicio en el descenso de la mortalidad por cáncer de mama se observa de forma más tardía a medida que se incrementa la edad.

Finalmente, en la figura 2 se representan para cada grupo de edad las tasas de mortalidad observadas, junto con la línea de la mortalidad estimada por el modelo (f), con el fin de ilustrar la versatilidad del modelo a la hora de describir de forma simultánea los patrones de evolución de todos los grupos de edad considerados.

Discusión

El objetivo de este trabajo era ejemplificar el uso de la metodología para analizar la evolución temporal de la mortalidad, basada en el ajuste de modelos lineales generalizados en el caso de series cortas, en el que un modelo de regresión de Poisson basta para capturar la dependencia entre observaciones consecutivas. La representación gráfica de las tasas de mortalidad por cáncer de mama ajustadas por edad, y de las tasas específicas por edad, no desvelaba con claridad si el patrón de evolución por edades es similar al conjunto de todas las edades. Sin embargo, el modelo lineal generalizado, ajustado a los datos observados, nos permite constatar que la evolución experimentada por la mortalidad por cáncer no ha sido lineal sino cuadrática y que, además, el patrón de evolución de todos los grupos de edad ha sido el mismo, si bien trasladado en el tiempo.

En este trabajo se presenta una serie de modelos muy sencillos, compatibles con unos patrones de evolución temporal determinados, que responden a unas

características genéricas que se han considerado frecuentes, o comunes, a diversas enfermedades o series de datos sanitarios. Sin embargo, se puede profundizar mucho más y elaborar modelos teóricos en función de otros parámetros, de acuerdo con los criterios que, según el conocimiento de la enfermedad o la observación previa de los datos empíricos, podamos establecer para cada caso. En este sentido, encontramos un excelente ejemplo en el trabajo de Alberdi y Díaz⁷, donde se parametriza el modelo incorporando una parte tendencial y otra cíclica, aprovechando para ésta las propiedades de las funciones trigonométricas, lo que permite ajustar un modelo donde *a priori* la relación entre las variables explicativas y la respuesta no era lineal.

El método presentado aquí difiere sobre otros procedimientos estadísticos también utilizados en estudios de series temporales. Por ejemplo, recientemente en la bibliografía han aparecido análisis de evolución temporal que utilizan la regresión *joinpoint*⁸ y segmentos de recta, estimando de forma simultánea los parámetros de las rectas y los puntos de cambio de éstas. Habitualmente, estos análisis se realizan sobre el logaritmo de las tasas estandarizadas utilizando modelos de regresión lineal normal por trozos. Un buen ejemplo de la regresión *joinpoint* lo podemos encontrar en el trabajo de Fernández et al⁹, donde se analizan las diferentes causas de mortalidad en Cataluña. En concreto, para el cáncer de mama, los autores identifican un cambio en la tendencia en el año 1992, resultado acorde con nuestro análisis; sin embargo, a partir de su análisis, no se puede discriminar si dicho patrón es similar para diferentes grupos de edad o no.

Los modelos utilizados en nuestra propuesta utilizan la distribución de Poisson, que es más natural que suponer normalidad, dado que el número de muertos es una variable discreta y no continua, si bien en la regresión *joinpoint* se puede abordar el caso de la Poisson aplicando mínimos cuadrados ponderados. Otra diferencia de nuestros modelos respecto a la regresión *joinpoint*, es que éstos utilizan menos parámetros, lo que facilita el estudio diferenciado por grupos de edad, si bien con la regresión *joinpoint* se podría efectuar un análisis estratificado por edad. Por último, obsérvese que los modelos que hemos utilizado capturan mejor el hecho de

Tabla 2. Modelo (f) ajustado en función de la edad, estimación del año en que se logra la tasa máxima por grupo de edad ($t_{e,tm\acute{a}x}$) y las tasas máximas estimadas ($m_{e,tm\acute{a}x}$). Mortalidad por cáncer de mama en mujeres (Cataluña, 1986-2000)

Grupo de edad (años)	Modelo ajustado	$t_{e,tm\acute{a}x}$	$m_{e,tm\acute{a}x}$
35-44	$m_{1,t} = \exp(-15227,58 + 15,3027t - 0,0038467t^2)$	1989	18,824
45-54	$m_{2,t} = \exp(-15227,58 - 12,5651 + [15,3027 + 0,006738]t - 0,0038467t^2)$	1990	43,591
55-64	$m_{3,t} = \exp(-15227,58 - 30,9297 + [15,3027 + 0,0016171]t - 0,0038467t^2)$	1991	65,861
65-74	$m_{4,t} = \exp(-15227,58 - 49,2184 + [15,3027 + 0,025480]t - 0,0038467t^2)$	1992	84,836
75-84	$m_{5,t} = \exp(-15227,58 - 62,9811 + [15,3027 + 0,032576]t - 0,0038467t^2)$	1993	123,923
> 84	$m_{6,t} = \exp(-15227,58 - 85,6324 + [15,3027 + 0,044156]t - 0,0038467t^2)$	1994	192,248

que la variación en las tasas de mortalidad no suelen ser fruto de una intervención puntual, como implícitamente suponen los modelos *joinpoint*, sino que son consecuencia de un gran número de mejoras paulatinas del sistema sanitario. Sin embargo, la regresión *joinpoint* puede ser más eficaz para identificar los cambios bruscos en la tendencia, como el ocurrido en la mortalidad por sida a partir de la aparición de las terapias altamente activas.

