*Biostatistics* (2007), **8**, 2, *pp.* 337–344
doi:10.1093/biostatistics/kxl013
Advance Access publication on June 29, 2006

# On the equivalence of case-crossover and time series methods in environmental epidemiology

YUN LU*

*Department of Biostatistics, Johns Hopkins Bloomberg School of Public health, 615 North Wolfe Street, Baltimore, MD 21205-2179, USA*
ylu@jhsph.edu

SCOTT L. ZEGER

*Department of Biostatistics, Johns Hopkins Bloomberg School of Public health, 615 North Wolfe Street, Baltimore, MD 21205-2179, USA*



## SUMMARY

The case-crossover design was introduced in epidemiology 15 years ago as a method for studying the effects of a risk factor on a health event using only cases. The idea is to compare a case's exposure immediately prior to or during the case-defining event with that same person's exposure at otherwise similar "reference" times. An alternative approach to the analysis of daily exposure and case-only data is time series analysis. Here, log-linear regression models express the expected total number of events on each day as a function of the exposure level and potential confounding variables. In time series analyses of air pollution, smooth functions of time and weather are the main confounders. Time series and case-crossover methods are often viewed as competing methods. In this paper, we show that case-crossover using conditional logistic regression is a special case of time series analysis when there is a common exposure such as in air pollution studies. This equivalence provides computational convenience for case-crossover analyses and a better understanding of time series models. Time series log-linear regression accounts for overdispersion of the Poisson variance, while case-crossover analyses typically do not. This equivalence also permits model checking for case-crossover data using standard log-linear model diagnostics.

*Keywords*: Air pollution; Case-crossover design; Environmental epidemiology; Log-linear model; Overdispersion; Poisson regression; Time series.

## 1. INTRODUCTION

The case-crossover design was introduced in epidemiology 15 years ago as a method for studying the effects of a risk factor on a health event using only cases (Maclure, 1991). The idea is to compare a case's exposure immediately prior to or during the case-defining event with that same person's exposure at otherwise similar "reference" times. Each person's exposure values comprise a matched set with a single case exposure during the event interval. Conditional logistic regression (CLR) is typically used to

*To whom correspondence should be addressed.

estimate an odds ratio as the measure of association (e.g. Bateson and Schwartz, 1999). The case-crossover design is attractive because it only involves cases and each case is compared to himself/herself, thereby controlling for time-invariant personal factors.

Maclure (1991) originally proposed that only the intervals before the one in which the event occurred can be used for reference. Greenland (1996) and Navidi (1998) pointed out that this choice produces a biased odds ratio estimate in the presence of a secular trend. As an alternative, Navidi (1998) proposed the "full-stratum" design such that all intervals other than the event interval can be used for reference. Bateson and Schwartz (1999) suggested a "symmetrical bidirectional" reference window that uses control intervals equidistant shortly before and after the event to control for bias induced by long-term and seasonal trends. Lumley and Levy (2000) and Janes *and others* (2005b) have shown that in the bidirectional design, CLR gives "overlap" biased estimates of the odds ratio because the reference windows are not chosen independently of the event time. They favor the use of prespecified reference windows or "time-stratified designs" (TSDs).

The substantial statistical interest in case-crossover designs reflects its common application in many subspecialties of epidemiology, including cardiovascular disease (e.g. Koton *and others*, 2004), HIV (e.g. Schneider *and others*, 2005), accidents (Hagel *and others*, 2005), and health service quality assessment (e.g. Polevoi *and others*, 2005). The number of papers per year that include "case-crossover" in their title or keywords as identified in a Science Citation Index search has grown from 4 to 66 papers between 1993 and 2005.

This work is motivated by our group's research on the effects of air pollution on morbidity or mortality where the case-crossover method is especially popular (e.g. Dominici *and others*, 2004, 2006; Wellenius *and others*, 2005; Zanobetti and Schwartz, 2005). Case-crossover methods are used to estimate the relative rate of events per unit increase in exposure, controlling for potential confounding variables through matching. For example, Zanobetti and Schwartz (2005) applied CLR to data from each of 21 regions to study the relative risk of emergency room admission for myocardial infarction associated with PM10 exposure (particulate matter 10 μm or smaller in aerodynamic diameter). This application and many others like it are characterized by the fact that the exposure for a given day is assumed to be the same for all persons.

