Research Article

Environmetrics

Received: 31 July 2008, Accepted: 10 August 2009, Published online in Wiley Online Library: 18 November 2009

(wileyonlinelibrary.com) DOI:10.1002/env.1031

Robustness of different regression modelling strategies in epidemiology: a time-series analysis of hospital admissions and air pollutants in Lisbon (1999–2004)

Susana M. Sarmento[a*], Tona G. Verburg[a], S. Marta Almeida[b], Maria C. Freitas[b] and Hubert Th. Wolterbeek[a]

Studies of the acute health effects of air pollution have used exposure windows of different spans and related them to single-day responses. Little is known about whether an increased response window span might be a viable alternative to single-day responses. Our aim is to compare a new model specification where both the exposure and response variables are represented as 7 day moving averages (CMA&CMA model) with the most widely used model specifications in the literature, where the response variable is usually a single-day, in terms of coefficients and their precision and robustness. To this end, daily series of 12 emergency-related hospital admissions and 6 air pollutants spanning 5.5 years in Lisbon were analysed through single-pollutant linear regression and, when necessary M-estimation. With our data, the CMA&CMA model yields coefficients that are very close to models where only the exposure variable is specified as a moving average whether the latter are estimated by OLS or robust M-estimation. In addition, the CMA&CMA model leads to more precise and robust estimates than other model specifications. The new model specification is a straightforward tool for adjusting weekend effects and errors. It is also analogous to robust estimation, with the added advantages of being sensitive to extreme values that are clustered in time, and leading to more precise and robust estimates without loss of high-frequency information. One drawback is the induction of autocorrelation in the residuals. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: epidemiology; air pollutants; moving averages; distributed lag models; robust M-estimation

1. INTRODUCTION

Environmental protection agencies recommend averaging times for each air pollutant (AP) on the basis of their usefulness for specific purposes (e.g. acute or chronic human health protection or vegetation protection) and on considerations of the time-scales at which APs fluctuate (WHO, 2005). For the purpose of human health protection, the choice of the averaging times for APs appears to be trapped in a circular reasoning because most epidemiological and toxicological studies tend to use the recommended averaging times whereas the recommended averaging times are based on epidemiological and toxicological studies. The regulatory agencies themselves acknowledge that some degree of subjectivity underlies the setting of this recommendation (WHO, 2005). Historically, the recommended averaging time for acute human health effects has been 24 h, which is the minimum time-unit for which clinical health data are routinely available. Recently, however, epidemiological studies have been using exposures longer than 1 day, either in the form of moving averages (MAs) or distributed lags (DLMs) and have consistently found that effects of APs increase as the time-span of the exposure window is increased (e.g. Goodman *et al*., 2004 and references therein). Such studies, though using exposures of several days or even weeks, always specify the health response variable as a single-day (known exceptions to use are Roberts, 2005; Schwartz, 2000a). In this context, we present a new (except for Schwartz, 2000a) model specification (CMA&CMA) where both the exposure and response variables are 7 day centred MAs. In this article we first present the a priori motivation for choosing the CMA&CMA model specification and then we compare it with the most widely used model specifications in the literature, where the response variable is always specified as a single-day, whereas the exposure variable is specified as either single-days at different lags, MAs or DLMs. Because our aim is to compare model specifications, which differ solely in the way the exposure and response variables are specified in time, our modelling strategies diverge somewhat from those commonly found in studies aimed at causal inference or prediction.

* *Correspondence to: S. M. Sarmento, Department of Radiation, Radionuclides and Reactors, Section RIH, Reactor Institute Delft, Delft University of Technology; Mekelweg 15, 2629 JB Delft, The Netherlands. E-mail: ssarmento@tudelft.nl*

a *S. M. Sarmento, T. G. Verburg, H. T. Wolterbeek*
Department of Radiation, Radionuclides and Reactors, Section RIH, Reactor Institute Delft, Delft University of Technology, Netherlands

b *S. M. Almeida, M. C. Freitas*
Neutron Activation in Environment, Nutrition and Epidemiology, Unity of Reactors and Nuclear Safety, Technological and Nuclear Institute, Portugal

2. METHODS

2.1. Data description

All data spanned the period between the 1st January 1999 and 30th June 2004.

Daily counts of hospital admissions (HAs) in 7 public hospitals in Lisbon were kindly provided by the Administração Central do Sistema de Saúde in Portugal and were aggregated into the following diagnostic categories (ICD9-CM): respiratory (460-519), circulatory (390-459) and cardiac diseases (390-429) and into the age-groups: <15, 15-64, >64 and total; yielding a total of 12 HAs categories.

Hourly concentrations of six APs: PM10, SO_2, NO, NO_2, CO and O_3 were obtained from the National Environmental Institute. For each monitoring station, daily concentrations were obtained by calculating the 24 h mean for PM10 and SO_2 and the daily 1 h maximum for the other APs. Spatial averaging across Lisbon was performed over the three central-site monitors that measured all APs throughout the entire study period: Avenida da Liberdade, Entrecampos and Olivais. These calculations were performed according to the recommendation of WHO (1999).

Daily mean temperature and relative humidity were obtained from the National Institute of Meteorology.

2.2. Model specifications

The aim of this paper is to compare the statistical output of the five model specifications described below, by running each model specification on the 72 HA–AP relationships available (12HAx6APs). The five model specifications differ in terms of the way the exposure and response variables are specified in time.

1. CMA&CMA: HAs and APs are both specified as 7 day centred moving averages (CMAs). Model includes only two terms: an AP and a constant, since weekday-related fluctuations are smoothed by the 7 days window of the CMA. This model specification is the reference model against which all other model specifications below are compared with, because this is the least known model specification.
2. O&CMA: HAs are specified as single-day values, whereas the APs are specified as CMAs. Model includes three terms: a constant, a workday dummy and an AP.
3. O&DLM: HAs and APs are both specified as single-day values. Model includes a constant, a workday dummy and seven terms for the AP lagged from 0 to 6 days. Estimation was performed without constraining the coefficients since we were only interested in the net-slope across all lags.
4. Single-day lags: HAs and APs are both specified as single-day values. Model includes a constant, a workday dummy and one term for the AP at one of 0 to 6 lags. Therefore, for this model specification and for each AP, seven single-day lag models were computed, one for each lagged value of the AP.
5. Best-lag: single-day lag model that, for a given HA–AP relationship, yielded the highest slope among the seven single-day lag models described in item 4. The use of the highest slope as the criterion for choosing the 'best' lag is a usual procedure in the literature (Lumley and Sheppard, 2000).

Because CMA&CMA and O&CMA models include backward lags, additional model specifications were also assessed, namely: FMA&PMA and O&PMA models. The results of these models are mentioned when relevant but their formal results are not presented for the sake of brevity. We preferred to present the results of models using CMA variables because CMA does not induce a shift in the series and, as a result, they are more strongly correlated with the original variables than FMA or PMA variables.

