

Comparing meta-analysis and ecological-longitudinal analysis in time-series studies. A case study of the effects of air pollution on mortality in three Spanish cities

M Saez, A Figueiras, F Ballester, S Pérez-Hoyos, R Ocaña, A Tobias

Abstract

Study objective—The objective of this paper is to introduce a different approach, called the ecological-longitudinal, to carrying out pooled analysis in time series ecological studies. Because it gives a larger number of data points and, hence, increases the statistical power of the analysis, this approach, unlike conventional ones, allows the complementation of aspects such as accommodation of random effect models, of lags, of interaction between pollutants and between pollutants and meteorological variables, that are hardly implemented in conventional approaches.

Design—The approach is illustrated by providing quantitative estimates of the short-term effects of air pollution on mortality in three Spanish cities, Barcelona, Valencia and Vigo, for the period 1992–1994. Because the dependent variable was a count, a Poisson generalised linear model was first specified. Several modelling issues are worth mentioning. Firstly, because the relations between mortality and explanatory variables were non-linear, cubic splines were used for covariate control, leading to a generalised additive model, GAM. Secondly, the effects of the predictors on the response were allowed to occur with some lag. Thirdly, the residual autocorrelation, because of imperfect control, was controlled for by means of an autoregressive Poisson GAM. Finally, the longitudinal design demanded the consideration of the existence of individual heterogeneity, requiring the consideration of mixed models.

Main results—The estimates of the relative risks obtained from the individual analyses varied across cities, particularly those associated with sulphur dioxide. The highest relative risks corresponded to black smoke in Valencia. These estimates were higher than those obtained from the ecological-longitudinal analysis. Relative risks estimated from this latter analysis were practically identical across cities, 1.00638 (95% confidence intervals 1.0002, 1.0011) for a black smoke increase of 10 µg/m³ and 1.00415 (95% CI 1.0001, 1.0007) for an increase of 10 µg/m³ of sulphur dioxide. Because the statistical power is higher than in the individual analysis more interactions were statistically significant,

especially those among air pollutants and meteorological variables.

Conclusions—Air pollutant levels were related to mortality in the three cities of the study, Barcelona, Valencia and Vigo. These results were consistent with similar studies in other cities, with other multicentric studies and coherent with both, previous individual, for each city, and multicentric studies for all three cities.

(J Epidemiol Community Health 2001;55:423-432)

Recently there has been an increasing use of quantitative methods that combine information from different studies on the same topic. In this sense, meta-analyses (MA) of published results have become an important part of epidemiological research. MA could be defined as a statistical pooling of results¹ to produce more precise inferences than individual studies and to explore heterogeneity of findings among different analyses. The usual MA approach is to obtain separate estimates from few individual studies by more or less standardised procedures and then to obtain a joint estimate by computing a weighted mean. Since the beginnings of 1990s, however, some limitations of MA²⁻³ have popularised an alternative approach known as pooled analysis²⁻⁵ (PA) of primary data,⁶ as well as Type III meta-analysis,⁷ meta-analysis of individual data⁸ or reanalysis.⁹ The term “pooled”, however, is not used by all authors in exactly the same sense. The paper by Blettner *et al.*,⁷ for instance, does say “pooling” but the methods suggested are in fact MA. Unlike MA, which uses the results of the individual studies, usually the estimates of the parameters of interest, PA combines raw (in the sense of primary) data from individual studies. This leads to relatively larger sample sizes. An example could clarify this argument. Suppose, for instance, that we are interested in combining 10 relative risks estimated from 10 time series individual studies, each one consisting of one year of daily observations. Whereas the sample size in MA would be 10, there would be 3650 (10 × 365) observations in the PA. As a consequence, the control of confounding and the detection of possible interactions are easily drawn from PA. Furthermore, because it is easier and more stable from a larger dataset, the derivation of more valid and precise conclusions regarding a particular exposure-disease relation is easier than from alternative MA.²⁻³

Statistics and Econometrics, Department of Economics, University of Girona, Spain
M Saez

Preventive Medicine and Public Health, Faculty of Medicine, University of Santiago de Compostela, Spain
A Figueiras

Institut Valencià d'Estudis en Salut Pública, IVESP, València, Spain
F Ballester
S Pérez-Hoyos

Escuela Andaluza de Salud Pública, Granada, Spain
R Ocaña

Clinical Epidemiology and Public Health Service, Hospital de la Santa Creu i de Sant Pau, Barcelona, Spain
A Tobias

Correspondence to: Dr Saez, Department of Economics, University of Girona, Campus de Montilivi 17071 Girona, Spain (msaez@gnomics.udg.es)

Accepted for publication 16 November 2000

The combination of information is particularly relevant in the case of time series ecological studies, in particular those on the short-term effects of air pollution on health.¹⁰⁻¹³ In such studies, the adverse effects on health (specifically mortality) of the suspected risk factors (usually air pollutants), although important, are usually small. As a presumable consequence of such moderate sizes, the results of individual studies to be combined may present some inconsistencies. For all these reasons, several multi-centre collaborative studies were launched over the past few years.^{14 15} In these studies, each centre (typically a city) analysed their data from an individual level (using standardising data collection procedures, definition of variables, questions and hypotheses for the individual studies) and then, the coordinator centre meta-analysed individual results.^{10 12 13} As one can see, the methods used by these multicentric studies are closer to MA than to PA, mainly because statistical analyses are not carried out on individual data. Possible reasons for this are that PAs are more expensive, time consuming and require more refined methodological and statistical methods than the alternative MA. Theoretically, however, a PA would be more appropriate as the effect of rare exposures could be examined. The control of confounding could be more precise and more interactions could be detected (for instance between air pollutants and meteorological variables) as well.

The objective of this paper is to introduce a different approach, which could be called ecological-longitudinal, in order to carry out PA in time series ecological studies. Because it gives a larger number of data points and, hence, increases the statistical power of the analysis, our approach, unlike conventional approaches, allows the complementation of aspects such as accommodation of random effect models, of lags, of interaction between pollutants and between pollutants and meteorological variables, that are not routinely complemented (although they should be implemented) in regular time series analysis and derived MA. The approach could be called longitudinal, because it is considered that, in our design, a given sample of "individuals"—that is, three Spanish cities, Barcelona, Valencia and Vigo—were treated as "subjects" observed over time. The label ecological refers to the ecological nature of the relations—that is, we estimate the effect of an exposure on a group, not on an individual basis. The new approach was illustrated by providing quantitative estimates of the short-term effects of air pollution on mortality in three of the Spanish cities participants in the EMECAM study,¹⁵ Barcelona, Valencia and Vigo, for the period 1992–1994. Furthermore, the results of the ecological-longitudinal approach were compared with those obtained by two complementary MA.

Methods

Our design is ecological, involving three Spanish cities, Barcelona (1 643 545 inhabitants), Valencia (749 796) and Vigo (274 574).

