

PROTOCOLO *EMECAM*: ANÁLISIS DEL EFECTO A CORTO PLAZO DE LA CONTAMINACIÓN ATMOSFÉRICA SOBRE LA MORTALIDAD *

Santiago Pérez-Hoyos (1), Marc Sáez Zafra, (2), M.^a Antonia Barceló (2), Coldo Cambra (3), Adolfo Figueiras Guzmán (4), José M.^a Ordóñez (5), Francisco Guillén Grima (6), Ricardo Ocaña (7), Juan Bellido (8), Lluís Cirera Suárez (9), Andrés Alonso López (10), Valentín Rodríguez (11), Tomás Alcalá Nalvaiz (12), Ferran Ballester Díez (1) por el Grupo *EMECAM*.

- (1) Institut Valencià d'Estudis en Salut Pública (IVESP). Direcció General de Salut Pública. Generalitat Valenciana.
 (2) Universitat de Girona. Departament d'Economia.
 (3) Departamento de Sanidad del Gobierno Vasco.
 (4) Universidad de Santiago. Facultad de Medicina.
 (5) Direcció General de Prevenció y Promoció de Salud. Comunidad de Madrid.
 (6) Area de Sanidad y Medioambiente. Ayuntamiento de Pamplona.
 (7) Escuela Andaluza de Salud Publica. Granada.
 (8) Centro Salud Pública Area 2. Castelló. Conselleria de Sanitat.
 (9) Centro Area Cartagena. Consejería de Sanidad de la Comunidad de Murcia.
 (10) Departamento de Salud y Consumo. Ayuntamiento de Vitoria-Gasteiz.
 (11) Direcció Regional de Salut Pública. Asturias.
 (12) Direcció General de Salut Pública. Aragón.
 (*) Este trabajo cuenta con una beca del Fondo de Investigaciones Sanitarias (Expediente núm 97/0051).

RESUMEN

El objetivo del presente trabajo es mostrar el protocolo de análisis elaborado dentro del proyecto *EMECAM*, ilustrando su aplicación en el efecto de la contaminación en la mortalidad en la ciudad de Valencia.

Se considera como variable respuesta la mortalidad diaria para todas las causas, excepto las externas. Las variables explicativas son las series diarias de diversos contaminantes (humos negros, SO₂, NO₂, CO, O₃). Como posibles variables de confusión se consideran factores meteorológicos, factores estructurales y casos semanales de gripe.

Se construye un modelo de regresión Poisson para cada una de las cuatro series de mortalidad en dos fases. En la primera se construye un modelo basal con las posibles variables de confusión. En una segunda se incluyen las variables de contaminación o sus retardos, controlando la autocorrelación residual con la inclusión de retardos de mortalidad.

El proceso de construcción del modelo basal sigue el siguiente orden: 1.º) Incluir los términos sinusoidales significativos hasta orden 6. 2.º) Incluir los términos significativos de temperatura o temperatura al cuadrado con sus retardos hasta orden 15. 3.º) Repetir el proceso con la humedad relativa. 4.º) Introducir los términos significativos de años del calendario, tendencia diaria y tendencia al cuadrado. 5.º) Los días de la semana como variables «dummy» se incluyen siempre en el modelo. 6.º) Incluir los días festivos, y de los retardos hasta 15 días de gripe aquellos que fueron significativos. Tras la reevaluación del modelo, se prueba cada uno de los contaminantes y sus retardos hasta orden 5. Se analiza el efecto por semestres incluyendo términos de interacción.

Palabras Clave: Contaminación. Mortalidad. Regresión Poisson. Series temporales.

ABSTRACT

The *EMECAM* Project Protocol: Short-term Effect of Air Pollution on Mortality

The aim of this study is to show the protocol of analysis which was set out as part of the *EMECAM* Project, illustrating the application thereof to the effect of pollution has on the mortality in the city of Valencia.

The response variables considered will be the daily deaths rate resulting from all causes, except external ones. The explicative variables are the daily series of different pollutants (black smoke, SO₂, NO₂, CO, O₃). As possible confusion variables, weather factors, structural factors and weekly cases of flu are taken into account.

A Poisson regression model is built up for each one of the four deaths series in two stages. In the first stage, a baseline model is fitted using the possible confusion variables. In the second stage, the pollution variables or the time lags thereof are included, controlling the residual autocorrelation by including mortality time lags.