2.3. Data manipulation

The 7 day centred moving average (CMA) database was created from the original (O) database by replacing each daily observation by the mean over $k = 7$ days (window) and attributing this mean to the $(k+1)/2$ day. The 7 day prior and forward moving average (PMA and FMA) were calculated in the same way but the mean was attributed to either the 7th (PMA) or 1st day (FMA) of the window. CMA, PMA and FMA were only calculated when there were no missing values within their windows.

Because our aim is to compare model specifications, it is essential that slopes are comparable between them. To achieve this we normalised all variables in all databases (original, CMA, PMA and FMA) to mean one. Slopes obtained from such normalised variables correspond to elasticities and may be converted into their original slopes by: Elasticity $= \beta\frac{\bar{X}}{\bar{Y}}$ (e.g. Lipfert, 1993).

Although our aim is not causal inference, we felt it important to adjust for any eventual seasonal patterns. We opted for month stratification because it requires few assumptions, adjusts for confounding, effect modification and non-linearities simultaneously and yields simple estimates that are easy to compare. Month stratification does not, however, adjust for potential inter-annual trends. Nevertheless, the method and extent of long-wave adjustment are believed not to be important for the purposes of this article as it will affect all model specifications being compared in a similar manner. Furthermore, because the results of the comparison between model specifications are identical for all month strata, we present results only for the January stratum (January1999, January2000,. . ., January2004). Since month stratification implicitly controls for weather effects, and temperature and humidity show very low variability in the January stratum (Table 1), these two variables were not considered in any of the analyses.

Adjustment for weekday effects was made by using a single dummy variable distinguishing weekends from workdays, which captures the major weekly fluctuation. To reiterate, because our aim is to compare model specifications rather than perform causal inference, it is irrelevant whether we use one dummy or the traditional six dummy variables for each weekday type, as long as all model specifications are treated in the same way the result of the comparisons remains unchanged (results not shown).

2.4. Statistical analyses and software

Linear (OLS) regression was used to estimate the 72 HA–AP relationships across the five model specifications. While most studies have used Poisson or Negative Binomial regression to constrain predictions to positive values, our aim is neither causal nor predictive research, thus this advantage is irrelevant (e.g. Chapter 10 Rothman, 2002). Furthermore, slopes from linear regression may be approximated to Poisson or Negative Binomial slopes times the mean of the dependent variable (e.g. page 89 Cameron and Trivedi 1998). As a check, we compared the slopes (using the approximation) and *t*-values of the slopes obtained from linear regression and Poisson regression with a paired-sample *t*-test: we found no statistically significant (1%) differences between these two types of regression for O&CMA and CMA&CMA models.

In a final comparison we performed M-estimation using Huber's and Tukey's Biweight weighing schemes. M-estimation was used because it has become a common procedure in the literature and one that has an impact on the value of all statistics including the slopes.

The statistical significance level for evaluation of each HA–AP relationship within each model specification was set at 1%, owing to the large number of significance tests performed; for all other tests a 5% significance level was used.

Analyses were performed in: Excel 2003, MatLab 7.0.1 and R 2.6.2.

3. RESULTS

Descriptive statistics are presented in Table 1.

3.1. Reasoning for the a priori choice of the CMA&CMA model

In this section we present the two major theoretical reasons that motivated us to evaluate the CMA&CMA model specification, namely: the ability of MAs to smooth potential influential values and noise in the data and the possibility that ecological studies may lack the temporal resolution to link responses and exposures on such a fine scale as single days. We end the section with a description of some anticipated disadvantages of the CMA&CMA model.

3.1.1. Noise and errors

The epidemiologist rarely participates in the sampling or recording of data and often has no access to retrospective information on the sources of inaccuracies and imprecision. For these reasons, major investments should be made on inspecting data quality and potentially adapt the analyses to these inspections. One such adaptation is the literature's regular use of M-estimation, which gives less weight to extreme values in the dependent variable (e.g. Samoli *et al*., 2001; Schwartz, 2000b).

Table 1. Descriptive statistics of the daily levels of APs ($\mu g\, m^{-3}$), temperature (°C), relative humidity (%) and hospital admissions counts in the January stratum of the original dataset

	N	Sum	Mean	Median	Min	Max	SD-O[a]	SD-CMA[a]	K-S statistic[b]
PM10	186		49.27	43.79	11.05	152.24	25.73	15.46	0.093
SO_2	180		11.07	5.67	0.04	124.47	15.18	10.82	0.234
NO	186		233.64	174.13	8.67	799.57	194.11	101.08	0.154
NO_2	186		95.11	87.18	24.67	270.67	46.21	32.31	0.132
CO	186		2910.09	1980.33	363.00	11102.67	2498.30	1669.07	0.168
O_3	153		40.67	40.00	4.00	77.50	17.67	12.78	0.064*
Temperature	186		11.07	11.25	4.50	16.75	2.53	1.86	0.056*
Humidity	186		79.81	81.13	48.75	98.00	10.25	6.77	0.083
Respiratory<15	186	382	2.05	2	0	8	1.78	0.86	0.185
Respiratory15-64	186	1954	10.51	10	3	23	3.98	1.63	0.104
Respiratory>64	186	1122	6.03	5	1	15	2.67	1.15	0.138
RespiratoryTotal	186	3458	18.59	18	5	34	5.78	3.01	0.076
Circulatory<15	186	2815	15.13	15	5	31	4.69	2.79	0.086
Circulatory15-64	186	1674	9.00	9	2	21	3.54	1.56	0.098
Circulatory>64	186	1082	5.82	6	0	17	2.79	1.21	0.100
Circulatory Total	186	5571	29.95	29	13	49	7.29	3.44	0.068*
Cardiac<15	186	2629	14.13	14	4	27	4.47	2.68	0.084
Cardiac15-64	186	905	4.87	5	0	14	2.58	1.07	0.130
Cardiac>64	186	866	4.66	4	0	14	2.43	1.07	0.125
Cardiac Total	186	4400	23.66	23	9	40	6.21	2.92	0.065*

[a]Standard deviation (SD) of the original (O) dataset and of the 7 days centred moving average dataset (CMA).
[b]Kolgomorov-Smirnov (K-S) statistic for the original dataset where * denotes that the null hypothesis of normality cannot be rejected at the 5% significance level.

In the case of APs, measurements require good quality assurance practices, as they are vulnerable to a wide range of technical problems and meteorological influences (EEA, 1998). Although daily AP concentrations are averaged over hourly measurements and monitors, which smoothes any eventual unusual observations, our data presents several instances of values that may be considered suspicious, as illustrated in Figure 1.

