

# On the Relationship between Time-Series Studies, Dynamic Population Studies, and Estimating Loss of Life Due to Short-Term Exposure to Environmental Risks

Richard T. Burnett,<sup>1</sup> Anup Dewanji,<sup>2</sup> Francesca Dominici,<sup>3</sup> Mark S. Goldberg,<sup>4</sup> Aaron Cohen,<sup>5</sup> and Daniel Krewski<sup>6</sup>

<sup>1</sup>Biostatistics and Epidemiology Division, Safe Environments Directorate, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario, Canada; <sup>2</sup>Applied Statistics Unit, Indian Statistical Institute, Kolkata, India; <sup>3</sup>Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; <sup>4</sup>Department of Medicine and Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada; <sup>5</sup>Health Effects Institute, Boston, Massachusetts, USA; <sup>6</sup>Department of Epidemiology and Community Medicine, R. Samuel McLaughlin Center for Population Risk Assessment, University of Ottawa, Ottawa, Ontario, Canada

There is a growing concern that short-term exposure to combustion-related air pollution is associated with increased risk of death. This finding is based largely on time-series studies that estimate associations between daily variations in ambient air pollution concentrations and in the number of nonaccidental deaths within a community. Because these results are not based on cohort or dynamic population designs, where individuals are followed in time, it has been suggested that estimates of effect from these time-series studies cannot be used to determine the amount of life lost because of short-term exposures. We show that results from time-series studies are equivalent to estimates obtained from a dynamic population when each individual's survival experience can be summarized as the daily number of deaths. This occurs when the following conditions are satisfied: *a*) the environmental covariates vary in time and not between individuals; *b*) on any given day, the probability of death is small; *c*) on any given day and after adjusting for known risk factors for mortality such as age, sex, smoking habits, and environmental exposures, each subject of the at-risk population has the same probability of death; *d*) environmental covariates have a common effect on mortality of all members of at-risk population; and *e*) the averages of individual risk factors, such as smoking habits, over the at-risk population vary smoothly with time. Under these conditions, the association between temporal variation in the environmental covariates and the survival experience of members of the dynamic population can be estimated by regressing the daily number of deaths on the daily value of the environmental covariates, as is done in time-series mortality studies. Issues in extrapolating risk estimates based on time-series studies in one population to estimate the amount of life lost in another population are also discussed. **Key words:** air pollution, dynamic population design, hazard function, mortality, Poisson regression, survival analysis, time-series studies.

Short-term exposure to ambient concentrations of combustion-related pollution has become a public health concern over the last decade largely because of numerous studies linking fluctuations in the daily number of deaths with daily variations in ambient air pollutants, such as particulate matter and ground-level ozone (Dominici et al. 2000; Katsouyanni et al. 1997; Moolgavkar 2002; Schwartz and Lee, 1999; Strieb et al. 2002). However, there has been considerable debate in interpreting the results from these studies in terms of estimating loss of life expectancy (McMichael et al. 1998). It has been argued that because no identifiable population of subjects was followed over time, the regression parameters obtained from the time-series studies could not be applied to population life-tables to determine the amount of life lost because of short-term pollution exposure (Kunzili et al. 2001; Rabl 2003).

Associations between daily values of pollution and mortality have been extensively examined using time-series study designs. Here, the daily variations in the number of deaths in a community are related to daily variations of

ambient air pollution levels using generalized additive models (Hastie and Tibshirani 1990), controlling for temporal trends, seasonality, and weather variables (Kelsall et al. 1999; Samet et al. 2000; Schwartz 1999). The time scale of exposure is on the order of a few days and therefore represents short-term exposure. Cumulative exposures have also been examined up to 2 months using distributive lag models (Zanobetti et al. 2001), frequency domain (Kelsall et al. 1999; Zeger et al. 1999), and time-domain regression techniques (Dominici et al. 2003). Time-series approaches have limitations in examining the effects of pollution exposure on mortality at longer time scales because these longer temporal averages of daily pollution values tend to be highly correlated with the seasonal component of the mortality time series and weather variables, making inference highly unstable.

In time-series studies, a common value of daily exposure is assigned to all subjects residing in a single community, and variations in exposure between subjects are omitted. Recently, time-series study design has been extended to

include multiple communities (Dominici et al. 2000; Katsouyanni et al. 1997, 2001), where community-specific estimates of effect are obtained from data in each community separately and then pooled among communities using random effects models (Burnett et al. 1995; Dominici et al. 2002; Katsouyanni et al. 1997, 2001).