The process of fitting the baseline model is as follows: 1) Include the significant sinusoidal terms up to the sixth order. 2) Include the significant temperature or temperature squared terms with the time lags thereof up to the 7th order. 3) Repeat this process with the relative humidity. 4) Add in the significant terms of calendar years, daily tendency and tendency squared. 5) The days of the week as dummy variables are always included in the model. 6) Include the holidays and the significant time lags of up to two weeks of flu. Following the reassessment of the model, each one of the pollutants and the time lags thereof up to the fifth order are proven out. The impact is analyzed by six-month periods, including interaction terms.

Key words: Pollution. Mortality. Poisson regression. Time series.

Correspondencia:

Santiago Pérez Hoyos.
 Institut Valencià d'Estudis en Salut Pública (IVESP).
 Joan de Garay, 21.
 46017 València.
 Correo electrónico: sperez@san.gva.es

INTRODUCCIÓN

El objetivo del proyecto EMECAM (Estudio Multicéntrico Español sobre la relación entre la Contaminación Atmosférica y la Mortalidad) es evaluar el impacto a corto plazo de la contaminación atmosférica sobre la mortalidad diaria por todas las causas, excepto las externas, y causas respiratorias y del aparato circulatorio, en un conjunto de ciudades españolas¹.

Para analizar el efecto a corto plazo de la contaminación atmosférica sobre la mortalidad se ha optado por llevar a cabo un análisis con datos de series temporales. Las consideraciones teóricas sobre la metodología del análisis de datos temporales se pueden encontrar detalladas en la revisión publicada en este número². Tal como se mencionaba en el artículo de presentación del proyecto¹, en un primer momento se adaptó el protocolo del proyecto *AP-HEA*³. Éste incluía, tras un análisis descriptivo, la identificación de un modelo basal por medio de una regresión lineal gaussiana para el logaritmo de las defunciones diarias. El ajuste posterior de las variables de contaminación y las identificadas en la primera fase gaussiana se efectuaba por medio de una regresión de Poisson. Como ya ha sido mencionado, el bajo número de muertes diarias en las ciudades de menor población, da lugar a que, para las causas seleccionadas, se hayan observado numerosos días sin defunciones. Por ello se ha optado por adaptar el protocolo de análisis a las circunstancias peculiares, utilizando modelos de Poisson desde el inicio.

El objetivo del presente texto es mostrar el protocolo de análisis elaborado dentro del proyecto *EMECAM*, ilustrando su aplicación en el efecto de la contaminación por partículas en suspensión (humos negros) en la mortalidad por todas las causas durante 1992-96 en la ciudad de Valencia.

Información

Variable respuesta

Como se ha mencionado en la presentación del proyecto¹, cada centro dispone de cuatro series diarias de mortalidad ocurridas en cada ciudad para las personas residentes en la misma: la mortalidad por todas las causas, exceptuando las externas, para todas las edades y para mayores de 70 años (CIE 9 001-799), la mortalidad por causas cardiovasculares (CIE 9 390-459) y la mortalidad por causas respiratorias (CIE 9 460-519).

Variables explicativas

Las variables explicativas son las series diarias de cada uno de los contaminantes disponibles en cada centro, construidas a partir de las estaciones captadoras consideradas admisibles tras la imputación de datos no registrados¹. Cada serie de datos de cada contaminante, bien en promedio diario o en máximo de 24 horas, ha sido considerada separadamente.

Factores de confusión

Como posibles variables de confusión se consideran la temperatura media diaria y la humedad relativa (factores meteorológicos), el número diario de casos de gripe, calculado a partir del número de casos semanales declarados dividido por 7, para controlar epidemias de gripe, y otras variables de estructura temporal como factores sinusoidales de estacionalidad anual, tendencias a largo plazo, fluctuaciones semanales y anuales, días festivos y eventos inusuales (la huelga médica, los Juegos Olímpicos en Barcelona o la Expo-92 en Sevilla, etc.)