HAs, on the other hand, are known to be particularly vulnerable to errors during data recording, but they are also subject to other more complex interferences. It must be borne in mind that the diagnostics are relatively unspecific to APs and even to acute AP exposures. Even after seasonal adjustment, a fraction of daily HAs may be due to APs acting at other time-scales (sub-daily to chronic) and another fraction may be due to other health determinants altogether (assuming a single causal factor). Errors, mixing of time-scales and health determinants, allied to the rare count nature of HAs which is summed rather than averaged, over time and hospitals, may contribute to the deterioration of data quality. Our HAs data, however, appear to be rather devoid of outliers and though they do present autocorrelation at frequencies $\leqslant 7$ days, part of their variability may be noise (Figure 2).

Since extreme values, which can be either much higher or much lower than the mean, as well as noise could have dubious impact on the results, the advantages of the CMA&CMA model specification are that it is able to smooth both these effects and in both the exposure and response variables. Furthermore, because each data point becomes an average of seven observations, the error associated with each observation is smaller. Consequently, one would expect an increased accuracy and stability of estimates with the CMA&CMA model specification compared to other model specifications. Another important property of MAs is that it smoothes extreme values that are isolated to a greater extent than extreme values that are clustered in time, as illustrated in Figures 1. This is desirable because, in the absence of further information, the first situation is more likely to reflect an error than a 'true' observation compared to the second situation. This property of MAs is not shared by robust estimation methods such as M-estimation.

3.1.2. Response duration or exposure duration?

A recent trend in acute epidemiological studies of air pollution has been to model single-day health responses with exposures of multiple days, either as MAs or DLMs. Such asymmetrical models are based on the assumption that the associations occur on a 1 day-to-1 day basis at the individual level, but owing to the likely existence of multiple susceptibility subgroups in large populations, different time-intervals (lags) are allowed to exist between the single-day's exposure and the single-day's response (Roberts, 2005). This reasoning implies that for a given exposure day, the health response has a duration of more than 1 day at the population level. It is unclear to us why then MAs are applied to the exposure variable instead of the response variable (DLMs would not be possible for the response variable). As Lipfert (1993) pointed out, for stationary time-series, summing or averaging the APs over a number of days is equivalent to adding responses for the same number of days.

Our scepticism with the approach described above stems from the difficulty in knowing, at least in ecological studies, whether what is being measured by such models is: (1) a 1 day-to-1 day individual-level relationship in a population with multiple induction periods (response duration), as is often argued; or (2) a 1 day-to-kdays individual-level relationships in a population with multiple exposure duration requirements. We feel that it might be a mixture of both for the reasons that follow.

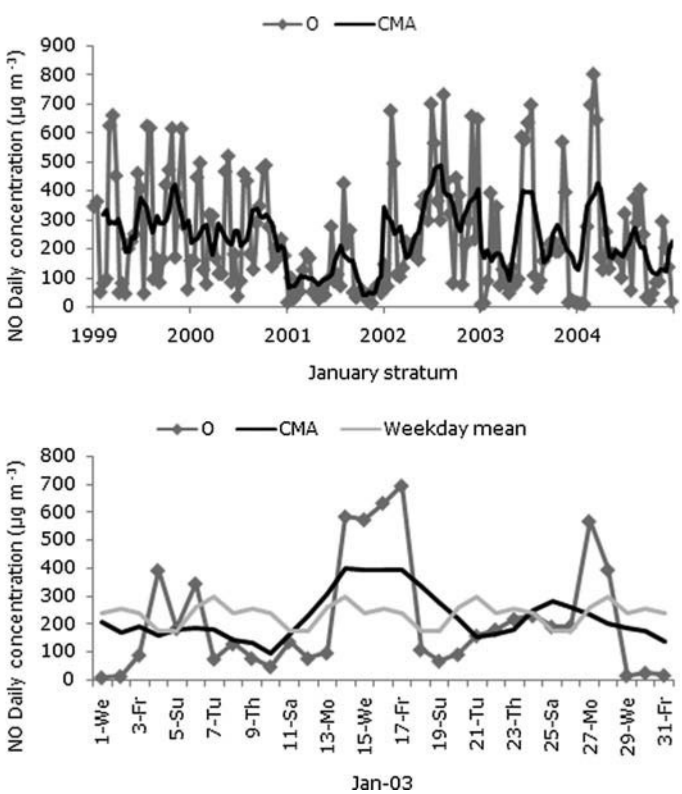

Figure 1. Sequential plot of daily levels of NO in the January stratum and close-up for the year 2003. Both the original data (O) and the CMA data are plotted as well as the mean concentration for each weekday type.

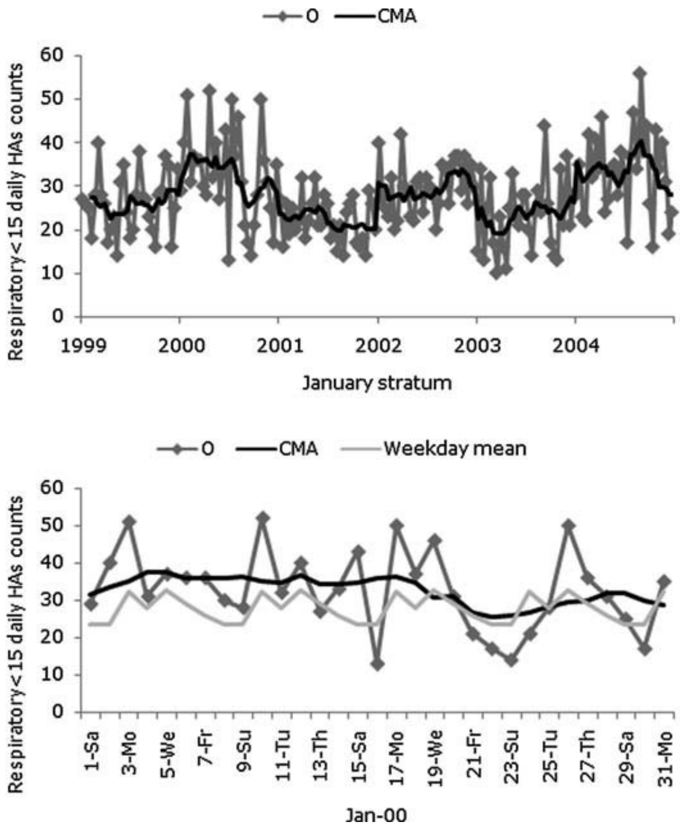

Figure 2. Sequential plot of daily levels of selected hospital admission counts of Respiratory<15 in the January stratum and close-up for the year 2000. Both the original data (O) and the CMA data are plotted as well as the mean count for each weekday type.