The effects of longer-term exposure scenarios on mortality can be examined by recording the survival experience of subjects in multiple communities adjusted for subject-specific risk factors. This design generates exposure variation between subjects that live in different communities and has been used in several studies (Dockery et al. 1993; Pope et al. 1995). Unlike time-series studies, a time-invariant multiyear population average exposure is assigned to all subjects residing in a single community. Thus, exposure is measured across communities even though individuals are followed in time.

Estimates of relative risk can be obtained from multicommunity cohort studies of air pollution and mortality (Burnett et al. 2001; Dockery et al. 1993; Krewski et al. 2000; Pope et al. 1995, 2002), in which spatial variations in ambient air pollution are related to spatial variations in survival, after adjusting for covariates obtained at the individual level, such as smoking habits. Estimates of relative risk from such studies have been incorporated into population life tables to determine the amount of life lost because of long-term exposure to ambient pollution (e.g., Brunekreef 1997). A common risk of death due to air pollution exposure and a common probability of death for those not exposed are implicitly assumed at any given age in this approach to estimating loss of life expectancy.

---

Address correspondence to R.T. Burnett, Biostatistics and Epidemiology Division, Safe Environments Directorate, Healthy Environments and Consumer Safety Branch, Health Canada, 4th Floor, Sir Charles Tupper Bldg., 2720 Riverside Dr., Ottawa, Ontario, Canada K1A 0K9. Telephone: (613) 952-1364. Fax: (613) 941-3883. E-mail: rick\_burnett@hc-sc.gc.ca

The views expressed in this manuscript are those of the authors and do not necessarily reflect those of the Health Effects Institute or its sponsors.



However, to place the effect of air pollution episodes in a population health perspective, it is also of interest to assess the impacts on longevity from short-term pollution exposures on the order of days to weeks.

In this article, we propose a survival model that jointly examines the effects of short- and long-term exposure to environmental risk factors on mortality. We also identify the conditions under which the effect estimates on survival associated with short-term exposures are equivalent to effect estimates from time-series studies. We identify sampling design characteristics under which daily time-series studies can be used to estimate the amount of life lost because of short-term exposure to environmental risk factors.

### Dynamic Population Studies

Consider a dynamic population design with the time to response defined as calendar time. Subjects are followed as long as they reside in a given community. Subjects can enter the study population either at birth or by immigration and leave the study population through death, emigration, or termination of follow-up. Further, suppose information is available on a subject's age at entry, sex, and race. In addition, subjects are interviewed periodically to obtain information on smoking habits, diet, occupation, education, and any other risk factors related to mortality. Measurements on several other environmental risk factors may also be available, such as ambient air pollution, aerobiologics, and weather.

The value of the environmental covariates recorded on day  $t$  for  $T$  days for the  $i$ th individual in the  $s$ th of  $S$  communities,  $z_s(i, t)$ , can be decomposed into three factors of the form

$$z_s(i, t) = [z_s(i, t) - z_{i_s}(\bar{t})] + [z_{i_s}(\bar{t}) - z_{i_s}(\bar{t})] + z_{i_s}(\bar{t}). \quad [1]$$

Here,  $z_{i_s}(\bar{t})$  is the average over the at-risk population in community  $s$ , and  $z_{i_s}(\bar{t})$  is the average of  $z_s(i, t)$  over a long period of time, such as years to decades, for the  $s$ th community on day  $t$ . The first term on the right side of Equation 1,  $P_{i_s}(t) = z_s(i, t) - z_{i_s}(\bar{t})$ , is the difference between the exposure value for the  $i$ th individual in the  $s$ th community and the average value on that day for all members of the community and represents the within-community variation in personal exposure. The second term,  $A_s(t) = z_{i_s}(\bar{t}) - z_{i_s}(\bar{t})$ , is the difference between the community average of personal exposures on day  $t$  and the long-term average for community  $s$ . Temporally varying exposure measures such as  $A_s(t)$  have been used in time-series studies with the population-average personal exposure values replaced by spatial averages of available fixed-site monitoring data. The spatially variable exposure

measures such as  $z_{i_s}(\bar{t})$  have been used in cohort studies of air pollution and mortality.

Further temporal decompositions of exposure can be made; for example, one could consider variation in daily averages within a month, monthly averages within a year, yearly averages within a decade, and so on (Dominici et al. 2002; Schwartz 1999; Zeiger et al. 1999). For the sake of simplicity of presentation, we restrict our discussion to a three-factor temporal decomposition given by Equation 1.