Modelo de análisis

La serie de datos diarios de mortalidad puede considerarse como distribuida aproximadamente como una distribución Pois-

son con sobredispersión⁴ (es decir la varianza es mayor que la media) y, habitualmente, con autocorrelación. La sobredispersión y la estructura de autocorrelación suelen ser consecuencia de factores exógenos, como la temperatura o la estructura estacional, más que de factores propios del número de defunciones. Así, para cada una de las series de mortalidad, cada centro construyó un modelo de regresión de Poisson para explicar las fluctuaciones de la mortalidad. El modelo de regresión de Poisson se construye como:

$$\log E(Y) = \alpha + \gamma C + \sum \beta_i X_i$$

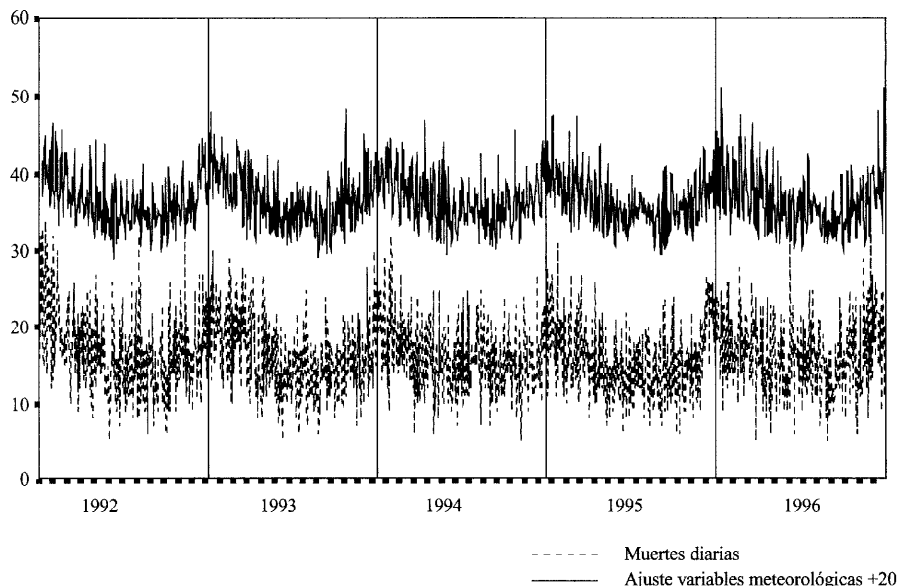
Donde E(Y) es el número esperado de defunciones diario; α es la constante del modelo; γ el efecto de cada contaminante o de sus retardos o de los promedios de los últimos días; C es el contaminante y β el efecto de cada una de las covariables X_i a controlar.

Para especificar el modelo de Poisson anterior se sigue una estrategia parecida a la metodología APHEA³. En primer lugar se identifica un modelo basal para cada una de las causas de muertes, a partir de las posibles variables de confusión entre la relación de la mortalidad y la contaminación. A diferencia de APHEA, este modelo basal se identifica directamente usando la regresión de Poisson con el programa EGRET⁵. Una vez identificado el modelo basal, se procede a extender este modelo a cada uno de los contaminantes y sus retardos. Finalmente, se intenta controlar la autocorrelación residual incluyendo términos autoregresivos de la mortalidad.

Tal como se ha mencionado, se procede a mostrar el método de análisis utilizando como ejemplo la serie de datos de mortalidad por todas las causas en el período 1992-1996. En la figura 1 se muestra esta serie, junto con la de humos negros, apreciándose una aparente correlación entre ambos.

Figura 1

Defunciones diarias por todas las causas menos externas en la ciudad de Valencia 1992-1996 y media diaria de humos negros



Identificación del modelo basal

Para la construcción del modelo basal se introducen progresivamente las variables estructurales y meteorológicas. Las variables en el modelo se incluyen utilizando un criterio estadístico conservador, consistente en la mejoría de la lejanía (*deviance*), determinada por una significación del cociente de verosimilitudes de $p < 0.10$. A continuación se muestra como se ha construido el modelo.

