

PROCESO DE ANÁLISIS DE DATOS EN UN ESTUDIO MULTICÉNTRICO SOBRE CONTAMINACIÓN ATMOSFÉRICA Y SALUD. PROYECTO EMECAS

M.P. Rodríguez¹, F. Ballester¹, C. Iñiguez¹, S. Pérez-Hoyos¹, M. Sáez², A. Daponte³, J. M. Ordóñez⁴, K. Cambra⁵, J. J. Guillén⁶, Á. Cañada⁷, e I. Aguinaga⁸ en nombre del grupo EMECAS

¹Unidad de Epidemiología y Estadística. Escola Valenciana d'Estudis per a la Salut. Conselleria de Sanitat. Generalitat Valenciana.

²Grup de Recerca en Estadística, Economia Aplicada i Salut (GRECS), Departament d'Economia, Universitat de Girona

³Escuela Andaluza de Salud Pública (EASP). Granada

⁴D. G. Salud Pública. Comunidad de Madrid

⁵D. de Sanidad del Gobierno Vasco, Vitoria

⁶Centro de área de Cartagena. Consejería de sanidad y política social. Cartagena

⁷Departamento Salud Pública. Dirección regional de salud pública.

Consejería de Servicios Sociales. Oviedo

⁸Departamento de Epidemiología. Área de Sanidad y Medio Ambiente.

Ayuntamiento de Pamplona

Objetivos

El proyecto EMECAS (Estudio Multicéntrico de los Efectos a corto plazo de la Contaminación Atmosférica en Salud) estudia la magnitud de la asociación entre contaminación atmosférica y morbilidad circulatoria, respiratoria y mortalidad. Participaron un total de 16 ciudades españolas: Barcelona, Bilbao, Cartagena, Castellón, Granada, Gijón, Huelva, Las Palmas, Madrid, Oviedo, Pamplona, Sevilla, Tenerife, Valencia, Vigo y Zaragoza, suponiendo un total de 10 millones de habitantes. Los centros correspondientes a las ciudades fueron los siguientes:

- Escola Valenciana d'Estudis per a la Salut (EVES) (Centro coordinador).
- Grup de Recerca en Estadística, Economia Aplicada i Salut (GRECS), Departament d'Economia, Universitat de Girona.
- Escuela Andaluza de Salud Pública, Granada.
- D. G. Salud Pública. Comunidad de Madrid.
- Centro de Salud Pública. Castellón.
- D. de Sanidad del Gobierno Vasco, Vitoria.

- D. Medicina Preventiva, Universidad de Santiago de Compostela.
- D. de Sanidad, Bienestar y Trabajo de Aragón, Zaragoza.
- Centro de Área de Salud Pública de Cartagena, Murcia.
- DG de Salud Pública de Asturias, Oviedo.
- Servicio de Salud, Ayuntamiento de Pamplona.
- DG de Salud Pública. Consejería de Sanidad de Canarias, Las Palmas.

Es esencial que en un estudio multicéntrico todos los centros sigan el mismo proceso, tanto en la creación de variables como en la metodología de análisis y tomen decisiones intermedias en base a un mismo criterio, en aras a obtener comparabilidad en los resultados y una adecuada interpretación de los estimadores combinados. En un momento avanzado del proyecto, salieron a la luz razones por las cuales dudar de la validez de los resultados de los modelos GAM obtenidos con el programa S-Plus 2000, dado que la implementación de la función gam en dicho programa, asume unos parámetros de convergencia por

En un momento avanzado del proyecto, salieron a la luz razones por las cuales dudar de la validez de los resultados de los modelos GAM obtenidos con el programa S-Plus 2000, dado que la implementación de la función gam en dicho programa, asume unos parámetros de convergencia por defecto excesivamente laxos (10 iteraciones máximo y precisión 10⁻³) y a problemas en el algoritmo de resolución. En otros estudios de contaminación atmosférica (NMMAPS, APHEA, etc.) reanalizaron los datos para valorar, en qué medida los resultados ya publicados pudieran quedar invalidados por diferir, de manera importante, de los obtenidos con parámetros de convergencia más estrictos. No hubo unanimidad en las conclusiones obtenidas, de modo que, decidimos repetir nuestros análisis e investigar si los resultados eran significativamente distintos. Presentaremos el proceso de análisis llevado a cabo en EMECAS.