Firstly, if AP exposure consisted of daily episodic events intercalated by absent or sub-threshold exposures, the idea that exposures on a single day could elicit health responses on any number of days would be straightforward. However, AP exposure is permanent and usually displays a healthy degree of autocorrelation in the short term. In this scenario, is it realistic to expect that we can distinguish the effect of 1 day's exposure from the effect of exposure on neighbouring days? Moreover, not only the level, but also the duration of exposure over several days is likely to be important for health responses (Cox, 2000), and different susceptibility subgroups in the population may also differ in this respect.

Secondly, our HAs display a curious weekly pattern. As shown in Figure 3, HAs counts are lower on weekends compared to working days, and the difference is statistically significant ($p(t) < 5\%$). With the exception of NO and NO_2, APs levels on weekends and working days are not statistically different. This is not an uncommon observation. In fact, even under the extreme pollution and weather conditions of London 1952, emergency-related HAs, but not mortality, displayed drops on Sundays (Lipfert, 1993). This suggested to us that the date of admission may be quite versatile because of variations in the functioning of hospital services and variations in the time interval between the development of symptoms and actual admission (i.e. latent period) due to personal circumstances and subjective perception of well-being (induction and latent periods as defined in Chapter 4 Rothman, 2002). The weekend effect could be the clearest manifestation of the impact of such factors but there is no reason why it should not occur to some extent every day. This further suggested to us that, even if the induction periods were known, such factors could lead to unpredictable shifts in admission dates of at least 1–3 days, which could effectively blur any attempts to attribute a particular day's admission to a particular day's exposure.

In this context, the advantage of the CMA&CMA model specification is that it bypasses the need to make assumptions regarding the underlying time-scale of associations and what they represent (e.g. induction periods or exposure duration?), which may be an impossible endeavour in ecological studies involving such large heterogeneous populations and such long follow-up periods. It does so by allowing each daily observation to embody the overall weekly context in which both exposures and responses arise.

A MA window of 7 days was selected because: (1) it can give considerable flexibility for varying exposure duration requirements and latent/induction periods; (2) 7 days is still considered an acute time-scale for associations (WHO, 2005) and (3) MAs of 7 days in both APs and HAs can adjust for biases arising from weekly fluctuations, thus bypassing the need for dummy variables for weekdays. It is also advisable to choose a model specification a priori rather than choose the 'best' one by resorting to multiple testing (Lumley and Sheppard, 2000; Smith *et al*., 2000).

3.1.3. Anticipated disadvantages of the CMA&CMA model

We anticipated two limitations of the CMA&CMA model specification: (1) loss of relevant information precisely in the high-frequency range of acute effects; and (2) the introduction of autocorrelation (however non-independence of the residuals does not affect the coefficients but only their standard errors and therefore significance test). In addition, this model specification deliberately abandons any attempt to locate the associations at the daily level, as each daily observation of both the exposure and response variables are now averages of 7 days. Having these

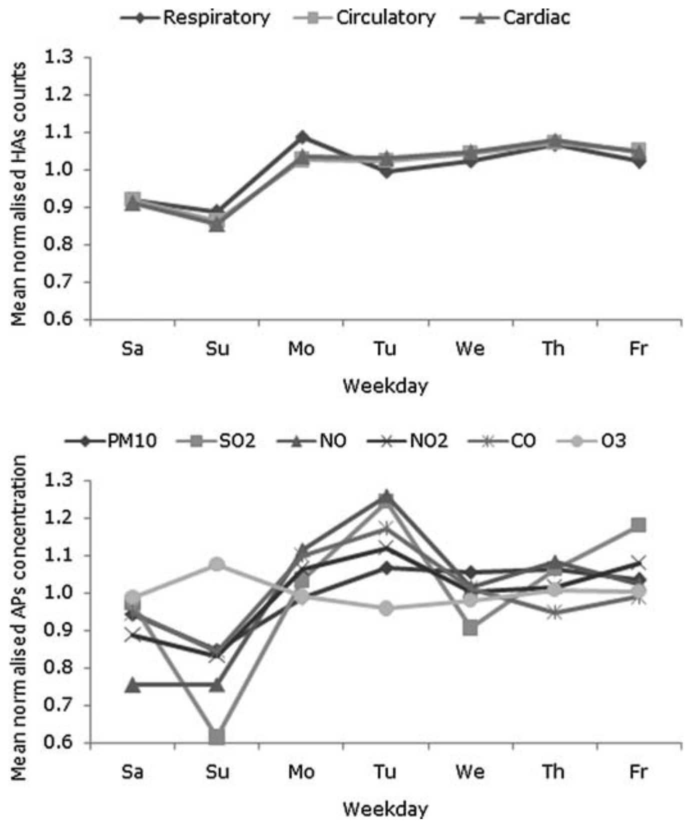

Figure 3. Mean hospital admission counts (total age-groups) (top graph) and mean APs' concentrations (bottom graph) for each weekday type in the January stratum, both normalised to mean one to help visualisation

advantages and disadvantages in mind, we proceeded to the comparison of the CMA&CMA model with the most widely used model specifications in the field.

3.2. Comparison of model specifications

Studies of the acute health effects of air pollution have used a wide range of specifications and time spans to represent exposure windows and lag-intervals. We compare most such specifications by keeping the span of the exposure window and lag-interval fixed at 7 and 0 days, respectively; except for the single-day lag model specification where the exposure window is of 1 day and the lag interval varies from 0 to 6 days.

This section begins with a comparison of the intercepts and slopes obtained from each model specification with those obtained from the CMA&CMA model. Then we proceed with an exploration of potential mathematical explanations for any differences or similarities found between model specifications.

All the results presented refer to the January stratum for the sake of brevity, since results and conclusions are unchanged when other month strata are analysed. OLS linear regression was used in all analyses and its estimates are statistically indistinguishable from those that would be obtained through maximum likelihood estimates (MLE) for Poisson or Negative Binomial regression.

3.2.1. General results

We wished to assess how the different model specifications compared to each other with respect to the intercept and slope. In order to do so, we compared each model specification at a time with the CMA&CMA model (reference model). The comparisons were performed by a simple linear regression where the y-variable was the intercepts or slopes obtained with the CMA&CMA model and the x-variable was the intercepts or slopes of one of the other alternative model specifications. From this regression, we calculated the degree of linear agreement (R^2) and slope between the intercepts (B_{int}) and slopes (B_{slopes}) obtained from each model specification relative to those obtained from the CMA&CMA model. Only those HA–AP relationships that showed a significant (1%) relationship for both model specifications being compared were considered in these regressions. The results of these comparisons are displayed in Figure 4.

In general, the slopes obtained from models with a 7 day exposure window (i.e. O&CMA and O&DLM) form a tight linear relationship that is fairly close to one, with the slopes estimated by the CMA&CMA model. On the other hand, all single-day lag models consistently estimate substatially lower slopes compared to all models that use a 7 day exposure window (i.e. CMA&CMA, O&CMA and O&DLM). A similar description applies to the comparison of estimated intercepts.