Estacionalidad

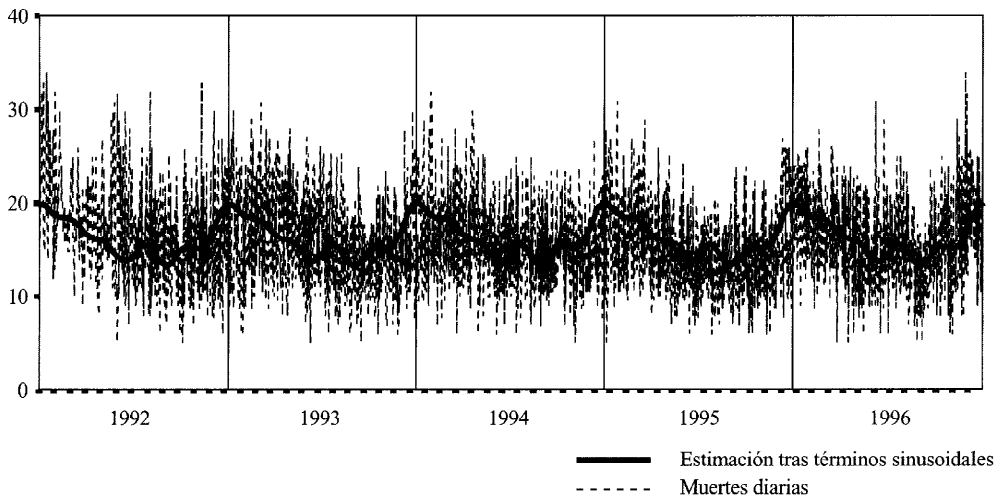
El control de la estacionalidad se lleva a cabo introduciendo términos sinusoidales del tipo $\sin\left(\frac{k2\pi t}{365}\right)$ y $\cos\left(\frac{k2\pi t}{365}\right)$, siendo k un valor de 1 al 6 y $t=1,2,\dots$ el número de

días transcurridos desde el 1 de enero de 1992. Este tipo de términos permite recoger desde ciclos anuales ($k=1$) hasta ciclos bimestrales ($k=6$). Para seleccionar los términos sinusoidales del modelo se decide incluirlos sucesivamente por parejas de seno y coseno, empezando por $k=1$. Para incluir una pareja se exige una mejora en la lejanía de $p < 0.10$. De este modo se incluyen los términos de orden 1, 2, 3, 4 y 5. En la figura 2 se observa como queda la serie tras ajustar los términos sinusoidales.

Temperatura

La relación de la temperatura con la mortalidad diaria no tiene por qué ser lineal. Además, el efecto puede ser retardado, ya que la temperatura de un día no tiene necesariamente que estar relacionada con la mortalidad del mismo día, sino que puede

Figura 2
Ajuste de variables sinusoidales para las defunciones diarias por todas las causas menos externas en la ciudad de Valencia 1992-1996



Modelo ajustado con términos sinusoidales $\sin\left(\frac{k2\pi t}{365}\right)$ y $\cos\left(\frac{k2\pi t}{365}\right)$ de orden $k=1, 2, 3, 4$ y 5 .

estarlo con la mortalidad de los días siguientes. Por ello, se calculan los retardos hasta orden 7 de la temperatura media diaria. Así mismo, se calcula la temperatura media diaria al cuadrado y sus correspondientes retardos hasta orden 7. Con ello no es necesario asumir la relación lineal entre la temperatura y la mortalidad, lo que permite otra forma funcional además de controlar el efecto retardado de la relación. Para decidir qué términos se incluyen en el modelo, con el fin de obtener el más parsimonioso posible, se sigue la siguiente estrategia:

En primer lugar se calcula el «*score test*» obtenido al ampliar el modelo con términos sinusoidales para cada una de las variables de temperatura mencionadas. El «*score test*» contrasta, si el coeficiente de regresión de cada una de las variables de temperatura es igual a cero al añadir esta variable al modelo ajustado hasta el momento, sin necesidad de ajustar cada uno de los modelos generados al añadir la variable de temperatura⁶. Aquella variable para la que se obtiene una mayor puntuación (significación) en el «*score test*» se selecciona y se ajusta junto con los términos sinusoidales elegidos. Si el modelo mejora su lejanía en una $p < 0.1$ se procede a extenderlo con un nuevo término de temperatura y se evalúa de nuevo el «*score test*», introduciendo la variable que tiene mejor puntuación. De este modo se sigue un proceso iterativo hasta que la inclusión de un nuevo término no mejore significativamente el modelo, considerando mejora significativa la de la lejanía del modelo con $p < 0.1$. Al efectuar este proceso en el ejemplo, se incluyen los retardos de orden 1, 2 y 7 de la temperatura media y el retardo de orden 1 de la variable temperatura media al cuadrado.

Humedad

En el caso de la humedad relativa, se efectúa un proceso semejante al seguido con las temperaturas. Se calculan los retardos de orden 1 a 7, tanto para la humedad como para su cuadrado. Se sigue el proceso iterati-

vo de ampliación del modelo utilizando el «*score test*» y añadiendo al modelo la variable de humedad de mayor puntuación, hasta que añadir una no mejore el modelo significativamente. Al efectuar el proceso se incluye la humedad del día y los retardos de orden 4 para la humedad y la humedad al cuadrado.