Barcelona and Valencia are located on the Mediterranean coast and Vigo on the Atlantic coast. Valencia is located on the central Mediterranean coast of Spain, Barcelona is on the northern Mediterranean coast and Vigo is a northern Atlantic coast city.

We carried out two types of analysis: individual analyses, in which cities were analysed separately; and combined analyses. The latter includes an ecological-longitudinal analysis, where the cities were considered as individuals measured over time, and two multicentric MA,^{10 12 13} one in which we completely standardised explanatory variables and functional forms in each one of the individual models and another where we permitted more flexibility in controlling (this is actually the approach followed by multicentric studies like APHEA¹⁴ or EMECAM¹⁵).

INDIVIDUAL ANALYSES

The response—that is, the dependent variable—is the daily total mortality excluding external causes (ICD-9 001–799) for each one of the cities—that is, Barcelona, Valencia and Vigo—during the period 1 January 1992 to 31 December 1994. Because on any given day only a small portion of the population dies and because of the number of deaths is a count, the daily number of deaths was considered a Poisson random variable.¹⁶ The Poisson process is not stationary over time, that is, the expected number of deaths is not constant but varies with time varying predictor variables. In this sense we can assume that, once predictors such as season, influenza epidemics and alike factors are controlled for, the daily number of deaths mainly depends on the average daily levels of two air pollutants, black smoke and sulphur dioxide.

It is important to note that the method to determine the levels of sulphur dioxide in Valencia and Vigo was that called the thornine technique (spectro-photometric method) from the local manual air pollution monitoring network, whereas in Barcelona it was an automatic network. Additional details can be found elsewhere.¹⁵

As is known, a basic issue in modelling is to control properly for potential confounding. We control for two observed confounders: (a) meteorological variables, average daily weather temperature and relative humidity in particular, in order to control for the effects of short-term variations in weather on mortality; and (b) the average daily number of influenza cases. Daily averaging was necessary in this latter case because influenza records were actually obtained weekly. Additional details on data and their sources can be found elsewhere.¹⁵

Because of high collinearity between air pollutants, two different regressions were specified in the individual analyses. Initially, the regressions were formulated following generalised linear models, GLM,¹⁷ with a log link, Poisson error and a over-dispersed variance.

Several modelling issues are worth mentioning now. Firstly, the relations between mortality and explanatory variables are usually non-linear. For instance, the dependence of mortality on temperature is usually U or V

shaped.^{18 19} In our case, the relation for all three cities looks like an imperfect V with a more pronounced left wind (see fig 1). In the absence of any hypothesis about the precise form of the relations, a flexible approach to covariate control is appropriate. In this sense we used generalised additive models, GAM.²⁰

The GAM extends the GLM by fitting non-parametric functions to estimate the relations between the response and the predictors. As these functions are unknown infinite dimensional parameters, we consider estimating them

by using natural cubic splines.²¹ A spline with k degrees of freedom (or, technically, with $k+2$ knots) for a particular explanatory variable would be similar to introducing k dummy variables for the covariate in the model, each one corresponding to an interval of the same size k .²² For the observed confounders we tried initially four degrees of freedom (equivalent to divide each variable into its four quartiles) although the amount of smoothing in the splines (that is, the degrees of freedom) was finally decided by means of the Akaike information criterion, AIC, corrected for non-parametric models.²³

Of course not all the relations between response and explanatory variables have to be non-linear. In this sense, we use approximate partial tests for the importance of the smooth for each term in the model.²⁰ For a single variable in the model, this would be equivalent to testing for a difference between a linear fit and a smooth fit that includes a linear term along with the smooth term.

Other unobserved confounders but with a systematic variation in time were controlled by means of a cubic spline of time ($t=1,2,3,\dots,1096$) with (initially) 36 degrees of freedom, that is the total number of months between 1992 and 1994.²² Note that this would be similar to introducing 36 dummy variables, each one corresponding to a different month between 1992 and 1994. In this way we controlled for long wavelength patterns—that is, trend and seasonality. However, short wavelengths, say lower than a month, were not controlled. For this reason we included calendar specific days in the GAM.

The second modelling issue we want to mention is the temporal behaviour of the relations. The effects of the predictors on the response may be immediate but, from a physiological point of view¹⁰⁻¹⁵ it is more reasonable to assume that they occur with some lag. Furthermore, because we have time series for all the variables, it is convenient to consider the existence of a very probable period of latency. One problem is how far back to go in exploring the lags. Too many lags could lead to identifying relations that have actually occurred by chance. Because we are trying to analyse short-term relations, restrictions of a week or less seem the most reasonable strategy. In this sense we allowed up to five lags for the air pollutants, up to seven lags for the meteorological variables and a moving average of 15 terms for influenza (it was equivalent to introducing up to 15 lags).

Thirdly, even controlling for unobserved confounders, for the temporal behaviour of the relations and allowing flexible smoothing functions it is possible that some residual autocorrelation remains after the estimation. This means imperfect control (mainly of short wavelengths lower than a month and other cyclic components with irregular period and amplitude²⁴). This autocorrelation was controlled by introducing up to six lags of the response, leading to an autoregressive Poisson GAM.²⁵⁻²⁷