Three conclusions may be derived from Figure 4.

First, O&CMA and O&DLM models yield very similar slopes (B_{slopes} =0.92) despite the slight mismatch in the days their exposure windows include (if we replace O&CMA by O&PMA, B_{slopes} = 0.96). This finding is intuitive and reflects the equivalence between

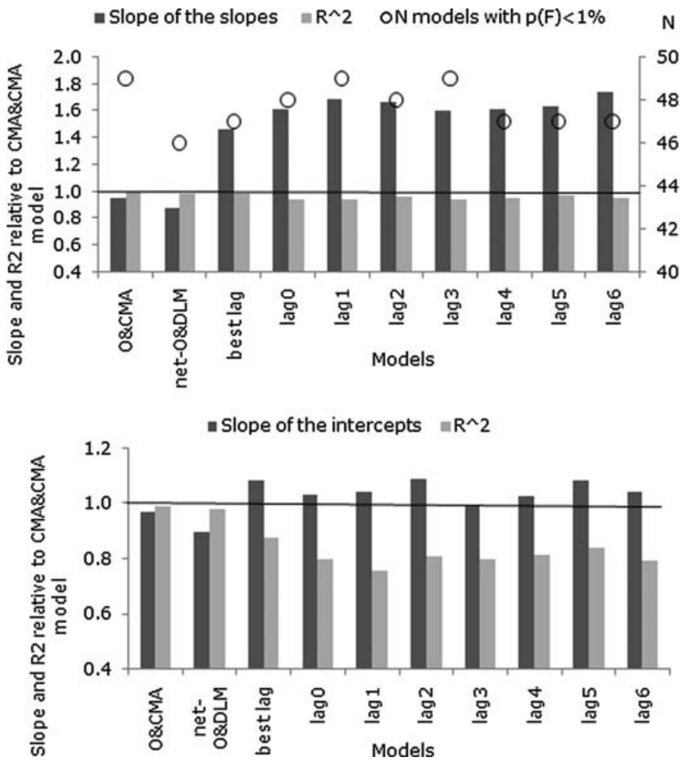

Figure 4. Relationship (evaluated by the R^2 and slope) between the slopes (top graph) and the intercepts (bottom graph) estimated by the CMA&CMA model and each of the alternative model specifications displayed on the *x*-axis. For models that included a weekend dummy variable, the mean intercept was used in the comparisons. The values of the slopes on the two graphs should be read as: increase (>1) or decrease (<1) of the slope or intercept of the CMA&CMA model per unit increase in the slope or intercept of a model on the *x*-axis. Only those HA–AP relationships that were statistically significant at 1% for each pair of models being compared were considered in the graphs; the number (N) of such HA–AP relationships is displayed in the first graph (total number of possible HA–AP relationships is 72).

'aggregating estimates or estimating aggregates' (Cox, 2000). Nevertheless, some authors appear not to have recognised this equivalence or appreciated its implications (e.g. Roberts, 2005; Braga *et al*., 2001).

Second, an exposure window of multiple days expressed either as a MA or as the net-slope of DLMs consistently yields substantially higher slopes than a single-day exposure window. This finding has been reported in numerous studies and in the next section we will attempt to find a mathematical explanation for it.

Finally, what is new in Figure 4 is that the concomitant averaging of the HAs in the CMA&CMA model, in addition to the averaging of the APs, does not lead to substantial changes in estimates compared to O&CMA and O&DLM models ($B_{slope} = 0.95$ and 0.87, respectively).

These results are qualitatively unchanged, when we replace CMA&CMA and O&CMA models by FMA&PMA and O&PMA models, respectively (not shown).

3.2.2. *Smoothing the exposure window: comparison of single-day lags with O&CMA models*

As described above, on average, slopes estimated from models that use a 7 day exposure window (CMA&CMA, O&CMA and O&DLM) are substantially higher than those obtained from single-day exposure windows (Figure 4). Because CMA&CMA models yield slopes that are very close to those obtained from O&CMA and O&DLM models, it must be concluded that the increase in slopes is mostly due to the averaging of the APs.

The tendency for models to display larger slopes as the span of the exposure window increases is a common observation in the literature (Goodman *et al*., 2004 and references therein). The prevailing explanation for this phenomenon is conceptual: longer exposure windows are able to capture single-day responses due to single-day exposures at multiple lag-intervals, where the latter are thought to reflect multiple induction periods in a heterogeneous population (Roberts, 2005; Goodman *et al*., 2004). This explanation has been the major justification for causal studies to report the optimum exposure window, i.e. the one that gives the largest slope.

We feel there are two major problems with this explanation. First, it is questionable whether it can be proven solely on the basis of ecological data. Secondly, we also find it questionable whether it is sound to compare models involving different exposure windows because: (1) although they pertain to the same exposure, averaging (CMA or DLM) leads to different datasets with different statistical properties; and (2) the associations may reflect the same phenomenon occurring at different time-scales, which may be equally valid but not directly comparable.

Assuming it is indeed sound to compare models with different exposure specifications we hypothesised whether mathematical explanations rather than the prevailing explanation described above might contribute to the fairly systematic finding that longer exposure windows yield larger slopes than shorter ones. Because O&CMA and O&DLM models yield similar estimates, we focused the comparison on O&CMA and single-day lag models. From a mathematical point of view, a MA is a smoothed version of the original dataset from which it

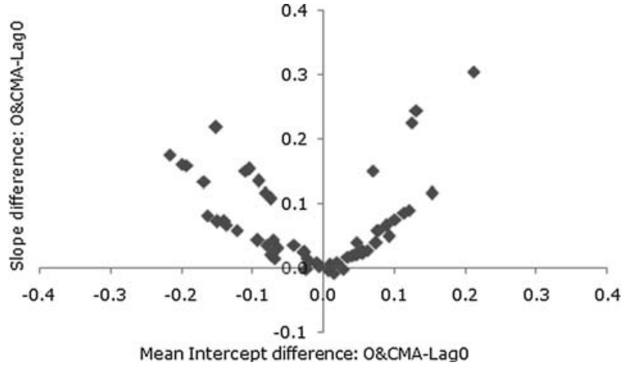

Figure 5. Relationship between the difference in mean intercept between the O&CMA model and the lag0 model (i.e. x-axis = mean intercept of O&CMA – mean intercept of lag0) with the difference in slope between the same two models (i.e. y-axis = slope of O&CMA – slope of lag0), for each HA–AP relationship that was significant for both models. The HA–AP relationships to the right side of the y-axis have negative slopes for both models whereas those to the left side have positive slopes

was calculated; therefore, the statistical properties of the MA dataset and the original dataset are rather different. Therefore, we investigated whether changes in the distribution (skewness and kurtosis) as well as in influential values and outliers between the two datasets might be associated with the different slopes. Our observations and a simulation were not able to pinpoint one of these distributional properties as a cause for the difference in slopes (not shown).