The relationship between the risk factors and survival is modeled by the hazard function  $\lambda_{i_s}^{(l)}(t)$  for the  $i$ th subject in the  $l$ th stratum in the  $s$ th community:

$$\lambda_{i_s}^{(l)}(t) = \lambda_0^{(l)}(t) \exp[\delta'_l \mathbf{x}_{i_s}^{(l)}(t) + \phi' P_{i_s}^{(l)}(t) + \beta' A_s(t) + \theta' z_{i_s}(\bar{t})] \quad [2]$$

where  $\lambda_0^{(l)}(t)$  is the baseline hazard function,  $\mathbf{x}_{i_s}^{(l)}(t)$  is a time-dependent vector of individual risk factors for the  $i$ th subject in the  $l$ th stratum in the  $s$ th community, and  $\phi$ ,  $\beta$ ,  $\theta$  are vectors of unknown parameters representing the logarithm of the relative risks associated with a unit change in the within-community between-individual variation in personal exposure [ $P_{i_s}(t)$ ], within-community temporal variation in population average exposure [ $A_s(t)$ ], and variation in the long-term average exposure between communities [ $z_{i_s}(\bar{t})$ ], respectively. We assume that  $\phi$ ,  $\beta$ , and  $\theta$  are constant for all strata. The  $\delta_l$  values are the log-relative risk for the individual risk factors that can vary by stratum.

Strata could be defined by groupings of age at entry, sex, and race, thereby allowing the baseline hazard function to depend on these risk factors. As a result, the effect of these risk factors on survival cannot be estimated. Note that we have indexed the personal exposure values by strata in order to uniquely identify subjects,  $P_{i_s}^{(l)}(t)$ . These values are incorporated into the hazard function in the same manner as the other individual-level risk factors,  $\mathbf{x}_{i_s}^{(l)}(t)$ .

We have assumed that each subject within a stratum has the same baseline hazard function,  $\lambda_0^{(l)}(t)$ , and that the association between exposure and survival is identical for all subjects. We therefore cannot distinguish subjects in terms of their sensitivity to die after adjusting for the known risk factors and strata. This model is referred to as a homogeneous survival model. We explore implications of this assumption on the ability to estimate loss in life expectancy, and some extensions to a heterogeneous survival model, in the "Discussion."

Our parameter of interest is  $\beta$ , which estimates the effect of temporal variation in exposure within a community,  $A_s(t)$ , on survival. The study design considered here can be described as following individuals' survival

experience over time within each community. Estimates of effects of the environmental covariates on survival can be made within each community separately, and a summary estimate of effect is given by pooling the community-specific estimates among communities. The longer-term average exposure values,  $z_{i_s}(\bar{t})$ , can then be absorbed into the baseline hazard function to form a community-specific baseline hazard of the form  $\lambda_{i_s}^{(l)}(t) = \lambda_0^{(l)}(t) \exp[\theta' z_{i_s}(\bar{t})]$ . This is a reasonable assumption because both  $\lambda_0^{(l)}(t)$  and  $\exp[\theta' z_{i_s}(\bar{t})]$  are expected to be smooth or slowly changing functions of time.

We consider estimation of  $\beta$  by using information from a single community, and we therefore omit the index  $s$ , for the sake of simplifying the notation. Of course, data from across communities can be combined for a pooled estimate of risk (Burnett et al. 1995; Dominici et al. 2002; Katsouyanni et al. 1997, 2002). Estimates of the baseline hazard function and regression parameters may be obtained by maximum likelihood methods. Under Equation 2, the log-likelihood function of  $\beta$ ,  $l(\beta)$ , is given by Cox and Oakes (1984)

$$l(\beta) \propto \sum_{l=1}^L \sum_{t=t_0}^T \left\{ \sum_{i \in D^{(l)}(t)} \ln \lambda_i^{(l)}(t) - \sum_{i \in A^{(l)}(t) \cup D^{(l)}(t)} \int_{t_0}^{t_i} \lambda_i^{(l)}(u) du \right\} \quad [3]$$

where  $t_0$  is the starting date of the study,  $T$  is the end date,  $t_i^{(l)}$  is the time the  $i$ th subject in the  $l$ th stratum entered the community, and  $D^{(l)}(t)$  and  $C^{(l)}(t)$  are the sets of individuals who died or were lost to follow-up on day  $t$  in the  $l$ th stratum, respectively. We define the population at risk by  $R^{(l)}(t) = D^{(l)}(t) \cup C^{(l)}(t)$ .