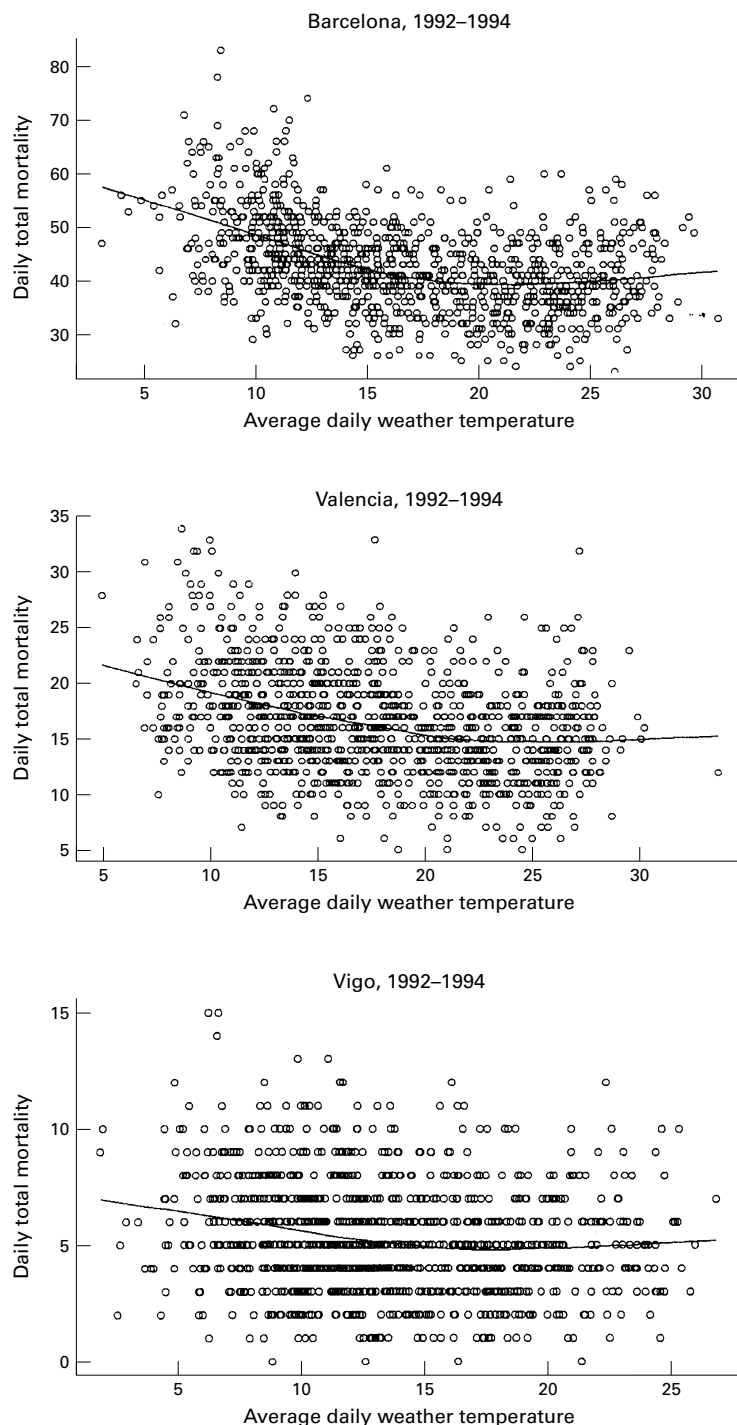


Figure 1 Relation between average daily weather temperature and daily total mortality in three Spanish cities: Barcelona, Valencia and Vigo, 1992-1994.

In all the individual models we allowed for interactions between weather temperature and humidity and between air pollutants and meteorological variables. For the choice of the final models (in particular, the number of lags of the explanatory variables) we also used the corrected AIC.²³ The individual autoregressive Poisson GAMs were estimated by the local scoring algorithm, which iteratively fits weighted additive models by backfitting.^{20 28 29} The backfitting algorithm is a Gauss-Seidel method for fitting additive models, by iteratively smoothing partial residuals. The algorithm separates the parametric from the non-parametric part of the fit, and fits the parametric part using weighted linear least squares within the backfitting algorithm.

COMBINED ANALYSES

Ecological-longitudinal analysis

Our fourth methodological issue named our approach. We have considered that in our design three “individuals”—that is, Barcelona, Valencia and Vigo—were observed on a daily basis for three years (1992–1994). It is important to point out that we assume that there was a highly consistent association between mortality and air pollution from location to another. Of course, cities are not actually individuals and instead of individual measures we have aggregated—that is, ecological measures. However, we called our approach ecological-longitudinal because cities were treated as individuals observed over time.

Our approach demanded the consideration of the existence of individual heterogeneity. The typical assumption that the response was generated by a parametric distribution function—that is, Poisson—identical for all individuals is not a realistic one in longitudinal designs. Ignoring such parameter heterogeneity among units leads to inconsistent and inefficient estimates of interesting parameters. For all these reasons, we estimated mixed models, Autoregressive Poisson GAMM, with fixed effects—that is, air pollutants—observed and unobserved confounders, and random effects. We introduced random intercepts and also random slopes for air pollutants and observed confounders, because we assume that not only the basal mortality but also the effects of predictors varied between cities.

$$\log(\mu_{it}) = \beta_{0i} + \beta_{1i}Z_{it} + \alpha X_{it}$$

$$\beta_{0i} = \beta_0 + b_{0i}$$

$$\beta_{1i} = \beta_1 + b_{1i}$$

where i denotes cities (i =Barcelona, Valencia and Vigo); t denotes time (t =1,...,1096; being 1=1 January 1992 and 1096=31 December 1994); μ_{it} = $E(y_{it} | Z_{it}, X_{it}, \beta_{0i}, \beta_{1i})$; y_{it} is the value of the response variable (total mortality) for the city i and the day t ; Z denotes air pollutants and observed confounders; X unobserved confounders; β_0 is a fixed intercept; and α and β_1 are fixed slopes.

In the ecological-longitudinal case we cannot assume that unobserved confounders but with a systematic variation in time—that is,

long wave trends—act exactly in the same way in the three cities. For this reason, rather than using only one cubic spline of time, as in the individual case, we controlled for this confounding by means of three linear trends, one for each city, each one equal to t_i , $t_i=1,2,3,\dots,1096$ in the corresponding city i and equal to 0 otherwise. As in the individual case, cubic splines of these variables with (initially) 36 degrees of freedom were tried.

Random effects, b_0 and b_1 , were assumed to be normally distributed with a zero mean and an unstructured covariance matrix G , that is we allowed correlation between random effects and a different variance for each one of them. In the case, for instance, of a random intercept, b_0 , and two random slopes, b_1 and b_2 ,

$$G = \begin{bmatrix} b_0 & b_1 & b_2 \\ \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} \\ \sigma_{02} & \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{matrix} b_0 \\ b_1 \\ b_2 \end{matrix}$$

where σ^2 denotes variances (for instance, σ_0^2 is equal to $\text{Var}(b_0)$) and σ covariances (for instance σ_{12} is equal to $\text{Cov}(b_1, b_2)$).

The models were estimated by penalised quasi likelihood, PQL.^{26 30 31} PQL maximises a quasi-likelihood by iteratively analysing a linearised pseudo variable using a weighted normal mixed model. It is worth mentioning that because the variance of the pseudo variable was estimated and not known there could be a small amount of downward bias in the PQL estimate of the variance of the fixed effects.

All the computations were carried out in S-Plus 4.5²⁹ where a macro for PQL estimation was programmed and is available from the authors.

Multicentric MA

As we mentioned above, two types of multicentric (that is, APHEA type) MA were carried out. In the first one (“flexible” hereafter), we permit in each location, the individual model contained its own number of lags for all the explanatory variables, air pollutants in particular, and the functional forms also could differed between individual models. In the second one (“standardised” hereafter), all the individual models contained the same explanatory variables, the same number of lags associated with these covariates, and also share exactly the same functional forms. In both cases combined estimates of the effects of air pollutants on daily total mortality were obtained as a weighted average, in which the weights were proportional to the inverse of the local variances, of the individual estimates, using the Woolf and DerSimonian-Laird approach^{32 33} (also called “random effects” MA).