An alternative approach to the analysis of daily exposure and case-only data is time series analysis (e.g. Kedem and Fokianos, 2002). Here, log-linear regression models express the expected total number of events on each day as a function of the exposure level and potential confounding variables. In time series analyses of air pollution, smooth functions of time and weather are the main confounders. The smooth function of time is typically modeled using a flexible parametric or nonparametric curve to represent longer term trends in the outcome due to changes in the population, its health behaviors, and services and to represent seasonality. Zeger *and others* (2006) and Bell *and others* (2004) present overviews of time series methods in general and with application to air pollution epidemiology specifically.

The current understanding is that case-crossover methods control for potential confounding "by design" while time series methods control by modeling (Bateson and Schwartz, 2001; Janes *and others*, 2005b; Mittleman, 2005; Zanobetti and Schwartz, 2005). In this way, case-crossover analysis apparently avoids the need to control through statistical modeling.

The relative merits of time series and case-crossover studies have been discussed by several recent papers in the environmental epidemiology literature. For example, Checkoway *and others* (2000) selected the case-crossover approach as an alternative to time series methods in order to make causal inferences about air pollution effects. Bateson and Schwartz (1999, 2001) demonstrated that strong confounding by seasonality could be controlled by design in the case–control approach.

In this paper, we demonstrate that when exposure is common to the cohort at each time, as in air pollution studies, the case-crossover approach is an application of log-linear time series analysis rather than an alternative approach. This equivalence has previously been noted in special cases by Levy *and others*

(2001) and by Janes *and others* (2005a). We show how the choice of reference intervals in the case-crossover design is equivalent to the choice of estimator for the confounding function of time in the time series analysis. Given this correspondence, we offer an alternate perspective on bias of inferences from case-crossover designs. We show that inferences from case-crossover designs based upon CLR do not account for overdispersion as is routinely done in time series analyses. The connection of case-crossover and time series analyses also sheds some new light on the time series applications.

## 2. General framework

Let $X_{it}$ be the exposure for person $i$ in interval $t$, $t = 1, \ldots, T$, and let $Y_{it}$ indicate whether subject $i$ has the event in interval $t$ (1, event; 0, not). Assume that the outcome $Y_{it} = 1$ is rare and that the probability that subject $i$ fails in interval $t$ is given by the relative risk model

$$\lambda_i(t, X_{it}) = \lambda_{0it} \exp(\beta X_{it}) = \lambda_{0i} \exp(\beta X_{it} + \gamma_{it}). \tag{2.1}$$

Each subject is assumed to have his/her own baseline risk $\lambda_{0it}$ at time $t$ consisting of two parts; $\lambda_{0i}$ is a constant frailty for person $i$ and $\exp(\gamma_{it})$ is the effect of unmeasured time-varying factors on his/her risk. The exposure $X_{it}$ is assumed to have a common effect on each individual, as quantified by the log relative risk $\beta$.

For air pollution and other similar studies, the population is assumed to have common exposure during each interval so that $X_{it} = X_t$.

### 2.1 *Time series analysis*

Denote the population from which cases arise by $\mathcal{I}$; hence, the observed number of events in interval $t$ is $Y_t = \sum_{i \in \mathcal{I}} Y_{it}$. The expected number of events is the sum over the population of the individual risks,

$$\mu_t = \sum_{i \in \mathcal{I}} \lambda_i(t, X_t) = \exp(\beta X_t) \sum_{i \in \mathcal{I}} \lambda_{0it} = \exp(\beta X_t + S_t), \tag{2.2}$$

where $\exp(S_t) = \sum_{i \in \mathcal{I}} \lambda_{0it} = \sum_{i \in \mathcal{I}} \lambda_{0i} \exp(\gamma_{it})$. The target of inference is the regression coefficient $\beta$, the common log relative rate of the event per unit change in the exposure. $S_t$ is a nuisance function that is the log of the total population baseline risk on each day $t$. The total risk integrates across the population the individual baseline risks and behaviors such as exercise, smoking, and seeking health care. It also represents factors that affect the population as a whole, such as influenza epidemics or improved medical services. In time series analysis, $S_t$ is assumed to be a smooth function of time and is modeled with parametric or nonparametric curves such as regression or smoothing splines (e.g. Kelsall *and others*, 1997). Because $S_t$ is not the scientific focus, most time series investigators examine the sensitivity of inferences about the exposure relative risk $\beta$ to the choice of model for $S_t$ (e.g. Dominici *and others*, 2004).