In a similar way, we hypothesised whether the reduction in the variance of the APs, induced by the O&CMA model, could have played a role. If we consider the simple case of just one independent variable, the slope is calculated by $b = \frac{\sum(x_i-\overline{X})(y_i-\overline{Y})}{\sum(x_i-\overline{X})^2} = \frac{\sum(x_i y_i)-\overline{XY}}{\sum(x_i x_i)-\overline{XX}}$. This formula suggests that a reduction in the variance of the independent variable may reduce the denominator to a greater extent than the numerator. However, it is difficult to make general statements because the extent to which the decrease in covariance might be due to the simple shrinkage of the x-values or due to an effective change in the structure of the variation of x relative to y is hard to pin down. Moreover, any generalisations are complicated by the fact that both the denominator and numerator are subtractions and a small difference between two large numbers can be very unstable.

Finally, we performed a comparison between the O&CMA and the single-day lag model (lag0) in a different way to the comparison performed in Figure 4. We calculated the difference in mean intercept between the two models and the difference in slope between the same two models and then scatter-plotted the differences. Figure 5 shows the resulting relationship. On the right hand-side of the y-axis, HA–AP relationships have a negative slope for both models being compared, whereas on the left hand-side they have positive slopes. The graph indicates that for most HA–AP relationships the greater the difference in mean intercept between the two models, the greater the difference in slope between the two models, and vice-versa. Why should such a relationship exist? A change in slopes does not necessarily have to be accompanied by a change in intercept and vice-versa.

In summary, this section attempted to explore potential mathematical explanations for the fact that models containing a 7 day exposure window display larger slopes than models with a 1 day exposure window. We have been unsuccessful at finding a definite explanation; but this may be due to the fact that we looked at the role of a single factor at a time (e.g. variance reduction or reduction of outliers or skewness) when it is likely that several factors act simultaneously to produce the change in slope. This issue should warrant further investigation in the future.

3.2.3. Smoothing the response window: comparison of CMA&CMA and O&CMA

The slopes estimated by the CMA&CMA models are slightly lower but very close to those obtained from O&CMA models ($B_{slopes} = 0.95$) (Figure 4). If we compare these models with single-day lag models, it is clear that averaging the HAs leads to a much smaller change in the slopes compared to smoothing just the APs. This result suggests two preliminary conclusions. Firstly, the original HAs dataset does not appear to contain influential values relative to the CMA HA dataset, because if it did they would be smoothed by the CMA which in turn would lead to more substantial changes in slopes. Secondly, the original HAs dataset appears to contain superfluous variation with little informative value (noise), since the variability lost by averaging the HAs into CMA does not impact the slopes to any great extent while substantially increasing the precision.

3.3. Evaluation of the CMA&CMA model

In this section we perform a more detailed comparison of the CMA&CMA model with the O&CMA model in terms of their sensitivity to noise and extreme observations in the response variable and to the deletion of single observations in the dataset.

3.3.1. Robust regression with the real dataset

Many epidemiological studies that use model specifications such as O&CMA have used M-estimation to handle influential observations in the response variable (e.g. Samoli *et al*., 2001; Schwartz, 2000b). It is unclear to us why authors prefer a robust regression method that targets extreme values in the response variable rather than methods that target extreme values in both the response and exposure variables or just in the exposure variable. Nevertheless, our aim is to compare the CMA&CMA model with the most widely used methods in the literature, which in this context would be M-estimation of the O&CMA rather than the OLS used in the previous sections.

Since both the CMA&CMA model and M-estimation share an ability to handle extreme values in the response variable, it was hypothesised that the slopes obtained from CMA&CMA models and O&CMA models might become more similar if the latter is estimated by M-estimation. Despite the apparent absence of extreme values in our HA dataset, it does appear to have a noisy pattern (Figure 2) which M-estimation may adjust for.

It is rarely stated in the literature which M-estimator was used, therefore we opted for using Huber's and Tukey's Biweight estimators in R. Figure 6 shows the relationship between the slopes obtained from the CMA&CMA model (y-axis) and the slopes obtained from the O&CMA model estimated by each robust estimator (x-axis), across statistically significant (1%) HA–AP relationships. It can be concluded that the slopes obtained by the CMA&CMA model estimated by OLS is fairly similar to the slopes obtained by the O&CMA models estimated by either robust estimator ($B_{slopes} = 0.95$ for either robust estimator). The fact that the slopes obtained by O&CMA estimated by OLS are very similar to those of O&CMA estimated by the robust estimators ($B_{slopes} = 0.99$ for Huber's and 0.97 for Tukey's) suggests that the HAs data is fairly devoid of influential values.

3.3.2. Robust regression with a simulated dataset

As explained before, both the CMA&CMA model and M-estimation are able to handle extreme values and noise in the response variable. However, this ability differs in one important aspect: while MAs are sensitive as to whether extreme values are consecutive or isolated in time (Figures 1 and 2), M-estimation is not. This ability is of importance when evaluating whether extreme values might be 'real' or errors and to what extent we are willing to allow them to influence the results of the analysis.

In order to have a clearer idea about the circumstances in which the results of the previous section might arise and about the influence of varying numbers of consecutive extreme values in the slopes obtained from the two model specifications (CMA&CMA and O&CMA) and three estimation methods (OLS, Huber's and Tukey's), we performed a simple simulation. A dataset consisting of 200 observations was used to create eight scenarios where: zero to seven consecutive observations were made extreme (the extreme values were approximately 2.5 times higher than the mean value of the 200 observations without extreme values). Figure 7 displays the slopes estimated by the two model

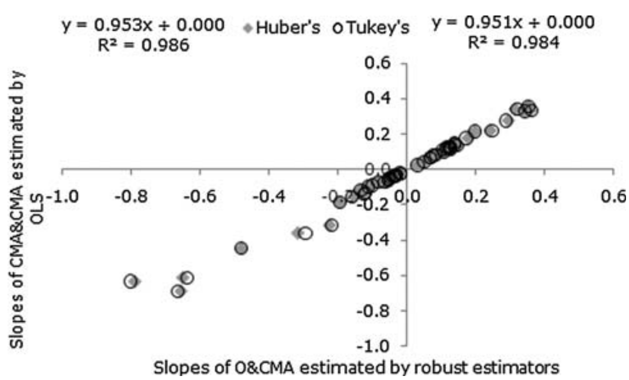

Figure 6. Relationship between the slopes obtained from the CMA&CMA model estimated by OLS (y-axis) and the slopes obtained from the O&CMA model estimated by either Huber's or Tukey's estimator (x-axis). Only those HA–AP relationships that were statistically significant at 1% for the CMA&CMA model were considered in the graph. The value of the slope and R^2 of both relationships are displayed on the graph