It is important to point out that in both, ecological-longitudinal and in multicentric analyses, there was a previous protocol¹⁵ of standardisation of the raw data that are being collected at the local level (exposure variables, confounders, effect variables, and alike), and of

Table 1 Demographic data, air pollution levels (24 hour average in $\mu\text{g}/\text{m}^3$), temperature (24 hour average in $^\circ\text{C}$) and relative humidity (24 hour average in %) in Barcelona, Valencia and Vigo, 1992–1994

City	Population			Air pollutants		Meteorological variables			
	Total	≥ 70 years (%)	Total daily mortality count mean (SD)	Black smoke mean (SD)	Sulphur dioxide mean (SD)	Temperature (SD)		Humidity (SD)	
						Cold*	Warm*	Cold*	Warm*
Barcelona	1 643 545	191 724 (11.7)	43.6 (8.6)	43.528 (19.312)	56.139 (48.666)	11.7 (2.7)	21.0 (4.2)	71.8 (4.9)	71.3 (5.7)
Valencia	749 796	75 256 (10.0)	16.4 (4.8)	48.801 (22.468)	34.483 (12.340)	14.3 (3.0)	22.9 (3.4)	61.0 (14.4)	65.3 (12.8)
Vigo	274 574	19 496 (7.1)	5.3 (2.4)	102.571 (40.281)	26.693 (23.357)	10.1 (2.8)	16.9 (3.7)	79.5 (11.7)	75.1 (12.3)

*Cold: November to April, warm May to October.

Table 2 Results of the (standardised) individual analysis for each city, Barcelona, Valencia and Vigo, 1992–1994

City	Black smoke*		Sulphur dioxide†	
	β	95% CI	β	95% CI
Barcelona	0.00069	0.00014, 0.00124	0.00028	0.00008, 0.00049
Valencia	0.00118	0.00037, 0.00199	-0.00129	-0.00259, -0.00001
Vigo	0.00068	0.00009, 0.00128	0.00102	-0.00005, 0.00210

*One lag, introduced linearly. †One lag, introduced linearly. Standardised models also contained the following explanatory variables with parameters statistically significant ($p < 0.1$): intercept, ns(t,36), friday, saturday, sunday, holidays, ns(temperature_{t,4}), ns(temperature_{t,1,4}), ns(temperature_{t,2,4}), ns(temperature_{t,3,4}), ns(temperature_{t,4,4}), ns(temperature_{t,2,4}):humidity_{t,2,2}, ns(temperature_{t,3,4}):humidity_{t,3,3}, ns(temperature_{t,4,4}):humidity_{t,4,3}, humidity_{t,2,2}, humidity_{t,3,3}, humidity_{t,4,3}, ma(influenza,15), total_{t,1,1}, total_{t,2,2}, total_{t,3,3}, total_{t,4,3}, total_{t,5,5} and total_{t,6,6}. Where ns(x,k) denotes a natural cubic spline for x with k degrees of freedom; day of the week an indicator for the day of the week; : denotes interaction, ma(15) a moving average of 15 terms, and total, daily total mortality.

Table 3 Results of the multicentric meta-analyses, Barcelona, Valencia and Vigo, 1992–1994

Meta-analysis	Black smoke		Sulphur dioxide	
	β	95% CI	β	95% CI
Flexible	0.00144	0.00021, 0.00266	-0.00023	-0.00101, 0.00106
Standardised	0.00092	0.00007, 0.00177	-0.00030	-0.00196, 0.00137

Individual models for flexible meta-analysis contained the following explanatory variables with parameters statistically significant ($p < 0.1$):
 Barcelona: ns(TREND,36), saturday, sunday, ns(temperature_{t,1,4}), ns(temperature_{t,2,4}), ns(temperature_{t,3,4}), ns(temperature_{t,4,4}), ns(temperature_{t,3,4}):humidity_{t,3,3}, ma(influenza,15), total_{t,1,1}, total_{t,4,4} and total_{t,6,6}.
 Valencia: ns(TREND,110), friday, saturday, ns(temperature_{t,1,4}), ns(temperature_{t,2,4}), ns(temperature_{t,3,4}), ns(temperature_{t,4,4}), ns(temperature_{t,3,4}):humidity_{t,3,3}, ma(influenza,15), total_{t,1,1}, total_{t,3,3} and total_{t,4,4}.
 Vigo: ns(TREND,36), saturday, holidays, ns(temperature_{t,4}), ns(temperature_{t,1,4}), ns(temperature_{t,2,4}), ns(temperature_{t,3,4}), humidity_{t,2,2}, humidity_{t,3,3}, ns(temperature_{t,2,4}):humidity_{t,2,2}, ma(influenza,15), total_{t,1,1}, total_{t,2,2}, total_{t,3,3}.
 Individuals models for standardised meta-analysis contained the variables shown in table 2.

Table 4 Results of the ecological-longitudinal analysis for three cities, Barcelona, Valencia and Vigo, 1992–1994

City	Black smoke*		Sulphur dioxide†	
	β	95% CI	β	95% CI
<i>Air pollutants introduced separately in two models</i>				
All three	0.00064	0.00021, 0.00106	0.00041	0.00011, 0.00072
Barcelona	0.00064		0.00041	
Valencia	0.00063		0.00041	
Vigo	0.00064		0.00041	
<i>Air pollutants introduced jointly in the same model</i>				
All three	0.00071	0.00034, 0.00108	0.00020	0.00009, 0.00031
Barcelona	0.00071		0.00020	
Valencia	0.00071		0.00020	
Vigo	0.00071		0.00020	

*One lag, introduced linearly. †Current value and one lag, introduced linearly. Models also contained the following explanatory variables with parameters statistically significant ($p < 0.1$): intercept, ns(t,36), friday, saturday, sunday, holidays, ns(temperature_{t,1,4}), ns(temperature_{t,2,4}), ns(temperature_{t,3,4}), ns(temperature_{t,4,4}), ns(temperature_{t,7,4}), ns(temperature_{t,2,4}):humidity_{t,2,2}, ns(temperature_{t,3,4}):humidity_{t,3,3}, ns(temperature_{t,4,4}):humidity_{t,4,3}, humidity_{t,2,2}, humidity_{t,3,3}, ma(influenza,15), total_{t,1,1}, total_{t,2,2}, total_{t,3,3}, total_{t,4,3}, total_{t,5,5} and total_{t,6,6}. Where ns(x,k) denotes a natural cubic spline for x with k degrees of freedom; day of the week an indicator for the day of the week; : denotes interaction; ma(15) a moving average of 15 terms and total, daily total mortality.

the way these data were analysed (variables transformation, core confounders to be accommodated, and alike).

Results

Table 1 shows the distribution of the mortality, air pollution and meteorological variables as well as additional demographic information on the three cities, Barcelona, Valencia and Vigo, during the period 1992–1994. Daily average mortality counts for Barcelona were 2.65 times higher than in Valencia and 8.23 times than in Vigo. These ratios did not correspond to those of the total population, in Barcelona it was 2.19 times higher than in Valencia and 5.98 times than in Vigo. The percentages of population over 70 years old were 11.7% in Barcelona, 10% in Valencia and 7.1% in Vigo. Among air pollutants, differences were in the same line of mortality only in the case of the daily average levels of sulphur dioxide (that is, the average level in Barcelona was 1.63 times higher than in Valencia and 2.10 times than in Vigo). Compare, however, with the case of black smoke, the daily average level was 2.10 times higher in Vigo than in Valencia and 2.36 times than in Barcelona. With respect to relative humidity there were not practical seasonal differences and, moreover, the behaviour among cities did not present a systematic pattern.