To estimate jointly $\beta$ and $S_t$, we assume that $Y_t$ follows a log linear model with mean $E(Y_t) = \mu_t$ and $\mathrm{Var}(Y_t) = \phi \mu_t$. For any chosen estimator $\hat{S}_t$ of $S_t$, we obtain the estimate $\hat{\beta}$ by solving the following estimating equation:

$$U(\beta) = \sum_{t=1}^{T} X_t (Y_t - \mathrm{e}^{\beta X_t} \exp(\hat{S}_t(\beta))) = \sum_{t=1}^{T} X_t (Y_t - \hat{\mu}_t(\beta)). \tag{2.3}$$

Note that $\hat{\mu}_t(\beta)$ will depend on the estimate of the nuisance function $\hat{S}_t(\beta)$, which also depends on $\beta$, so that joint estimation typically involves iteration. We choose the estimate of $\beta$ that makes the observed

number of events $Y_t$ on each day $t$ on average equal to the model-based predicted value $\hat{\mu}_t(\beta)$. Inferences about $\beta$ are made robust to the Poisson assumption by allowing the variance of the data to exceed its mean using the method of "quasi-likelihood" or by using a robust variance estimator (Liang and Zeger, 1986; McCullagh and Nelder, 1989; White, 1982; Zeger, 1988).

### 2.2 *Case-crossover design*

In the case-crossover approach, the exposure of cases in interval $t_i$ is compared to the exposures from a set of reference periods. We denote the event interval by $t_i$ and the set of reference periods by $W(t_i)$. For example for day 10, $W(10)$ might be {8, 9, 10, 11, 12}. The key assumption of a case-crossover design is that the time-varying effect $\gamma_{ij}$ is constant for all $j$ within the reference window $W(t_i)$, i.e. $\gamma_{ij} = \gamma_{ij'}$ for $j, j' \in W(t_i)$.

Conditional on an individual being a case within a prespecified reference window $W(t_i)$, the probability $p_{it_i}$ that subject $i$ fails at time $t_i$ is

$$p_{it_i} = P\left(T_i = t_i | \mathbf{X}, W(t_i), \sum_{m=1}^{T} Y_{im} = 1\right) = \frac{P\left(T_i = t_i, \sum_{m=1}^{T} Y_{im} = 1 | \mathbf{X}, W(t_i)\right)}{\sum_{j \in W(t_i)} P\left(T_i = j, \sum_{m=1}^{T} Y_{im} = 1 | \mathbf{X}, W(t_i)\right)}$$

$$= \frac{\lambda_{0i} \exp\left(\beta X_{it_i} + \gamma_{it_i}\right)}{\sum_{j \in W(t_i)} \lambda_{0i} \exp(\beta X_{ij} + \gamma_{ij})} = \frac{\exp\left(\beta X_{it_i}\right)}{\sum_{j \in W(t_i)} \exp(\beta X_{ij})}, \tag{2.4}$$

which is free of the time-constant effect $\lambda_{0i}$ and time-varying effects $\gamma_{ij}$ using the case-crossover assumption that $\gamma_{ij}$ is constant for all $j$ within the reference window $W(t_i)$.

As Janes *and others* (2005b) have pointed out, this probability is not correct if the reference window depends on $t$, e.g. in the symmetric bidirectional design (SBD). However, (2.4) can still be used to construct an estimating equation for $\beta$.

If we assume subjects are independent, the likelihood function is

$$L(\beta) = \prod_{i=1}^{n} p_{it_i} = \prod_{i=1}^{n} \left[ \frac{\exp\left(\beta X_{it_i}\right)}{\sum_{j \in W(t_i)} \exp(\beta X_{ij})} \right]. \tag{2.5}$$

The estimating equation for $\beta$ is

$$U(\beta) = \sum_{i=1}^{n} U_i(\beta) = \sum_{i=1}^{n} \left[ X_{it_i} - \sum_{m \in W(t_i)} X_{im} \frac{\exp(\beta X_{im})}{\sum_{j \in W(t_i)} \exp(\beta X_{ij})} \right]. \tag{2.6}$$

This estimating equation is the sum over subjects of the difference between each subject's exposure at the index time $t_i$ and a weighted average of exposures at all times in the reference window $W(t_i)$ (Janes *and others*, 2005a). By solving (2.6), we estimate $\beta$ by the value that on average makes the relative risk weighted average of exposures on reference days equal to the exposure on the event days.