Métodos

En la creación de variables, cada centro recogió datos de su ciudad y construyó, en base a unas decisiones previas establecidas, su propia base de datos. Las variables a construir fueron:

- Variables respuesta (indicadoras de salud):
 - *mortalidad*: valores diarios de mortalidad por todas las causas excluyendo externas (mtot), por causas circulatorias (mcir), cardíacas (mcar), cerebrovasculares (mcbs), isquémicas (misq), respiratorias (mres), por todas las causas excepto externas en personas mayores de 70 años (mm70).

- *morbilidad circulatoria*: valores diarios de ingresos urgentes por todas las enfermedades del aparato circulatorio (ecir), enfermedades cardíacas (ecar), cerebrovasculares (ecbs), isquémicas (eisq).
- *morbilidad respiratoria*: valores diarios de ingresos urgentes por todas las enfermedades respiratorias (eres), asma (easm), EPOC (eepo), y por enfermedades digestivas (edig), ésta última se incluyó para ser variable control.

La selección de los casos por enfermedades específicas se realizó siguiendo la clasificación CIE-9.

- Variables indicadoras de contaminación atmosférica:
 - promedios de 24 horas del total de partículas en suspensión (tsp), humos negros (bs), partículas de diámetro inferior a 10μ (pm10), SO₂ (so224, o so2m si el captador era manual), y NO₂ (no224)
 - medias móviles de 8 horas de CO (co8), ozono (o38).
 - máximo de 1 hora de SO₂ (so2a1 sólo para estaciones automáticas), NO₂ (no21), CO (co1), ozono (o31).
- Variables meteorológicas:
 - promedio de temperatura máxima y mínima diaria (temp).
 - promedio de los valores de humedad a las 0, 7, 13, 18h (hum).
 - presión atmosférica (pa).
- Covariables:
 - casos diarios de gripe (infl).
 - Días festivos (hol), días de la semana (wday), días inusuales (unev).
 - olas de calor (hotw).
 - tendencia temporal (trend).

Una vez construidas las bases, se realizó para todas ellas, un descriptivo numérico básico (media, desviación típica, mínimo, máximo, y percentiles 10, 25, 50, 75, 90) estratificando por año, y considerando todo el periodo y separando en semestre frío y cálido. A nivel central se realizó un examen más exhaustivo mediante gráficas de secuencia, y el cálculo y representación gráfica de los valores promedio anuales, mensuales y semanales, por si existiera alguna tendencia temporal, y de las tasas, en las que no existe confusión debida a diferencias en la población.

Para tener en cuenta en los análisis el efecto retardado de la contaminación, se estimaron modelos para los retardos del 0 al 3 de cada indicador de contaminación,

y para los calculados como medias diarias y de 8 horas, se realizaron también los modelos para los promedios de los retardos 0-1 y 2-3. En el caso del ozono sólo se analizó el periodo cálido. La metodología utilizada para la creación de dichos modelos, fueron modelos aditivos generalizados (GAM) de Poisson, controlando previamente por confusión y sobredispersión. Para llevar los análisis a cabo, se empleó el paquete estadístico S-Plus 2000.

Se obtuvo tal cantidad de modelos que para continuar con la fase posterior del estudio, en la que se valoraba la confusión entre contaminantes introduciendo uno más a los modelos ya creados, hubo que restringirse a un subconjunto de ellos, seleccionando los mejores indicadores para valores promedio y máximos de las medidas diarias de los contaminantes. Nos quedamos con los lags que resultaran más positivos y significativos en mayor cantidad.

El análisis de sensibilidad se hizo comparando los resultados obtenidos para las causas de morbilidad circulatoria de 15 de las 16 ciudades participantes. Se repitieron un total de 594 modelos por causa con parámetros de convergencia más estrictos (número máximo de iteraciones 1000, y precisión 10⁻⁸). Se obtuvieron medidas de sesgo para cada causa, tales como el cambio absoluto y porcentual en los coeficientes y errores estándar, tanto para las estimaciones individuales en cada ciudad, como para los valores del análisis combinado. Se representaron los coeficientes y los errores estándar obtenidos para parámetros por defectos versus parámetros más estrictos. Los resultados se presentan como mediana e intervalo de percentiles 5 y 95 (P_{50} , [P_5 , P_{95}]). Nos restringimos a los valores combinados que son los de mayor interés.