Otras aproximaciones intentan diferenciar en el propio modelo los efectos etiológicos y de exposición o avances terapéuticos, introduciendo la variable cohorte (año de nacimiento)^{1,10-14}. Una buena aplicación de estos modelos, llamados edad-período-cohorte, la encontramos en el trabajo de López-Abente et al¹⁰. No obstante, estos modelos, a pesar de que hace más de una década que se utilizan, aún hoy son objeto de debate por la dependencia lineal de las tres variables y la consiguiente dificultad de diferenciar los efectos.

Últimamente han aparecido trabajos¹⁵⁻¹⁸ en los que se utilizan modelos aditivos generalizados, regresión polinómica local, regresión por *splines* o modelos polinómicos fraccionales, entre otros, para modelar la no linealidad inherente en algunas relaciones estadísticas. Dichos modelos están especialmente indicados para el análisis descriptivo y exploratorio de datos, y han mostrado ser eficientes en cuanto al control de la confusión; en cambio, su interpretación en algunos casos puede ser más compleja.

Los resultados de la evolución de la mortalidad por cáncer de mama muestran un incremento significativo en la década de los ochenta, así como una reducción en los últimos años, compatible con las tendencias observadas en otros estudios de nuestro entorno^{9,10}. El año de cambio ha sido distinto entre los grupos de edad. No es raro pensar que el efecto beneficioso de las intervenciones sanitarias, terapéuticas y asistenciales afectan de forma diferente a los distintos grupos de edad y, en particular, que los primeros beneficiados sean los grupos de edad jóvenes.

Agradecimientos

A los miembros del Registro de Mortalidad de Cataluña por su inestimable colaboración.

El presente trabajo ha contado con la financiación parcial del ISCIII (Red de Centros RECESP C03/09).

Bibliografía

- Banegas JR, Alberdi JC, Rodríguez-Artalejo F, Domínguez V, Del Rey Calero J. Análisis edad-período-cohorte de la mortalidad cardiopatía isquémica en España 1965-1985. *Gac Sanit.* 1992;6:97-104.
- Pérez-Hoyos S, Fayos G. Evolució de la mortalitat per suïcidi al País Valencià 1976-1990. *Gac Sanit.* 1996;10:4-11.
- Frome EL. The analysis of rates using Poisson regression models. *Biometrics.* 1983;39:655-74.
- McCullagh P, Nelder JA. *Generalized Linear Models.* London: Chapman and Hall; 1989.
- Breslow NE, Day NE. *Statistical methods in cancer research. Vol II. The design and analysis of cohort studies.* Lyon: IARC Scientific Publications; 1987.
- Breslow NE. Extra-Poisson variation in log-linear models. *Appl Statist.* 1984;33:38-44.
- Alberdi JC, Díaz J. Modelización de la mortalidad diaria en la Comunidad Autónoma de Madrid (1986-1991). *Gac Sanit.* 1997;11:9-15.
- Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med.* 2000;19:335-51.
- Fernández E, González JR, Borràs JM, Moreno V, Sánchez V, Peris M. Recent decline in cancer mortality in Catalonia (Spain). A joinpoint regression analysis. *Eur J Cancer.* 2001;37: 2222-8.
- López-Abente G, Pollán M, Aragonés N, Pérez-Gómez B, Llácer A, Pérez J, et al. Tendencias de la mortalidad en España, 1952-1996. Efecto de la edad, de la cohorte de nacimiento y del período de muerte. Madrid: Instituto de Salud Carlos III; 2002.
- Puig X, Gispert R, Puigdefábregas A, Pérez G, Mompert A, Doménech J. Mortalidad por demencias en Cataluña: un problema de salud emergente. *Med Clin (Barc).* 2002;118:455-9.
- Holford TR. Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annu Rev Public Health.* 1991;12:425-57.
- Osmond C, Gardner MJ. Age, period and cohort models applied to cancer mortality rates. *Stat Med.* 1982;1:245-59.
- González JR, Llorca FJ, Moreno V. Algunos aspectos metodológicos sobre los modelos edad-período-cohorte. Aplicación a las tendencias de mortalidad por cáncer. *Gac Sanit.* 2002;16:267-73.
- Iñiguez C, Pérez-Hoyos S, Ballester F, Sáez M. Comparación de dos métodos en el análisis del efecto a corto plazo de la contaminación atmosférica en salud. *Gac Sanit.* 2003;17:283-8.
- Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol.* 1999;28:964-74.
- Wahba G. *Spline Models for Observational Data.* Philadelphia: Society for Industrial and Applied Mathematics; 1990.
- Fan J, Gijbels I. *Local Polynomial Modelling and Its applications.* London: Chapman and Hall; 1996.