If we assume common exposure, $X_{it} = X_t$, (2.6) can be rewritten as

$$U(\beta) = \sum_{t=1}^{T} X_t \left[ Y_t - \mathrm{e}^{\beta X_t} \sum_{m \in \mathcal{R}(t)} \frac{Y_m}{\sum_{j \in W(m)} \exp(\beta X_j)} \right]$$

$$= \sum_{t=1}^{T} X_t [Y_t - \mathrm{e}^{\beta X_t + \hat{S}_t}] = \sum_{t=1}^{T} X_t (Y_t - \hat{\mu}_t^{(cc)}). \tag{2.7}$$

In the supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org), we give the derivation of (2.7). Here, $\mathcal{R}(t)$ is the set of days containing day $t$ in their reference window. For the SBD and TSD, but not more generally, this set is identical to the reference set for day $t$ itself, that is $\mathcal{R}(t) = W(t)$.

In (2.7), $\hat{S}_t$ is the weighted average of numbers of events across days $m$ that have day $t$ in their reference window. The weight for $Y_m$ is the probability of having an event on day $t$ given the reference window $W(m)$.

The case-crossover equation (2.7) is a special case of the time series equation (2.3) in which $S_t$ is estimated by a weighted average of the observed numbers of events for those intervals $m$ that include interval $t$ in their reference windows. The weights are determined by the conditional probabilities that an event occurs in $t$ given that it occurs within the window.

Two special cases are worth considering further: TSD and SBD. For TSD, time is divided *a priori* into strata $s(t) = 1, \ldots, S$. The reference window for day $t$ is the set of days in its stratum (Lumley and Levy, 2000). Levy *and others* (2001) previously pointed out that the time-stratified case-crossover design leads to the same estimate as obtained from a Poisson regression with dummy variables indicating the strata. The score equation can be written as

$$\sum_{i=1}^{n} U_i(\beta) = \sum_{t=1}^{T} X_t \left[ Y_t - \mathrm{e}^{\beta X_t} \frac{\sum_{m \in s(t)} Y_m}{\sum_{j \in s(t)} \exp(\beta X_j)} \right] = \sum_{t=1}^{T} X_t (Y_t - \hat{\mu}_t^{(a)}), \tag{2.8}$$

where $\hat{\mu}_t^{(a)} = \mathrm{e}^{\beta X_t} \sum_{t \in g(t)} Y_t / \sum_{t \in g(t)} \exp(\beta X_t)$ is the expected number of events on day $t$. Note that $\exp(\hat{S}_{s(t)}) = \sum_{t \in g(t)} Y_t / \sum_{t \in g(t)} \exp(\beta X_t)$ is the maximum likelihood estimator of $\exp(S_{s(t)})$. The smooth function of time is assumed to be a step function with a separate level of population baseline risk for each prespecified stratum. Whether to expect the total population baseline risk to change abruptly at each stratum boundary as assumed in this design is a question specific to each application. However, if it does not, assuming $S_t$ is a step function may introduce bias in the estimator or the pollution log relative risk $\beta$.

In the SBD, symmetric control days close to the event time are used. As the simplest example, define the controls as the days immediately before and after the event day. Then the score equation can be written as

$$\sum_{i=1}^{n} U_i(\beta) = \sum_{t=1}^{T} X_t \left[ Y_t - \mathrm{e}^{\beta X_t} \sum_{m=t-1,t,t+1} \frac{Y_m}{\mathrm{e}^{\beta X_{m-1}} + \mathrm{e}^{\beta X_m} + \mathrm{e}^{\beta X_{m+1}}} \right] = \sum_{t=1}^{T} X_t (Y_t - \hat{\mu}_t^{(b)}).$$

This is equivalent to using a locally weighted running-mean smoother to estimate $S_t$ in time series analysis.

## 3. Discussion

This paper has shown that the CLR estimating equation used to obtain the case-crossover estimate of relative risk is a special case of the time series log-linear model estimating equation when exposure is common across subjects in each interval. Time series and case-crossover analyses simply offer different parameterizations for $S_t$.