Assuming that the covariate values are constant within a day and writing the limits of integration in Equation 3 as a sum of integrals between consecutive days, we can rewrite the log-likelihood (Equation 3) as

$$l(\beta) \propto \sum_{l=1}^L \sum_{t=t_0}^T \left\{ y^{(l)}(t) [\ln \lambda_0^{(l)}(t) + \beta' A(t)] + \left( \sum_{i \in D^{(l)}(t)} \delta'_l \mathbf{x}_i^{(l)}(t) + \phi' P_{i_s}^{(l)}(t) - \Lambda_0^{(l)}(t) \mathfrak{R}^{(l)}(t) e^{\beta' A(t)} \right) \right\} \quad [4]$$

where  $y^{(l)}(t)$  is the number of subjects who died on day  $t$  in the  $l$ th stratum,  $\Lambda_0^{(l)}(t) = \int_{t_0}^t \lambda_0^{(l)}(u) du$  is the cumulative baseline hazard function, and

$$\mathfrak{R}^{(l)}(t) = \sum_{i \in R^{(l)}(t)} e^{\delta'_l \mathbf{x}_i^{(l)}(t) + \phi' P_{i_s}^{(l)}(t)}$$

is the effect of individual covariates on survival averaged over the population at risk. Algebraic



calculations that lead from Equation 3 to Equation 4 are detailed in the Appendix. We are interested primarily in estimating the association between  $A(t)$  and survival,  $\beta$ , while treating as nuisance parameters the underlying hazard functions,

$\{\Lambda_0^{(l)}(t), l = 1, \dots, L\}$ ; and the effects of the individual covariates,  $\delta_l$ , and personal exposure,  $\phi$ , on survival. The maximum likelihood estimate of  $\beta$  can be obtained as the solution to the score function equation

$$\begin{aligned} \frac{\partial l(\beta)}{\partial \beta} &= S(\beta) \\ &= \sum_{t=t_0}^T A(t) [y(t) - \Theta(t)e^{\beta' A(t)}] \\ &= 0 \end{aligned} \tag{5}$$

where

$$\Theta(t) = \sum_{l=1}^L \Lambda_0^{(l)}(t) \mathfrak{R}^{(l)}(t)$$

and

$$y(t) = \sum_{l=1}^L y^{(l)}(t).$$

The function  $\Theta(t)$  represents the daily baseline hazard function multiplied by the effects of time-varying individual covariates averaged over the at-risk population. First, it is reasonable to model  $\mathfrak{R}^{(l)}(t)$  as a smooth function in time because it is a summation of the individual covariate effects over the at-risk population in the  $l$ th strata. For example, the effect of the number of smoked cigarettes on survival may vary markedly from day to day for any single individual, but the average effect of smoking on survival in the population at risk should vary relatively smoothly over time. Second, we would also expect to model the cumulative baseline hazard function,  $\Lambda_0^{(l)}(t)$ , as a smooth function of time. Therefore, it is reasonable to assume that  $\Theta(t)$  is a smooth function of time.

Finally,  $\Theta(t)\exp[\beta' A(t)]$  can be interpreted as the expected number of deaths on day  $t$ . To show this, we note that the conditional probability of dying on day  $t$  for the  $i$ th subject in the  $l$ th stratum is given by

$$1 - \exp\left(-\int_{t-1}^t \lambda_i^{(l)}(u) du\right) \approx \int_{t-1}^t \lambda_i^{(l)}(u) du \tag{6}$$

(Cox and Oakes 1984), with the approximation being reasonable because of the small daily death rate in North American and European cities ( $\approx 10^{-5}$ ). The expected number of deaths on day  $t$  is given by the probability of death summed over all individuals at risk in the community on day  $t$ , or

$$\sum_{l=1}^L \sum_{i \in \mathfrak{R}^{(l)}(t)} \int_{t-1}^t \lambda_i^{(l)}(u) du = \Theta(t)e^{\beta' A(t)}. \tag{7}$$