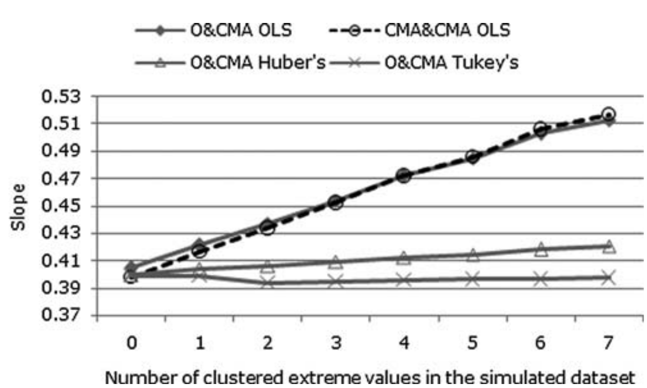

Figure 7. Slopes estimated from a simulated dataset consisting of 200 observations where 0 to 7 consecutive observations were made extreme (x-axis). The extreme observations were approximately 2.5 times higher than the mean of the dataset with 0 extreme values. The slopes for these eight scenarios are presented for the CMA&CMA model estimated by OLS, and for the O&CMA model estimated by OLS, Huber's and Tukey's estimator

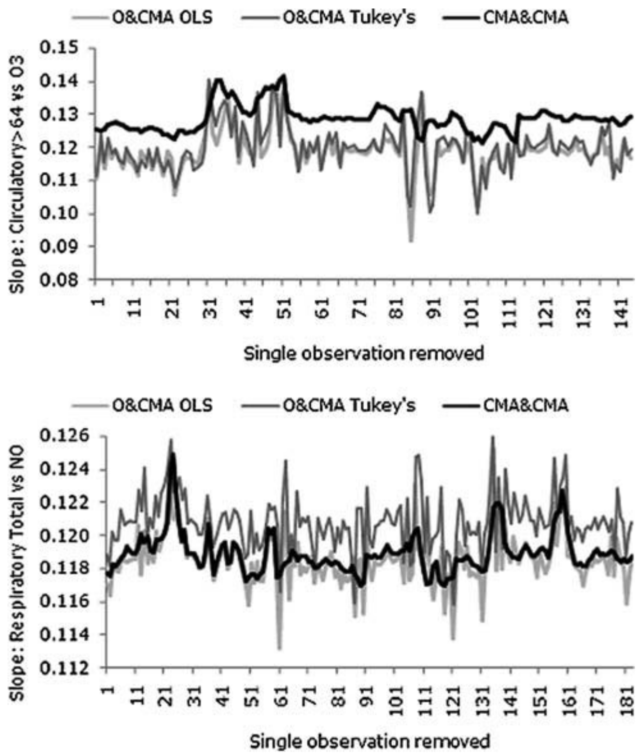

Figure 8. Change in slope when a single observation is removed across the range of observations available in the dataset (*x*-axis). This is exemplified for two HA–AP relationships: Circulatory>64 with O_3 (top graph) and respiratory total with NO (bottom graph). The slopes were obtained from the CMA&CMA model estimated by OLS and from the O&CMA model estimated with Tukey's estimator

specifications and three estimation methods, across the eight scenarios. The two robust estimators reveal a striking difference in their sensitivity to the number of extreme values: O&CMA_Tukey's estimator being the least sensitive whereas O&CMA_Huber's estimator occupies an intermediate position between O&CMA_Tukey's and O&CMA_OLS.

It is interesting to note that in the scenario where the response variable has no extreme values, the slope estimated by CMA&CMA is very similar to that obtained from O&CMA with M-estimation, or OLS estimation. This is in fact what was found with our real HA–AP dataset in the previous section. However, when the number of consecutive extreme values is low (say less than 5), the slope obtained from CMA&CMA gradually approaches the slope obtained from O&CMA_OLS; whereas, the robust estimators give comparatively lower and lower slopes. Finally, as the number of consecutive extreme values increases the slope of the CMA&CMA model gradually becomes larger and larger than the slope of O&CMA_OLS; whereas, the robust estimators remain conservative. The fact that CMA&CMA models may give slopes that are larger than those obtained from O&CMA_OLS indicates that when extreme values are abundant instead of smoothing, the CMA&CMA 'broadens' extreme values over neighbouring observations.

3.3.3. *Robustness of estimates*

In a final comparative analysis we tested an important property of any model specification: that its estimates are robust to small changes in the dataset. One way of investigating this is to remove a single observation across the range of available observations and see how that impacts the slope. This is exemplified for two HA–AP relationships in Figure 8 where it can be observed that the slopes obtained by the CMA&CMA model estimated by OLS shows much smaller fluctuations than the slopes obtained by the O&CMA model, whether the latter is estimated by OLS or by Tukey's estimator.

4. DISCUSSION

The new model specification, CMA&CMA, was conceived a priori on the basis of considerations regarding data quality and characteristics, the ecological design of the study and the processes that might interfere with acute health effects.

This model was then compared with the most widely used model specifications (namely: single-day lags, O&CMA and O&DLM models) and estimation methods (M-estimation) in the literature, by taking as a sample the 72 HA–AP relationships available. Results were presented only for the January stratum since results were unchanged for the other month strata. All analyses were performed by linear regression as this method leads to the same estimates (after an approximation) as Poisson regression and our aim is a comparison of methods, not causal inference.

To the best of our knowledge only one article has reported the use of MAs in both the response and exposure variable (Schwartz, 2000a) but the reasons for doing so and its consequences relative to other model specifications were not made explicit. Apart from Roberts (2005) who proposed the use of moving total counts of mortality as a substitute for MAs in cases where PM measurements are not available daily, no other authors have mentioned the use of aggregated response variables.

Three warning words must be given before proceeding. First, we have compared statistics from different model specifications without knowing which (if any) is the most accurate; the only criteria we are able to evaluate are the precision, relative change in estimates and the latter's robustness in the face of small changes to the dataset. It must be emphasised that, in the absence of independent experimental evidence to guide the choice of model specification, statistical-based decisions can lead to strong biases and ultimately meaningless associations (Lumley and Sheppard, 2000; Smith *et al*., 2000; Chen *et al*., 1999). Secondly, the results presented here refer to the general properties of the 5 model specifications across 72 HA-AP relationships, and they may differ for specific individual HA-AP relationships. Thirdly, the results presented here need not be reproducible in all datasets, in fact different results may be expected depending on the characteristics of the data as was shown through a simulation.

Our results show that smoothing the HAs, in addition to the APs, does not lead to loss of information at the fine time-scales where acute effects are conventionally expected to occur. This is evidenced by the fact that the CMA&CMA model yields slopes that are very close to those estimated by the O&CMA model estimated by either OLS, or M-estimation. It can be concluded that, our particular HAs dataset was fairly devoid of influential values, and for this reason both models (O&CMA and CMA&CMA) and both estimation procedures (OLS and M) gave very similar estimates.