In table 2 we show the results of the analysis of the (standardised) individual models (results from the flexible individual models hardly differed). In the case of black smoke, although they were not statistically different, the estimates of the relative risks ($\exp(\beta)$) varied between Valencia and the rest of the cities. The highest relative risk corresponded to Valencia, 1.01185 (95% confidence intervals 1.0004, 1.0020)—that is, an increase of $10 \mu\text{g}/\text{m}^3$ in black smoke augmented 1.185% the probability of a death on a given day. Relative risks were practically identical for Barcelona and Vigo, slightly higher for Barcelona (1.00696 (95%CI 1.0001, 1.0012) v 1.00685 (95%CI 1.0001, 1.0013)). Relative risks for sulphur dioxide were very heterogeneous: 0.99871 (95%CI 0.9974, 0.9999) for Valencia; 1.00282 (95%CI 1.0001, 1.0005) for Barcelona; and 1.01031 (95% CI 0.9999, 1.0021) for Vigo.

The results of the two MA are shown in table 3. The effects for the flexible meta-analysis were higher than in the standardised, in the case of black smoke (1.01445 (95% CI 1.0002, 1.0027) v 1.00927 (95% CI 1.0001, 1.0018) for a $10 \mu\text{g}/\text{m}^3$ increase). It is probably a consequence of the great heterogeneity found in the estimated individual coefficients associated

with sulphur dioxide. When combined, relative risks were not statistically significant (0.9997, 95% CI 0.9990, 1.0011 in both cases).

In table 4, finally, we show the results of the ecological-longitudinal analysis. Note that the relative risks were practically identical across cities, 1.00638 (95% CI 1.0002, 1.0011) for an increase of $10 \mu\text{g}/\text{m}^3$ of black smoke, and 1.00415 (95% CI 1.0001, 1.0007) for an increase of $10 \mu\text{g}/\text{m}^3$ of sulphur dioxide. In fact, note that the slopes did not seem very different in both air pollutants (fig 2). Results were very different from those of the MA, in particular for sulphur dioxide. Moreover, because the statistical power was higher than in the individual analysis, we included several lags of sulphur dioxide and also the two air pollutants

in the same model. In fact the associated coefficients were statistically significant in all the cases. Estimated relative risk for sulphur dioxide was lower (1.00200 (95% CI 1.0001, 1.0003) for an increase of $10 \mu\text{g}/\text{m}^3$) and for black smoke higher (1.00712 (95% CI 1.0003, 1.0011) for a $10 \mu\text{g}/\text{m}^3$ increase). Note, however, that the sum of the effects of both pollutants, that is the exponentiation of the sum of the slopes, was practically the same (1.00913 air pollutants jointly *v* 1.01053 each pollutant separately) suggesting there was no independent effect of sulphur dioxide once black smoke was in the model. Correlations between these two pollutants in the different cities, however, were only moderate (Vigo 0.456; Barcelona 0.482; Valencia 0.625).

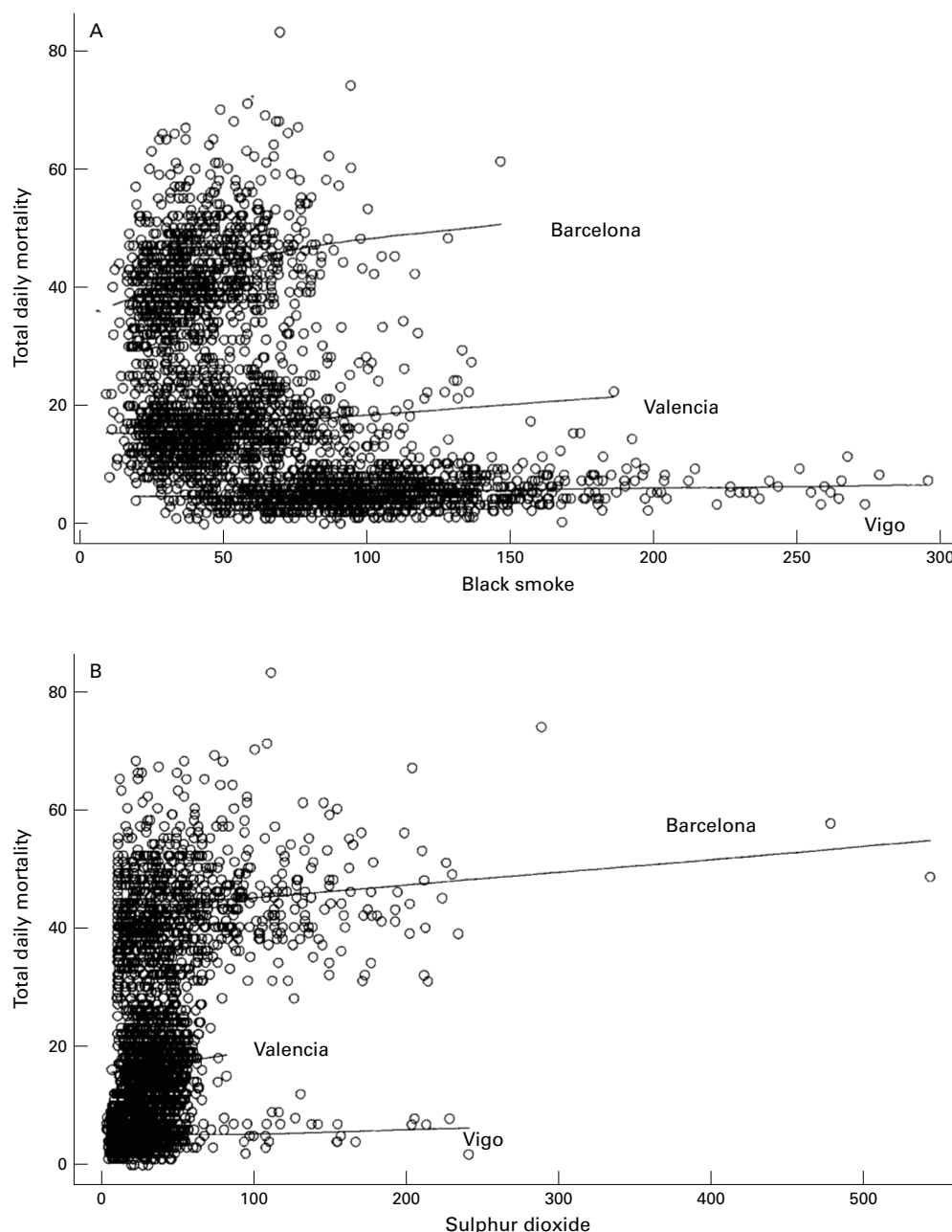


Figure 2 (A) Relation between average daily levels of black smoke and daily total mortality, Barcelona, Valencia and Vigo, 1992–1994. (B) Relation between average daily levels of sulphur dioxide and daily total mortality, Barcelona, Valencia and Vigo, 1992–94.