### Time-Series Model

Several investigators have estimated associations between daily variations in population average exposure such as ambient air pollution and mortality using a time-series approach (Katsouyanni et al. 1997; Samet et al. 2000). This approach assumes that the number of daily deaths in the  $l$ th stratum,  $y^{(l)}(t)$ , follows a Poisson distribution with the expected number of deaths on day  $t$  given by

$$E\{y^{(l)}(t)\} = \Phi^{(l)}(t)e^{\gamma' A(t)}, \tag{8}$$

where  $\Phi^{(l)}(t)$  is the daily baseline number of deaths for the nonexposed population in stratum  $l$ , and  $\gamma$  is the relative rate of daily mortality for a unit change in  $A(t)$ , assumed identical for all strata. Assuming that the counts are independent among strata, the log-likelihood function of  $\gamma$  is proportional to

$$\begin{aligned} l(\gamma) \propto \sum_{l=1}^L \sum_{t=t_0}^T \{y^{(l)}(t) [\ln \Phi^{(l)}(t) + \gamma' A(t)] \\ - \Phi^{(l)}(t)e^{\gamma' A(t)}\}. \end{aligned} \tag{9}$$

The score function for  $\gamma$ ,  $S(\gamma)$ , is given by

$$S(\gamma) = \sum_{t=t_0}^T A(t) [y(t) - \Phi(t)e^{\gamma' A(t)}], \tag{10}$$

where  $\Phi(t) = \sum_{l=1}^L \Phi^{(l)}(t)$  is the daily baseline mortality averaged across strata.

### Comparison of the Two Models

Score Equations 5 and 10 suggest that if the population-average baseline hazard function times the average effect of individual-level covariates on survival for the population at risk,  $\Theta(t)$ , and the population average baseline mortality,  $\Phi(t)$ , are both modeled using the same unknown function of time, then our modeling approaches to the dynamic cohort and time-series designs provide identical estimates of the effects of environmental covariates. This analytical approach is reasonable because both quantities represent the expected number of deaths on day  $t$  when  $A(t) = 0$  in their respective designs.

Commonly, in the analysis of time series,  $A(t)$  is estimated by the daily average of the concentrations observed at fixed-site ambient monitoring stations. For some air pollutants such as fine particulate matter, aggregated measures provide reasonable estimates of the population average of personal exposure values (Zeger et al. 2000). Biased estimates of the effects of the environmental exposures will occur because of measurement error if these aggregate measures provide poor estimates of the average of the personal exposures of the at-risk group with the amount of bias dependent on the amount of error in measuring exposure (Zidek et al. 1996).

### Discussion

We have demonstrated that dynamic population study and time-series designs provide the same relative rate estimates of mortality associated with exposure to air pollution under the following conditions: *a*) the environmental covariates vary in time and not between individuals; *b*) on any given day, the probability of death is small; *c*) each subject of the at-risk population has the same probability of death after adjusting for known risk factors; *d*) all members of at-risk population share a common effect of environmental covariates on mortality; and *e*) the population-average baseline hazard function and association between risk factors and death can be approximated adequately by smooth functions of time. In other words, if conditions *a*–*e* hold, then each individual's survival experience can be summarized as the daily number of deaths.

In addition, the equivalency of the estimates of  $\beta$  and  $\gamma$ , obtained from  $S(\beta)$  and  $S(\gamma)$ , respectively, depend on whether  $\Theta(t)$  and  $\Phi(t)$  are modeled as the same nonstochastic function, possibly involving some unknown parameters. A challenge of time-series studies is the lack of a clear-cut method to choose the smooth time function to eliminate long-term and seasonal trends in the data, and different estimation methods can lead to different results. For example, we have suggested previously (Goldberg et al. 2000) that the smooth function be selected so that the residual time series is consistent with a white noise process. It seems clear now that estimates of the air pollution effect are sensitive to the method of modeling time and weather, although this sensitivity can vary by location and season depending on how these variables are correlated.

Although we have estimates of the effects of long-term exposure to ambient air pollution on survival (Dockery et al. 1993; Pope et al. 1995) based on variations in exposure between communities and estimates of shorter-term exposure on mortality (Samet et al. 2000) based on daily variations in exposure within a community, we as yet have no direct estimates of the total effect of exposure to ambient air pollution from all time scales based on the same study. The sum of these effects gives estimates for two of the three components described in Equation 1. A few studies have attempted to estimate the effects of personal exposure to air pollution on mortality. Variations in personal exposure estimates are generated as a function of a subject's residence within a community or geographic region (Abbey et al. 1999). One could sum the estimates of effect from these time-series mortality and cohort studies to obtain a total estimate of effect. However, they are based on different populations and exposure data. The time-series studies use mortality data covering the entire population, whereas the cohort studies



are not necessarily representative of the target population. It is desirable to obtain joint estimates of risk based on personal variation of exposure within a community, temporal variation within a community, and spatial variation between communities obtained from a multi-community dynamic population study using a unified survival model. Model formulations for a joint analysis of time-series and cohort studies have been recently discussed (Zeger et al. In press).