En la figura 3 se observa la serie de mortalidad estudiada tras ajustar los términos sinusoidales y las variables de temperatura y humedad.

Años de estudio

Debido a que el proyecto EMECAM tiene por objeto evaluar la asociación diaria entre mortalidad y contaminación y no otros tipos de relación temporal, y con el fin de controlar posibles diferencias por año se crea una variable indicador para cada año, excepto para el primero, que es seleccionado como referencia, siendo en el ejemplo el año 1992. Estas variables toman el valor 1 cuando las observaciones corresponden a ese mismo año y 0 cuando corresponden a otros años. Se incluyen en el modelo de forma conjunta (en bloque) y sólo si se produce una mejora de la lejanía en $p < 0.1$. En el ejemplo han sido incluidas.

Tendencia

Además de la estacionalidad y diferencias anuales, puede haber una tendencia secular en la serie que no necesariamente tiene que ser lineal. Con el fin de controlarla se crean dos variables. Una es el número de días transcurridos desde el primer día de la serie (tendencia) y otra su cuadrado. Se añade al modelo la variable que entre las dos mejore la lejanía. En nuestro caso ambas variables mejoran el modelo.

Días de la semana

Tanto la mortalidad como la contaminación tienen un patrón semanal. Por ello se crean 6 variables indicadoras, una para cada

Tabla 1

Ajuste del modelo basal de regresión de Poisson para las defunciones diarias por todas las causas menos externas en la ciudad de Valencia 1992-1996

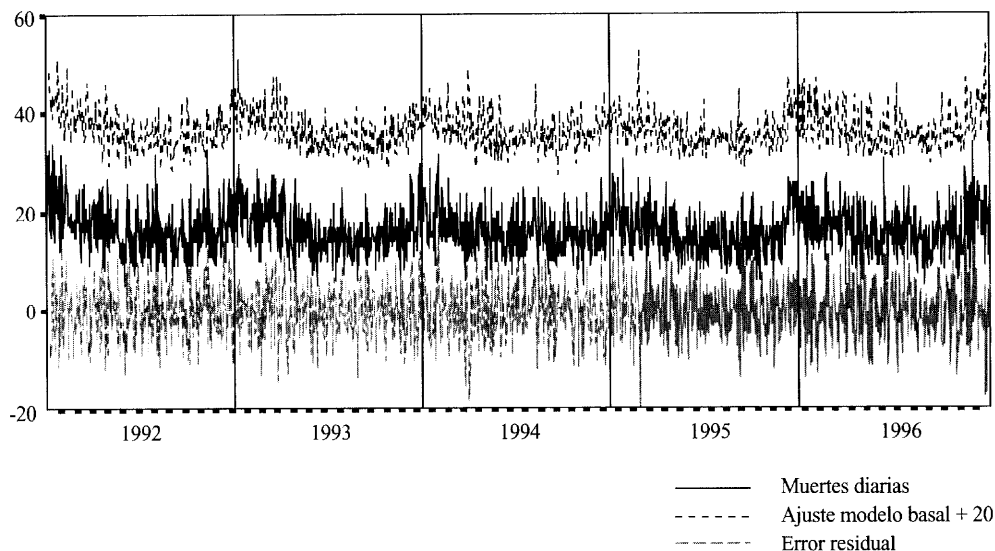
Variables	Modelo Basal			Modelo final con Humos negros		
	Coficiente	Error Estándar	p	Coficiente	Error Estándar	p
Constante	3.441	145	<.001	3.362	.149	<.001
sin1	.03295	.0283	.244	.03582	.0284	.208
cos1	.09121	.0300	.002	.08336	.0303	.006
sin2	.01772	.0160	.269	-.01418	.0161	.380
cos2	.01924	.00931	.039	.01752	.00934	.061
sin3	-.02434	.0113	.032	-.02279	.0113	.045
cos3	.01312	.00879	.136	.01417	.00885	.110
sin4	.01210	.00984	.219	.01217	.00982	.217
cos4	.02407	.00873	.006	.02282	.00874	.009
sin5	-.02050	.00927	.027	-.01901	.00930	.041
cos5	.02542	.00845	.003	.02460	.00848	.004
temperatura (7)	-.008677	.00276	.002	-.00826	.00277	.003
temperatura*temperatura (1)	.001014	.00028	<.001	.0009642	.000281	<.001
temperatura (2)	-.009431	.00362	.009	-.008971	.00364	.017
temperatura (1)	-.01711	.0100	.088	-.01575	.0101	.117
humedad	.0008927	.00046	.052	.0008074	.000464	.082
humedad*humedad (4)	.00007690	.000027	.004	.00008189	.0000271	.003
humedad (4)	-.01084	.00339	.001	-.01144	.00340	<.001
año 93	.1011	.0668	.130	.07863	.0676	.245
año 94	.1914	.125	.126	.1567	.126	.214
año 95	.2257	.182	.215	.1832	.183	.318
año 96	.2821	.239	.239	.2353	.241	.328
tendencia	-.0004849	.000199	.015	-.0003965	.000203	.050
tendencia*tendencia	.0000001547	.000000596	.009	.0000001256	.0000000607	.038
martes	-.01771	.0217	.415	-.02827	.0225	.208
miércoles	.002668	.0217	.902	-.01124	.0228	.622
jueves	-.02718	.0218	.213	-.004263	.0230	.064
viernes	-.02285	.0218	.294	-.03732	.0230	.104
sábado	-.04041	.0219	.064	-.05281	.0226	.020
domingo	-.02223	.0217	.306	-.02551	.0218	.242
gripe	.0004613	.000144	.001	.0004374	.000145	.003
gripe (2)	-.000345	.000153	.025	-.0003386	.000154	.027
gripe (8)	.0001692	.0000942	.072	.0001706	.0000942	.070
Humos negros (1)				.0007017	.000360	.051
Muertes (3)				.002389	.00136	.080
Lejanía 1974.836 con 1779 grados de libertad			Lejanía 1967.773 con 1777 grados de libertad			