Higher statistical power also led to more interactions that became statistically significant with respect to individual models, especially those between air pollutants and meteorological variables.

Discussion

DISCUSSION OF THE RESULTS

As was mentioned above, as an illustration of the new approach we are introducing in this paper, we have obtained quantitative estimates of the short-term effects of air pollution on mortality in three Spanish cities (Barcelona, Valencia and Vigo). Results have been different in individual and in combined analyses, and also between the two types of combined analyses.

Individual results showed a high heterogeneity. The problem was that the adverse effects of the air pollutants on mortality were very small and not always consistent. There were differences, however, between the two air pollutants. The relations between black smoke and total daily mortality were always statistically significant, showing a consistent dynamic behaviour (an increase in the black smoke levels of the previous day was associated with an augment in the mortality in the current day). Relative risks for a 10 $\mu\text{g}/\text{m}^3$ increase of black smoke were estimated from 1.007 (Barcelona and Vigo) to 1.012 (Valencia), although, in fact, differences were not statistically significant. These findings coincide with those obtained previously using GAM (Vigo)³⁴ and also GLM models^{35 36} (even with respect to the number of lags statistically significant). High heterogeneity was found in the individual results for sulphur dioxide. In this sense, not only relative risk was five times higher in Vigo than in Barcelona, but also it was protective for Valencia. These results were also found previously using other different methods.^{34 35 36} It must take into account that sulphur dioxide was not measured similarly in the different locations, thus highly precluding comparisons of local results without further evaluation.

These individual results were combined using, first, two type of multicentric MA. Random effects MA take into account the heterogeneity by widening the confidence intervals of the joint estimate. In this sense, the very high heterogeneity found in the individual effects of sulphur dioxide has led to non-significant combined effects. Nevertheless, this should not be the most suitable strategy. When there is heterogeneity, one should not try to lump individual results together, but rather should try to understand factors that explain the discrepancies, using a meta-regression, for instance.

Results have been very different using our ecological-longitudinal approach. Although part of the difference between results could be ascribed to the different confounder models we think that, under this approach, we could better estimate, in the sense of more efficiently, the short-term effects of air pollution on mortality. We found a statistically significant association between the two air pollutants, black smoke and sulphur dioxide, and mortality for all

KEY POINTS

- Air pollution levels were related to mortality in the three Spanish cities of the study, Barcelona, Valencia and Vigo.
- Whereas the relations between black smoke and mortality were consistent, high heterogeneity was found for sulphur dioxide.
- A 10 $\mu\text{g}/\text{m}^3$ increase in the daily levels of black smoke was associated with a 0.7% increase in the daily number of deaths for all causes.
- It seems that there was no independent effect of sulphur dioxide once black smoke was in the model.

causes. A 10 $\mu\text{g}/\text{m}^3$ increase in the daily levels of black smoke was associated with a 0.7% increase in the daily number of deaths for all causes. The effect was practically the same when the effects of sulphur dioxide were controlled for in the same equation. Estimated relative risks for sulphur dioxide although always statistically significant, differed when this air pollutant was analysed independently (0.4% increase in mortality for a 10 $\mu\text{g}/\text{m}^3$ augment in sulphur dioxide) or together with black smoke in the same model (0.2% increase in mortality). As was mentioned above, it seems that there was no independent effect of sulphur dioxide once black smoke was in the model. When black smoke is included in the model, the effect of sulphur dioxide is dramatically reduced. Furthermore, sulphur dioxide was measured differently across cities, leading to a poor comparability of sulphur data across cities.

Recently, several multicentric MA in European cities^{10 12 13} (that is, APHEA project) and also in US^{11 37} cities, have reported significant associations between air pollution and daily mortality. In all the cases, MA were of the flexible type—that is, the effects of air pollutants found in individual models (in which data collection procedures and the definition of some variables were standardised) were combined by means of a weighted average (the weights were proportional to the local variances). The results of the APHEA project MA^{10 12 13} were very similar to those found in this paper. In the APHEA MA, relative risks in western cities for each one of the air pollutants considered here, black smoke and sulphur dioxide, were estimated at close to 1.006. The independent effects (with respect to the other air pollutant) were slightly lower, 1.004 for black smoke and 1.005 for sulphur dioxide. Results from US MA^{11 37} differed in magnitude with those found here. However, as Moolgavkar and Luebeck³⁸ point out, the main limitation of the US MA³⁷ was that they did not use the same indicator for particles in all cities, but converted different particulate air pollution indicators (black smoke, total suspended particles) by means of ad hoc factors. The generalised adequacy of these conversion factors has been questioned in other recent studies.³⁸

In 1998, other multicentric MA carried out in Canada.³⁹ The number of daily deaths for non-accidental causes and daily concentrations of ambient gaseous air pollutants (nitrogen dioxide, sulphur dioxide, ozone and carbon monoxide) were obtained in 11 cities from 1980 and 1991. The relation between mortality and air pollutants was estimated in each city, using Poisson GAM. The statistical method was very similar to that described above, in this paper, as standardised individual model, allowing multiple pollutant in the same regression. To obtain summary risks, Burnett *et al*,³⁹ however, simply averaged risks among cities (in fact the procedure is called a “fixed effects” MA). Nitrogen dioxide had the largest effect on mortality with a 4.1% increased risk ($p < 0.01$), followed by ozone at 1.8% ($p < 0.01$), sulphur dioxide at 1.4% ($p < 0.01$) and carbon monoxide at 0.9% ($p = 0.04$). In this study, however, comprehensive particulate data were not available, leading to some doubt about the independent effect of nitrogen dioxide, considering the correlation between these pollutants.⁴⁰

POOLED ANALYSES VERSUS INDIVIDUAL ANALYSES

Our approach possesses several major advantages over conventional time series designs. Pooled analyses like ours give a larger number of data points, increasing the degree of freedom and hence improving the statistical power of the analyses. This higher power allows the complementation of aspects such as accommodation of random effect models, of lags, of interaction between pollutants and between pollutants and meteorological variables, that are hardly complemented in regular time series analyses.

As a first consequence, the effects of a presumably high collinearity among explanatory variables might be noticeably attenuated. In fact, as is known, a typical solution of high multicollinearity lies in increasing the sample size. In this sense, consider the results of the individual analyses previous to the multicentric MA. There was a very important problem of collinearity in all the individual models (for instance, correlation coefficients between black smoke and sulphur dioxide were close to 0.45 and between black smoke and temperature close to -0.5). The reasons are that, on one hand, air pollutants share sources of emission and, on the other hand, present the same seasonal behaviour. Because high collinearity increased standard errors it is very probable that statistical significance of the parameters could be masked in many cases. Note, in this sense, that individual models used in the flexible meta-analysis differed—that is, included variables and functional forms—from those used in the standardised MA. As a consequence, results cannot be directly compared and, therefore, inferences might be difficult to generalise. But, moreover, some of the observed explanatory variables that were omitted because their coefficients seemed to be not significant,⁴¹ could actually be relevant. This fact led to, at least, a very complex model, in the sense of the control of unobserved

confounders and of the residual autocorrelation, and, perhaps, to those observed inconsistencies. See, in this sense, the individual models for sulphur dioxide and that of Valencia, in particular.