We have considered values only of the environmental covariates defined on a single day. However, the estimates of effect between the two epidemiologic designs are equivalent if values of environmental covariates are defined as several-day averages or distributed lag models (Zanobetti et al. 2001). This model formulation is also resistant to mortality displacement by a few days or weeks (Zeger et al. 1999), a phenomenon in which air pollution plays a role in advancing the time of death by a relatively short period. However, the day-to-day variation in the temporal summary estimate of exposure will decrease as the number of days included in its calculation increases, thus decreasing the ability to detect effects on mortality. Furthermore, this summary measure of exposure could become confounded in time with the baseline hazard function if a large number of time lags are used, resulting in unstable parameter estimates (Dominici et al. 2003; Zeger et al. In press). Consequently, time-series studies have limitations in investigating the association between long-term exposure to environmental covariates, such as air pollution, and mortality. Studies in which individual exposures vary, either within a community or between communities, are required to estimate the effects of longer-term exposure on mortality.

In the absence of any other information in addition to the daily count data, the baseline hazard functions,  $\lambda_0^{(l)}(t)$ , and the regression parameters for the individual covariates,  $\delta_l$ , cannot be estimated. Estimation of the  $\delta_l$  values requires information on the individual covariates,  $\mathbf{x}_i^{(l)}(t)$ , which in turn is required to estimate  $\lambda_0^{(l)}(t)$ . Estimates of all these quantities are needed to estimate the amount of life lost because of exposure to the environmental covariates in this study population. The exposure effect estimate from a time-series study is therefore not sufficient to determine the amount of life lost.

However, our results show that relative risks due to exposure to the environmental covariates estimated from studies employing either a dynamic population or time-series design can then be applied to the hazard function to determine the amount of life lost under specific exposure scenarios assuming a homogeneous survival model. Age- and sex-dependent number of deaths and number of persons surviving specific ages are required to construct population-based

life tables. These quantities are used to determine the baseline hazard function for specific populations (normally for entire countries). Here, age is the time variable for the hazard function. Survival probabilities also vary with age, sex, and race, and therefore separate estimates of the effects of environmental covariates on survival should be made by these categories.

A fundamental assumption in these calculations is that the relative risks estimated in the study population can be applied uniformly to all members of the population used in deriving the life tables (viz., there is no effect modification between individual characteristics and ambient air pollution). This assumption may not be valid, as evidenced, for example, from the findings of a reanalysis of the Harvard Six Cities and the American Cancer Society studies (Krewski et al. 2000), in which an interaction was found between attained education (a measure of socioeconomic status) and level of air pollution. In addition, the effects of short-term exposure to several environmental covariates such as ambient air pollution, weather, and aerobiologies on survival may be modified by host conditions. For example, Goldberg et al. (2000, 2001) have shown that persons with certain medical conditions, such as congestive heart failure, are more susceptible to air-pollution-related death than is the general population. Their survival experience may also be different from that of the average person in that their disease condition reduces their life expectancy. Information on disease status can be incorporated into the survival model by defining an

individual-level covariate as an indicator function of the presence/absence of a disease, which would vary with time. The interaction between the disease state indicator and air pollution would provide a means of assessing the effect modification of host conditions on air-pollution-related deaths.

Incorporating the influence of disease condition on the relative risks of environmental covariates into estimates of the amount of life lost would require disease-specific life tables. Such life tables could be determined from national longitudinal population health surveys linked to mortality (Tambay and Carlin 1995). These life tables provide estimates of the expected life span of an individual with a disease condition by age. Incorporation of individual covariates (which is not possible in time-series study designs) is therefore important to capture this difference in susceptibility.