Resultados

Los descriptivos de las bases de datos nos permitieron encontrar errores que no fueron detectados previamente. En alguna serie de morbilidad, para alguna ciudad, se observó gracias a la gráfica de secuencia un cambio de nivel brusco de un año para otro; en otros casos no tan llamativos, pero igualmente importantes, se vieron caídas en los valores al final de la serie. Para algún contaminante la gráfica que resultó no era representativa del comportamiento esperado inherente al contaminante, se detectó que el problema venía por una creación errónea del indicador de contaminante. Para la gripe se encontró una ciudad con una tasa excesivamente elevada en comparación al resto. Dichos problemas se comunicaron a los centros correspondientes y se subsanaron en la medida de lo posible, o se procedió a la exclusión eventual o parcial de la serie.

De todos los modelos de un contaminante obtenidos, aquellos con lags 0 y 1 y el promedio 0-1 fueron los que, en general, resultaron más adecuados excepto para el ozono, cuyo efecto fue más retardado, siendo el lag 3 y el promedio 2-3 los elegidos para continuar con modelos de dos contaminantes.

Del estudio de sensibilidad, obtuvimos que para los estimadores combinados, el

cambio porcentual medio fue de un 13.17% [0.71%, 220.31%] para ecir, 12.57% [0.36%, 147.12%] para ecar, 13.34% [0.27%, 359%] para eisq, 16.85% [0.86%, 93.71%] para ecbs. Dichos cambios porcentuales conllevan cambios absolutos no muy importantes. Los percentiles 95 de la diferencia absoluta en los coeficientes fueron 0.0007 y 0.0008 para ecir y ecar respectivamente, y 0.0012 y 0.0036 para ecbs y eisq respectivamente. A pesar de ser un poco mayores en éstas causas, en el 90% de los casos, los valores difieren en la cuarta cifra decimal.

Conclusiones

Un análisis descriptivo exhaustivo permitió detectar ciertas anomalías. Tras su corrección y re-análisis se observaron diferencias en los resultados poniendo de manifiesto la importancia de tener datos correctos y comparables para evitar introducir más variabilidad en los resultados.

Con intención de obtener estimaciones más precisas, creemos que sería mejor realizar análisis con los parámetros más estrictos, no obstante, los pequeños cambios producidos en nuestros resultados no invalidan los obtenidos anteriormente con parámetros por defecto. Las diferencias importantes encontradas en otros estudios, pueden deberse en parte a otras causas, no sólo a los parámetros de convergencia.

Agradecimientos

Este estudio ha contado con financiación del Fondo de Investigaciones Sanitarias del Ministerio de Sanidad (FIS 00/0010).

Referencias

- Ballester, F., Sáez, M., Perez-Hoyos, S., Iñiguez, C., Gandarillas, A., Tobias, A., Bellido, J., Taracido, M., Arribas, F., Daponte, A., Alonso, E., Cañada, A., F., G.-G., Cirera, L., Perez-Boillos, M. J., Saurina, C., Gómez, F. y Tenías, J. M., on behalf of the EMECAM group (2002). The EMECAM project: a multi-center study on air pollution and mortality in Spain. Combined results for particulates and for sulphur dioxide. *Occupational and Environmental Medicine* 59(5):300–308.
- Der-Simonian, R. y Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* 7:177–188.
- Dominici, F., McDermott, A., Zeger, S. L. y Samet, J. M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology* 156:193–203.
- Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., Tertre, A. L., Monopolis, Y., Rossi, G., Zmirou, D., Ballester, F., Boumghar, A., Anderson, H. R., Wojtyniak,

- B., Paldy, A., Braunstein, R., Pekkanen, J., Schindler, C. y Schwartz, J. (2001). Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 european cities within the aphea2 project. *Epidemiology* 12:521–531.
- Sáez, M., Ballester, F., Barceló, M. A., Perez-Hoyos, S., Bellido, J., Tenías, J. M., Ocaña, R., Figueiras, A., Arribas, F., Aragonés, N., Tobias, A., Cirera, L. y Cañada, A. M., on behalf of the EMECAM group (2002). A combined analysis of the short-term effects of photochemical air pollutants on mortality within the EMECAM project. *Environmental Health Perspectives* 110:221–228.