The second, and perhaps more important advantage derived from higher statistical power, is that it allows us to analyse a number of important hypotheses that cannot be investigated using standard time series analyses. With a larger sample size the effect of rare exposures can also be examined, which is often not possible in single studies. New hypotheses may be formulated with these types of pooled analyses and specific subgroups may also be analysed. In this sense, note that the ecological-longitudinal approach was the only one that allowed jointly considering, in the same model, the effects of more lags of the same pollutants, the effects of more than one pollutant and the consideration of possible interactions between meteorological variables and air pollutants.

MULTICENTRIC MA VERSUS

ECOLOGICAL-LONGITUDINAL ANALYSES

PA of epidemiological studies have recently become more common as an alternative of MA in a great variety of settings: clinical trials⁴²; cohort studies⁵; case-control studies.⁹ However, to our knowledge, there has not been an attempt to apply these techniques to ecological-time series designs.

The actual strategies followed by the multicentric studies^{10 12 13} and also those of US cities,^{11 37} those we called flexible MA, were not actually “pure” MA, but a mix between MA and prospectively planned PA (or Type IV meta-analysis⁷). The objective was to reduce the severe limitations of the MA of published papers and heterogeneity between studies in particular. Because analyses differ considerably in their designs, data collection methods and the definition of the exposure and confounder variables, the heterogeneity between studies used to be extremely high. Hence, those multicentric studies followed standardising data collection procedures, definition of variables, questions and hypotheses for the individual studies (that is, as in Type IV MA). However, because the costs for Type IV MA are very high, once results were obtained, using a (variance based) weighted average of the individual estimates a combined estimate is computed—that is, individual results are simply pooled (as in Type II MA). The problem is that because the range of geographical, climatic and socio-demographic patterns between cities (centres) used to be too wide, separate studies did not adjust exactly for the same confounding factors. In this sense, the functional form of the covariates entered in the individual models, the indicator variables that fit unusual events or controls for influenza epidemics; and even the explanatory variables included in the models (in particular the number of lags when the model is dynamic) depend on local conditions.

Our argument is that the high heterogeneity between studies has dramatically widened the confidence interval of the joint estimates.³⁹ This could have questioned the inferences

drawn from multicentric studies, in the sense that they could not be as general as one had expected. Note, for instance, the differences of the results of the MA and of the ecological-longitudinal approach. Because of instability resulting from very high heterogeneity, MA were not able to detect the presumably statistical significance of the effects of air pollutants on total mortality.

Our approach is closely related to PA. As in these PA, we reanalysed individual data based on primary studies. In fact, we consider the same inclusion criteria for all studies, a unified definition of the variables (covariates included, functional form and number of lags), and new statistical modelling of the pooled dataset. The key difference of our approach, however, is that we allow for heterogeneity between the effects of different studies. That is to say, we assume (and try to estimate) that each study has its own true exposure effect (and not necessarily the same for all studies as assumed by PA). At the same time, we assume that there is random distribution of these exposure effects around a central effect (the only of interest in PA). This assumption requires the use of proper statistical techniques, in particular those derived from longitudinal studies.

SOME LIMITATIONS OF THE ECOLOGICAL-LONGITUDINAL ANALYSIS

Despite its advantages, our approach could present some limitations, in particular from a statistical and from an epidemiological point of view.

Firstly, from a statistical point of view, as we mentioned above, we controlled for unobserved confounders with a systematic variation, by means of a cubic spline of time. Of course, amplitudes and periods of the cyclic behaviours of these confounders differ between cities. For this reason we introduced calendar specific days in our ecological-longitudinal model and, more important, three flexible functions of time, one for each city. The problem was that control was not perfect. For this reason there remained some residual autocorrelation (up to order six). Precisely, the autoregressive GAMMs were able to capture this serial correlation. However, it was possible that observed predictors (for instance the demographic structure of the city or several socioeconomic variables) could explain part of this (residual) heterogeneity. The modelling of such over-dispersion could improve the efficiency of the estimates. At any rate this is a topic that deserves further research.

Secondly, the residual autocorrelation was controlled for in the same form in the three cities (introducing up to six lags of the dependent variable). It was possible, however, that there were “cluster” effects, that is residual autocorrelation (heterogeneity) specific to each city. In this sense, a more flexible approach able to capture this “clustering” could also provide more efficient estimates.

Another limitation to our approach was the use of natural cubic splines with equally spaced knots versus the alternative more flexible use of smoothing splines, in the estimation of the

non-parametric functions in the GAMMs. The problem is that as the evaluation of the corresponding log-quasi-likelihood requires numerical integration, except for the Gaussian case, it is often difficult to calculate full natural cubic smoothing spline estimators by directly maximising the associated penalised log-quasi-likelihood. Lin and Zhang,⁴³ however, have recently proposed a double penalised quasi-likelihood to make an approximate inference. Although this topic deserves further consideration, we think that the use of the corrected AIC in the choice of the amount of smoothing, adapting it (that is, the degrees of freedom) to the actual dataset, could have minimised the relative higher rigidity of our approach.

From an epidemiological point of view, our principal limitation was that our approach assumed that all the variables reflect the same phenomenon in the different “individuals” (cities). This assumption could be hardly attainable in the case of black smoke, for instance, because the indicator of such a pollutant could be measuring a different mixture of particles in each different city. Furthermore, although observed confounders were considered random effects in our ecological-longitudinal analysis, some interactions, in particular those of meteorological variables and air pollutants, could not have been totally modelled, from an individual level. In this sense, a multicentric MA might have been more flexible to local conditions.

At any rate, to assess the effects of air pollutants on health in more than one geographical area, we think that the strategy followed in this paper may be a very valid approach. Firstly, it is important that common rules be strictly stated a priori, so that data across different cities have some common features (for example, data collection, core confounders, data transform and alike). Secondly, individual analyses for each city (we recommend autoregressive Poisson GAM models) should be carried out. This step would allow one to detect possible sources of heterogeneity. Then, depending on the size of such heterogeneity, flexible MA might be undertaken. The last, and perhaps the most important, step should be the ecological-longitudinal analysis, for which we recommend following the procedure described in this paper, in particular the autoregressive Poisson GAMM models. Results from the two combined analyses—that is, MA and ecological-longitudinal—should be compared, to assess the actual importance of heterogeneity, to better control confounding, for more precise detection of interactions and to analyse further epidemiological hypotheses.

We would like to thank to the editors, specially to Idefonso Hernández, and two referees of an earlier draft of this paper for their comments and suggestions. Any remaining errors or omissions are our own.