The use of time-series mortality studies to estimate the amount of life lost because of short-term ambient air pollution exposures has been criticized (Kunzili et al. 2001; McMichael et al. 1998; Rabl 2003). However, those authors suggest that it is appropriate to estimate from cohort studies the amount of life lost. We have shown that under certain conditions time-series studies can be viewed as dynamic population studies and that estimates of life lost can be obtained from time-series studies in a manner similar to that used in cohort studies. However, we did have to assume a homogeneous survival model. It is likely that people dying from short-term exposures to

### Appendix

Here we give detailed calculations that lead to the log-likelihood (Equation 4) from the initial definition (Equation 3). Writing the limits of integration in Equation 3 as a sum of integrals between consecutive days, we have

$$\sum_{t=0}^T \left[ \sum_{i \in R^{(l)}(t) \cup D^{(l)}(t)} \int_{t_0 \vee t_0}^{t_1 \wedge t_1} \lambda_i^{(l)}(u) du \right] = \sum_{t=0}^T \left[ \sum_{i \in R^{(l)}(t)} \int_{t-1}^t \lambda_i^{(l)}(u) du \right], \tag{A1}$$

with  $R^{(l)}(t)$  representing the set of individuals in stratum  $l$  at risk in the community on day  $t$ . Further note that

$$\begin{aligned} \sum_{i \in R^{(l)}(t)} \ln \lambda_i^{(l)}(t) &= \sum_{i \in R^{(l)}(t)} \ln \left[ \lambda_0^{(l)}(t) e^{\beta' A(t) + \delta_l' \mathbf{x}_i^{(l)}(t) + \phi' P_i^{(l)}(t)} \right] \\ &= y^{(l)}(t) \left[ \ln \lambda_0^{(l)}(t) + \beta' A(t) \right] + \left[ \sum_{i \in R^{(l)}(t)} \delta_l' \mathbf{x}_i^{(l)}(t) + \phi' P_i^{(l)}(t) \right], \end{aligned} \tag{A2}$$

where  $y^{(l)}(t)$  is the number of subjects who died on day  $t$  in the  $l$ th stratum. Assuming that the covariate values are constant within a day, we have

$$\begin{aligned} \int_{t-1}^t \lambda_i^{(l)}(u) du &= \int_{t-1}^t \lambda_0^{(l)}(u) e^{\beta' A(u) + \delta_l' \mathbf{x}_i^{(l)}(u) + \phi' P_i^{(l)}(u)} du \\ &= \left[ \int_{t-1}^t \lambda_0^{(l)}(u) du \right] \times e^{\beta' A(t) + \delta_l' \mathbf{x}_i^{(l)}(t) + \phi' P_i^{(l)}(t)}. \end{aligned} \tag{A3}$$

Equation 4 follows directly by substituting Equation A2 and Equation A3 in Equation 3.



environmental covariates such as ambient air pollution are more vulnerable to dying and therefore do not have the same expected residual lifetime as an average person their age. Similar concerns arise with the cohort studies in that long-term exposure to air pollution could be affecting only those persons with pre-existing diseases or some other vulnerabilities (e.g., low education). It is therefore important to develop heterogeneous survival models for both short- and long-term exposure and to conduct epidemiologic studies to both identify vulnerable populations and subgroups sensitive to environmental exposures. Estimates of the heterogeneity of survival and effect of environmental exposures on mortality coupled with disease-specific life tables will enable use to determine reasonable estimates of the amount of life lost because of environmental exposures.