El valor entre paréntesis indica el retardo de la variable.

La variable sin k = $\sin\left(\frac{k2\pi t}{365}\right)$ y cos k = $\cos\left(\frac{k2\pi t}{365}\right)$

Figura 4

Ajuste del modelo Poisson para las defunciones diarias por todas las causas menos externas en la ciudad de Valencia 1992-1996



Cambra (Bilbao); M Taracido, JM Barros, I Castro, A Figueiras, A Montes, E Smyth (Vigo); JM Ordóñez, E Aranguez, I Galán, AM Gandarillas (Madrid); I Aguinaga, MY Floristan, F Guillén, MS Laborda, MA Martínez, MT Martínez, PJ Oviedo (Pamplona); A Daponte, R Garrido de la Sierra, JL Gurucelain, P. Gutiérrez, JA Maldonado, JL Martín, JM Mayoral, R Ocaña, J Serrano (Granada); JB Bellido, A Arnedo, F González (Castellón); JJ Guillén, Ll Cirera, L García, E Jiménez, MJ Martínez, S Moreno, C Navarro (Cartagena); MJ Pérez, A Alonso, JJ Estíbalz, MA García-Calabuig, (Vitoria); A Cañada, C Fernández, F Fernández, V García, I Huerta, V Rogríguez (Asturias); F Arribas, M Navarro, C Martos, MJ Rabanaque, E Muniesa, JM Abad, JI Urraca (Zaragoza); y J Sunyer como asesor del proyecto.

BIBLIOGRAFIA

1. Ballester F, Sáez M, Alonso ME, Taracido M, Ordóñez JM, Aguinaga I et al. El proyecto EMECAM: Estudio Multicéntrico Español sobre la relación entre la contaminación atmosférica y la mortalidad. Antecedentes, participantes, objetivos y material y métodos. *Rev Esp Salud Publica* 1999; 73:165-173.
2. Sáez M, Pérez-Hoyos S, Tobias A, Saurina C, Barceló MA, Ballester F Métodos de series temporales en los estudios epidemiológicos sobre contaminación atmosférica. *Rev Esp Salud Publica* 1999; 73:133-143.
3. Katsouyanni K, Schwartz J, Spix C, Touloumi G, Zmirou D, Sanobetti et al. Short term effects of air pollution on health: A European approach using epidemiologic time series data: The APHEA protocol. *J Epidemiol Community Health* 1996;50 (Suppl 1):S12-S18.
4. Schwartz J, Spix C, Touloumi G, Bachárová L, Barumamdzadeh T, leTertre et al. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *J Epidemiol Community Health* 1996;50(Suppl 1):S3-S11.
5. Egret User Manual. Seattle Washington: Statistics and Epidemiology Research Corporation; 1991.
6. Breslow NE, Day NE. *Statistical Methods for Cancer Research Vol II. The design and analysis of cohort studies.* Lyon: IARC; 1987.