Funding: this project was funded in part by two grants from Fondo de Investigaciones Sanitarias, FIS 97/0051 and FIS 00/0010.

Conflicts of interest: none.

1 Gordon I, Boffetta P, Demers PA. A case study comparing a meta-analysis and a pooled analysis of studies of sinonasal cancer among wood workers. *Epidemiol* 1998;9:518–24.

- 2 Friedenreich CM. Methods for pooled analyses of epidemiological studies. *Epidemiology* 1993;4:295-302.
- 3 Friedenreich CM, Brant RF, Riboli E. Influence of methodological factors in a pooled analysis of 13 case-control studies of colorectal cancer and dietary fiber. *Epidemiology* 1994;5:66-79.
- 4 Leroy V, Newell ML, Dabis F, et al. International multicentre pooled analysis of late postnatal mother-to-child transmission of HIV-1 infection. Ghent International Working Group on Mother-to-Child Transmission of HIV. *Lancet* 1998;352:597-600.
- 5 Hunter DJ, Spiegelman D, Adami HO, et al. Cohort studies of fat intake and the risk of breast cancer—a pooled analysis. *N Engl J Med* 1996;334:356-61.
- 6 Checkoway H. Data pooling in occupational studies. *J Occup Med* 1991;33:1257-60.
- 7 Blettner M, Sauerbrei W, Schlehofer B, et al. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol* 1999;28:1-9.
- 8 Steinberg KK, Smith SJ, Stroup DF, et al. Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. *Am J Epidemiol* 1997;145:917-25.
- 9 Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet* 1996;347:1713-27.
- 10 Katsouyanni K, Touloumi G, Spix C, et al. Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. *BMJ* 1997;314:1658-63.
- 11 Samet JM, Zeger SL, Berhane K. *Particulate air pollution and daily mortality: replication and validation of selected studies*. Health Effects Institute Report. Cambridge, MA: Health Effects Institute, 1995.
- 12 Touloumi G, Katsouyanni K, Zmirou D, et al. Short term effects of ambient oxidants exposure on mortality: a combined analysis within the APHEA project. *Am J Epidemiol* 1997;146:77-185.
- 13 Zmirou D, Schwartz J, Saez M, et al. Time-series analysis of air pollution and cause-specific mortality. *Epidemiology* 1998; 9:495-503.
- 14 Katsouyanni K, Zmirou D, Spix C, et al. Short-term effects of air pollution on health. An European approach using epidemiologic time series data. The APHEA project: background, objectives, design. *Eur Respir J* 1995;8:1030-8.
- 15 Ballester F, Saez M, Alonso ME, et al. The EMECAM project. Spanish multicentric study on the relationship between air pollution and the mortality. Background, participants, objectives and methods. [In Spanish]. *Rev Esp Salud Pública* 1999;73:165-76.
- 16 Schwartz J, Spix C, Touloumi G, 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):S3-11.
- 17 McCullagh P, Nelder JA. *Generalized linear models*. London: Chapman and Hall, 1989.
- 18 Saez M, Sunyer J, Castellsagué J, et al. Relationship between weather temperature and mortality: a time series analysis approach in Barcelona. *Int J Epidemiol* 1995;24:576-82.
- 19 Ballester F, Corella D, Pérez-Hoyos S, et al. Mortality as a function of temperature. A study in Valencia, Spain, 1991-1993. *Int J Epidemiol* 1997;26:551-61.
- 20 Hastie TJ, Tibshirani RJ. *Generalized additive models*. London: Chapman and Hall, 1990.
- 21 Green PJ, Silverman BW. *Nonparametric regression and generalized linear models*. London: Chapman and Hall, 1994.
- 22 Kelsall JE, Samet JM, Zeger SL, et al. Air pollution and mortality in Philadelphia, 1974-1988. *Am J Epidemiol*, 1997;146:750-62.
- 23 Hurvich CM, Simonoff JS, Tsai CL. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society Ser B*, 1998;60:271-93.
- 24 Saez M, Figueiras A. Re: Air pollution and mortality in Philadelphia, 1974-1988. *Am J Epidemiol* 2000;151:106-7.
- 25 Besag J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Ser B* 1974;36:192-236.
- 26 Diggle PJ, Liang KY, Zeger SL. *Analysis of longitudinal data*. Oxford: Oxford University Press, 1994.
- 27 Saez M, Pérez-Hoyos S, Tobias A, et al. Time series methods in the epidemiological studies regarding air pollution. [In Spanish]. *Rev Esp Salud Pública*, 1999;73:33-144.
- 28 Chambers JM, Hastie TJ. *Statistical models in S*. Pacific Grove, CA: Wadsworth and Brooks Cole Advanced Books and Software, 1992.
- 29 Data Analysis Products Division. *S-PLUS 4 guide to statistics*. Seattle: MathSoft, 1997.
- 30 Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993;88:9-25.
- 31 Patterson HD, Thompson R. Recovery of inter-block information when block sizes are unequal. *Biometrics* 1971;58: 545-54.
- 32 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
- 33 Dickersin K, Berlin JA. Meta-analysis: state of the science. *Epidemiol Rev* 1992;14:154-76.
- 34 Taracido M, Figueiras A, Castro I. The short-term impact of air pollution on the mortality. Results of the EMECAM project in the city of Vigo, 1991-1994. [In Spanish]. *Rev Esp Salud Pública*, 1999;73:275-82.
- 35 Tenias JM, Pérez-Hoyos S, Molina R, et al. The short-term impact of air pollution on the mortality. Results of the EMECAM project in the city of Valencia, 1994-1996. [In Spanish]. *Rev Esp Salud Pública*, 1999;73:267-74.
- 36 Saurina C, Barceló MA, Saez M, et al. The short-term impact of air pollution on the mortality. Results of the EMECAM project in the city of Barcelona, 1991-1995. [In Spanish]. *Rev Esp Salud Pública*, 1999;73:199-208.
- 37 Schwartz J, Dockery DW, Neas LM. Is daily mortality associated specifically with fine particles? *Journal of Air Waste Management Association* 1996;46:927-39.
- 38 Moolgavkar S, Luebeck E. A critical review of the evidence on particulate air pollution and mortality. *Epidemiology* 1996;7:420-8.
- 39 Burnett RT, Cakmak S, Brook JR. The effect of the urban ambient air pollution mix on daily mortality rates in 11 Canadian cities. *Can J Public Health* 1998;89:152-6.
- 40 Bates DV. Air pollution: the environmental Phoenix. *Can J Public Health* 1998;89:149-50.
- 41 Saez M, Barceló MA. A criteria to omit superfluous variables in regression models. [In Spanish]. *Gac Sanit* 1998;12:281-5.
- 42 The trial Fibrillation Investigators. The efficacy of aspirin in patients within a trial fibrillation. *Arch Intern Med* 1997;157:1237-40.
- 43 Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society Ser B* 1999;61:381-400.