The time-stratified case-crossover design is equivalent to Poisson regression with indicator variables for strata (Levy *and others*, 2001). The smooth function of time $S_t$ is assumed to be a step function with different levels of total population baseline risk for each stratum. The symmetric bidirectional case-crossover design is equivalent to Poisson regression using a locally weighted running-mean smoother to estimate $S_t$.

The equivalence of the case-crossover and time series methods improves our understanding of both methods and provides computational convenience. Most case-crossover analyses use CLR for estimation.

When the number of time intervals and the number of controls for each case are large (e.g. full-stratum design), standard CLR is computationally inefficient by comparison with Poisson regression.

Each case-crossover design corresponds to a model (or estimator) for $S_t$. The equivalence of case-crossover and time series methods permits model checking for case-crossover data using standard log-linear model diagnostic tools (McCullagh and Nelder, 1989).

When the same estimating equation is used for a time series and case-crossover analysis, that is, the same estimator of $S_t$ is used, the two methods can give different standard errors. This is because time series analysis allows for overdispersion of the Poisson variance, while case-crossover design uses the exact Poisson variance to calculate the standard error. In some applications, the Poisson assumption may not be valid.

This connection also informs our interpretation of time series analysis. For example, in Dominici *and others* (2004) time series models are used to estimate a PM effect on daily mortality. The degrees of freedom to estimate $S_t$ with a regression spline are allowed to vary ninefold from 2.3 to 21 degrees of freedom per year, yet the standard error of the pollution effect changes little. For matched case–control studies, there is little change in the standard error when the number of exactly matched controls per case is beyond roughly four (McCullagh and Nelder, 1989). In a case-crossover design, this corresponds to four control days per event day, or equivalently 90 degrees of freedom per year, which is much greater than the entire range included by Dominici *and others* (2004). This point only considers precision; the actual choice of degrees of freedom is obviously a trade-off between bias and precision.

The connection between case-crossover estimates obtained by CLR and by time series methods is an example of the connection between logistic and log-linear Poisson regression (McCullagh and Nelder, 1989). A related connection is between Poisson regression and the Cox proportional hazards model with time-invariant covariates estimated by CLR, as discussed by Clayton (1988, 1991). The Cox model is approximated by a log-linear Poisson model for the number of events in small intervals of follow-up time. The number of events is regressed on the covariates plus indicator variables for bins with log person-time in each bin as an offset. Clayton exploits this connection to develop Bayesian formulations of frailty and other extensions of the basic Cox model.

This connection is also apparent in our work whereby a hazard model with individual frailties is the basis for a log-linear regression for the binned event counts. The overdispersion in our time series model reflects the influence of unmeasured causes of mortality that vary over time in a manner that is not accounted for by the assumed model for $S_t$. These factors are population analogues to frailties in the survival context.

In this paper, we only focused on exposures common to all subjects. In many applications of the case-crossover design, exposures vary among subjects. The connection between case-crossover and time series method in this case is the topic of further study.

## ACKNOWLEDGMENTS

The authors are grateful to partial support from the National Institute for Environmental Health Sciences (NIEHS) grant ES012054-03 and the NIEHS Center in Urban Environmental Health grant P30 ES 03819. *Conflict of Interest:* None declared.

## REFERENCES

BATESON, T. F. AND SCHWARTZ, J. (1999). Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. *Epidemiology* **10**, 539–44.

BATESON, T. F. AND SCHWARTZ, J. (2001). Selection bias and confounding in case-crossover of environmental time-series data. *Epidemiology* **12**, 654–61.

BELL, M. L., SAMET, J. M. AND DOMINICI, F. (2004). Time-series studies of particulate matter. *Annual Review of Public Health* **25**, 247–80.

CHECKOWAY, H., LEVY, D., SHEPPARD, L., KAUFMAN, J., KOENIG, J. AND SISCOVICK, D. (2000). A case-crossover analysis of fine particulate matter air pollution and out-of-hospital sudden cardiac arrest. *Research Report 99*. Boston, MA: Health Effects Institute.

CLAYTON, D. (1988). The analysis of event history data: a review of progress and outstanding problems. *Statistics in Medicine* **7**, 819–41.

CLAYTON, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47**, 467–85.

DOMINICI, F., MCDERMOTT, A. AND HASTIE, T. J. (2004). Improved semiparametric time series models of air pollution and mortality. *Journal of the American Statistical Association* **99**, 938–48.