## REFERENCES

- Abbey DE, Nishino N, McDonnell WF, Burchette RJ, Knutsen SF, Beeson WL, et al. 1999. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *Am J Resp Crit Care Med* 159:373–382.
- Burnett RT, Ma R, Jerrett M, Goldberg MS, Cakmak S, Pope CA III, et al. 2001. The spatial association between community air pollution and mortality: a new method of analyzing correlated geographical cohort data. *Environ Health Perspect* 109(suppl 3):375–380.
- Burnett RT, Ross WH, Krewski D. 1995. Nonlinear random effects regression models. *Environmetrics* 6:85–99.
- Brunekreef B. 1997. Air pollution and life expectancy: is there a relation? *Occup Environ Med* 54:781–784.
- Cox DR, Oakes D. 1984. *Analysis of Survival Data*. New York:Chapman and Hall.
- Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fay ME, et al. 1993. An association between air pollution and mortality in six U.S. cities. *N Engl J Med* 329:1753–1759.
- Dominici F, Daniels M, Zeger SL, Samet JM. 2002. Air pollution and mortality: estimating regional and national dose-response relationships. *J Am Statist Assoc* 97:100–111.
- Dominici F, McDermott A, Zeger SL, Samet JM. 2003. Airborne particulate matter and mortality: time-scale effects in four US cities. *Am J Epidemiol* 157:1053–1063.
- Dominici F, Samet JM, Zeger SL. 2000. Combining evidence on air pollution and daily mortality from the twenty largest US cities: a hierarchical modeling strategy. *J R Stat Soc Ser A* 163:263–302.
- Goldberg MS, Bailar JC III, Burnett R, Brook J, Tamblin R, Bonvalot Y, et al. 2000. Identifying Subgroups of the General Population That May Be Susceptible to Short-Term Increases in Particulate Air Pollution: A Time Series Study in Montreal, Quebec. Cambridge, MA:Health Effects Institute.
- Goldberg MS, Burnett R, Bailar JC III, Tamblin R, Ernst P, Flegel K, et al. 2001. The identification of persons with cardio-respiratory conditions who are at risk of dying from the acute effects of ambient air particles. *Environ Health Perspect* 109(suppl 4):487–494.
- Hastie T, Tibshirani RJ. 1990. *Generalized Additive Models*. New York:Chapman and Hall.
- Katsouyanni K, Touloumi G, Samoli G, Gryparis A, LeTertre A, Monopoli Y, et al. 2001. Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APAAZ project. *Epidemiology* 12:521–533.
- Katsouyanni K, Touloumi G, Spix C, Balducci F, Medina S, Rossi G, et al. 1997. 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. *Br Med J* 314:1658–1663.
- Kelsall J, Zeger SL, Samet JM. 1999. Frequency domain log-linear models; air pollution and mortality. *J R Stat Soc Ser C Appl Stat* 48:331–344.
- Krewski D, Burnett RT, Goldberg M, Sieminski J. 2000. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality: A Special Report of the Institute's Particle Epidemiology Reanalysis Project. Cambridge, MA:Health Effects Institute.
- Kurzi N, Medina S, Kaiser R, Quenel P, Horak F Jr, Studnicka M. 2001. Assessment of deaths attributable to air pollution: should we use risk estimates based on time series or on cohort studies? *Am J Epidemiol* 153:1050–1055.
- McMichael A, Anderson HR, Brunekreef B, Cohen AJ. 1998. Inappropriate use of daily mortality analyses to estimate longer-term mortality effects of air pollution. *Int J Epidemiol* 27:450–453.
- Moolgavkar SH. 2002. Air pollution and daily mortality in three U.S. counties. *Environ Health Perspect* 108:777–784.
- Pope CA III, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J Am Med Assoc* 278:1132–1141.
- Pope CA III, Thun MJ, Namboordiri MM, Dockery DW, Evans JS, Speizer FE, et al. 1995. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am J Respir Crit Care Med* 151:669–674.
- Rabl A. 2003. Interpretation of air pollution mortality: number of deaths or years of life lost? *J Air Waste Manag Assoc* 53:41–50.
- Samet JM, Zeger SL, Dominici F, Currier F, Coursac I, Dockery D, et al. 2000. The National Morbidity, Mortality, and Air Pollution Study (HEI Project No. 96–7): Morbidity and Mortality from Air Pollution in the United States. Cambridge, MA:Health Effects Institute.
- Schwartz J. 1999. Harvesting and long term exposure effects in the relation between air pollution and mortality. *Am J Epidemiol* 151:440–448.
- Schwartz J, Lee JT. 1999. Reanalysis of the effects of air pollution on daily mortality in Seoul, Korea: a case-crossover design. *Environ Health Perspect* 107:633–636.
- Stieb DM, Judek S, Burnett RT. 2002. Meta-analysis of time-series studies of air pollution and mortality: effects of gases and particles and the influence of cause of death, age, and season. *J Air Waste Manag Assoc* 52:470–484.
- Tambay J-L, Catlin G. 1995. Sample design of the National Population Health Survey. *Health Rep* 7:29–38.
- Zanobetti A, Schwartz J, Samoli E, Gryparis A, Touloumi G, Atkinson R, et al. 2001. The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement. *Epidemiology* 13:87–93.
- Zeger SL, Dominici F, McDermott A, Samet JM. In press. Modeling the health effects of environmental exposures: particulate air pollution and mortality in urban America. In: *Monitoring the Health of the Populations: Statistical Methods for Public Health Surveillance* (Brookmeyer R, Stroup D, eds). Oxford, UK:Oxford University Press.
- Zeger SL, Dominici F, Samet JM. 1999. Harvesting-resistant estimates of pollution effects on mortality. *Epidemiology* 89:171–175.
- Zeger SL, Thomas D, Dominici F, Cohen A, Schwartz J, Dockery D, et al. 2000. Measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Perspect* 108:419–426.
- Zidek JV, Wong H, Le ND, Burnett R. 1996. Causality, measurement error and multicollinearity in epidemiology. *Environmetrics* 7:441–451.