However, a simulation revealed that such a result may only occur under certain circumstances. When the response variable has several extreme values that are consecutive in time, CMA&CMA slopes are closer to those of O&CMA_OLS and higher than those of O&CMA_robust. When the response variable has several extreme values that are not consecutive in time, CMA&CMA slopes are closer to those of O&CMA_robust and lower than those of O&CMA_OLS. This difference emphasises the fact that CMA&CMA model is able to deal with extreme values that are more likely to be ‘‘real’’ (due to their being consecutive in time) differently from those that are less likely to be ‘‘errors’’ (because they are sporadic), a property that is not shared by M-estimation which down weighs extreme values regardless of their proximity in time.

In what concerns the precision and robustness of the slopes, the CMA&CMA model outperforms the O&CMA model even if the latter is estimated by M-estimation. The increased precision and robustness of the slopes obtained from CMA&CMA models may stem from two sources: the fact that each data point is based on seven observations rather than just one and the introduction of serial correlation as the Durbin-Watson statistic is always inferior to one for the CMA&CMA dataset whereas it is usually very between one and two for the O&CMA dataset. The extent to which each of these factors contributes to the increased precision and robustness of the CMA&CMA models is difficult to disentangle.

It was shown by means of a simulated dataset that different M-estimators differ greatly in their sensitivity to the number of clustered extreme values. While this is not new, the fact that many authors do not report the M-estimator they used (e.g. Samoli *et al*., 2001; Schwartz, 2000b) allied to the fact that the major statistical packages for robust regression (R and S-PLUS) differ in their default weighing methods, appears to overlook the great impact that this decision may have on effect estimates intended for causal inference. In addition, it is important to consider that, while M-estimators have different sensitivities to the number of extreme values in the response variable they are, contrarily to MAs, completely blind as to whether they are also clustered in time. Both of these properties are important for evaluating, in the absence of additional information, whether extreme values are ‘real’ or errors.

In conclusion, the CMA&CMA model estimated by OLS has the following advantages for datasets similar to ours: (1) it yields estimates that are substantially more robust and more precise compared to other model specifications and two major M-estimators; (2) MAs are sensitive to the number of clustered extreme values, a property that not all M-estimators share; (3) MAs are sensitive as to whether extreme values are clustered in time, a property that none of the M-estimators share; (4) the CMA&CMA model is easy to use and does not require specialised software; (5) when the window of the CMA&CMA model is of 7 days, it can adjust for systematic weekly oscillations, thus avoiding the need to include additional terms for weekdays, and can adjust for effect modification (which dummy variables cannot). In the case of datasets with different properties from ours, the CMA&CMA model may serve as a tool to assess data properties and quality more closely. We had anticipated two potential disadvantages of the CMA&CMA model. First, the loss of relevant high frequency information due to the smoothing of HAs, but this was shown to be unfounded. The second was the introduction of autocorrelation in the residuals, which is confirmed as the Durbin-Watson statistic is always less than one.

The pursuit of the daily level of analysis in time-series studies appears to have grown out of convention, the availability of data on a daily basis, and the statistical advantages offered by a large dataset. In this article we have tried to introduce another criteria for the time-unit of analysis, that of data quality. There could be instances when single-day health data at the population level is simply not reliable enough and/or contains no additional information compared to aggregations over more than 1 day, even though we may have strong reasons to believe that associations do occur on a daily basis at the individual level. As shown in this study, the averaging of the HAs in time may in fact be advantageous as it increases the precision of the estimates without distortion or loss of the underlying daily signal. However, with such aggregations we lose the ability to locate or attribute the associations with single days and we introduce autocorrelation in the residuals.

Acknowledgements

This study was supported by the Fundação para a Ciência e Tecnologia–Portugal through a PhD grant (SFRH/BD/8773/2002) and a research project grant (POCI/AMB/55878/2004). We are thankful to the following Portuguese institutes who provided the data: Agência do Ambiente, Instituto de Meteorologia, Administração Central do Sistema de Saúde.

REFERENCES

Braga ALF,Saldiva PHN,Pereira LAA,Menezes JJC,Conceição GMS,Lin CA,Zanobetti A,Schwartz J,Dockery DW.2001.Health effects of air pollution exposure on children and adolescents in São Paulo, Brazil.*Pediatric Pulmonology***31**:106–113.

Chen C,Chock DP,Winkler SL.1999.A simulation study of confounding in generalised linear models for air pollution epidemiology.*Environmental Health Perspectives***107**:217–222.

Cox LH.2000.Statistical issues in the study of air pollution involving airborne particulate matter.*Environmetrics***11**:611–626.

EEA.1998.*Guidance report on preliminary assessment under EC air quality directives*. *Technical Report No 11*. Copenhagen: European Environmental Agency.

Goodman PG,Dockery DW,Clancy L.2004.Cause-specific mortality and the extended effects of particulate pollution and temperature exposure.*Environmental Health Perspectives***112**:179–185.

Lipfert FW.1993.A critical review of studies of the association between demands for hospital services and air pollution.*Environmental Health Perspectives***101**(Suppl.2):229–268.

Lumley T,Sheppard L.2000.Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses.*Environmetrics***11**:705–717.

Roberts S.2005.Using moving total mortality counts to obtain improved estimates for the effect of air pollution on mortality.*Environmental Health Perspectives***113**:1148–1152.

Rothman KJ.2002.Epidemiology: an Introduction.Oxford University Press: New York.

Samoli E,Schwartz J,Wojtyniak B,Touloumi G,Spix C,Balducci F,Medina S,Rassi G,Sunyer J,Bacharova L,Anderson HR,Katsouyami K.2001.Investigating regional differences in short-term effects of air pollution on daily mortality in the APHEA project: a sensitivity analysis for controlling long-term trends and seasonality.*Environmental Health Perspectives***109**:349–353.

Schwartz J.2000a.Harvesting and long-term exposure effects in the relation between air pollution and mortality.*American Journal of Epidemiology***151**:440–448.

Schwartz J.2000b.Assessing confounding, effect modification and thresholds in the association between ambient particles and daily deaths.*Environmental Health Perspectives***108**:563–568.

Smith RL,Davis JM,Sacks J,Speckman P,Styer P.2000.Regression models for air pollution and daily mortality: analysis of data from Birmingham, Albama.*Environmetrics***11**:719–743.

WHO.2005.Air Quality Guidelines. Global Update 2005.World Health Organization: Copenhagen.

WHO.1999.Monitoring ambient air quality for health impact assessment, *WHO Regional Series, European Series, No 85*.World Health Organization: Copenhagen.