DOMINICI, F., PENG, R. D., BELL, M. L., PHAM, L., MCDERMOTT, A., ZEGER, S. L. AND SAMET, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Journal of the American Medical Association* **295**, 1127–34.

GREENLAND, S. (1996). Confounding and exposure trends in case-crossover and case-time-control designs. *Epidemiology* **7**, 231–9.

HAGEL, B. E., PLESS, I. B., GOULET, C., PLATT, R. W. AND ROBITAILLE, Y. (2005). Effectiveness of helmets in skiers and snowboarders: case-control and case crossover study. *British Medical Journal* **330**, 281–3.

JANES, H., SHEPPARD, L. AND LUMLEY, T. (2005a). Case-crossover analyses of air pollution exposure data: referent selection strategies and their implications for bias. *Epidemiology* **16**, 717–26.

JANES, H., SHEPPARD, L. AND LUMLEY, T. (2005b). Overlap bias in the case-crossover design, with application to air pollution exposures. *Statistics in Medicine* **24**, 285–300.

KEDEM, B. AND FOKIANOS, K. (2002). *Regression Models for Time Series Analysis*. Hoboken: John Wiley & Sons, Inc.

KELSALL, J. E., SAMET, J. M., ZEGER, S. L. AND XU, J. (1997). Air pollution and mortality in Philadelphia, 1974–1988. *American Journal of Epidemiology* **146**, 750–62.

KOTON, S., TANNE, D., BORNSTEIN, N. M. AND GREEN, M. S. (2004). Triggering risk factors for ischemic stroke—a case-crossover study. *Neurology* **63**, 2006–10.

LEVY, D., LUMLEY, T., SHEPPARD, L., KAUFMAN, J. AND CHECKOWAY, H. (2001). Referent selection in case-crossover analyses of acute health effects of air pollution. *Epidemiology* **12**, 186–92.

LIANG, K. Y. AND ZEGER, S. L. (1986). Longitudinal data-analysis using generalized linear-models. *Biometrika* **73**, 13–22.

LUMLEY, T. AND LEVY, D. (2000). Bias in the case-crossover design: implications for studies of air pollution. *Environmetrics* **11**, 689–704.

MACLURE, M. (1991). The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* **133**, 144–53.

MCCULLAGH, P. AND NELDER, J. A. (1989). *Generalized Linear Models*, 2nd edition. London: Chapman & Hall.

MITTLEMAN, M. A. (2005). Optimal referent selection strategies in case-crossover studies—a settled issue. *Epidemiology* **16**, 715–6.

NAVIDI, W. (1998). Bidirectional case-crossover designs for exposures with time trends. *Biometrics* **54**, 596–605.

POLEVOI, S. K., QUINN, J. V. AND KRAMER, N. R. (2005). Factors associated with patients who leave without being seen. *Academic Emergency Medicine* **12**, 232–6.

SCHNEIDER, M. F., GANGE, S. J., MARGOLICK, J. B., DETELS, R., CHMIEL, J. S., RINALDO, C. AND ARMENIAN, H. K. (2005). Application of case-crossover and case-time-control study designs in analyses of time-varying predictors of T-cell homeostasis failure. *Annals of Epidemiology* **15**, 137–44.

WELLENIUS, G. A., BATESON, T. F., MITTLEMAN, M. A. AND SCHWARTZ, J. (2005). Particulate air pollution and the rate of hospitalization for congestive heart failure among medicare beneficiaries in Pittsburgh, Pennsylvania. *American Journal of Epidemiology* **161**, 1030–6.

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrika* **50**, 1–26.

ZANOBETTI, A. AND SCHWARTZ, J. (2005). The effect of particulate air pollution on emergency admissions for myocardial infarction: a multicity case-crossover analysis. *Environmental Health Perspectives* **113**, 978–82.

ZEGER, S. L. (1988). A regression model for time series of counts. *Biometrika* **75**, 621–9.

ZEGER, S. L., IRIZARRY, R. A. AND PENG, R. D. (2006). On time series analysis of public health and biomedical data. *Annual Review of Public Health* **27**, 57–79.

[*Received March 11, 2006; first revision May 31, 2006; second revision June 15, 2006; third revision June 21, 2006; accepted for publication June 27, 2006* ]