Effect of Air Pollution on Lung Cancer: A Poisson Regression Model Based on Vital Statistics

Toshiro Tango

Division of Theoretical Epidemiology, Department of Epidemiology, The Institute of Public Health, Tokyo, Japan

This article describes a Poisson regression model for time trends of mortality to detect the long-term effects of common levels of air pollution on lung cancer, in which the adjustment for cigarette smoking is not always neccessary. The main hypothesis to be tested in the model is that if the long-term and common-level air pollution had an effect on lung cancer, the death rate from lung cancer could be expected to increase gradually at a higher rate in the region with relatively high levels of air pollution than in the region with low levels, and that this trend would not be expected for other control diseases in which cigarette smoking is a risk factor. Using this approach, we analyzed the trend of mortality in females aged 40 to 79, from lung cancer and two control diseases, ischemic heart disease and cerebrovascular disease, based on vital statistics in 23 wards of the Tokyo metropolitan area for 1972 to 1988. Ward-specific mean levels per day of SO_2 and NO_2 from 1974 through 1976 estimated by Makino (1978) were used as the ward-specific exposure measure of air pollution. No data on tobacco consumption in each ward is available. Our analysis supported the existence of long-term effects of air pollution on lung cancer. — Environ Health Perspect 102(Suppl 8):41–45 (1994)

Key words: air pollution, ischemic heart disease, cerebrovascular disease

Introduction

Very high levels of air pollution have been shown to have some short-term effects on human health. One typical example is the smog episode in London 1952 (*1–4*). A sophisticated study design and analysis method is not neccessary to detect these short-term health effects. By contrast, there is uncertainty about the long-term effects of common levels of air pollution, because there are several difficulties in analyzing those effects, including the lack of quantitative data on individual cumulative exposures, the misspecification of the relevant unknown latency, and the existence of many confounders that might be misclassified or measured with errors. Further, most studies exploring the association of air pollution and lung cancer are based on cross-sectional ecologic study design (*5–8*).

A natural approach to examining the long-term health effects of common levels of air pollution will be to compare the time trend of incidence or mortality of the target disease among regions having different air pollution exposure levels and to detect the subtle changes, if any, expected to be seen in the patterns. Following this idea, we have designed the retrospective cohort study based on vital statistics data in the Tokyo metropolitan area between 1972 and 1988. As air pollution measures, SO_2 and NO_2 were used. The death registration system in Japan is and has been reasonably complete and reliable.

Recently, Trichopoulos et al. (*9*) compared the time trends of standardized lung cancer mortality between Athens and the rest of Greece, taking into account tobacco consumption trends, but they failed to detect the effect of air pollution on lung cancer mortality. Their approach is similar to that of this article but their design and analysis seem to be not sophisticated enough to detect the subtle change.

Materials and Methods

Mortality and Population Data

Mortality data, restricted to females aged 40 to 79 in 23 wards of the Tokyo metropolitan area outlined in Figure 1, were read from the mortality tapes of Japanese Vital Statistics for the years 1972 through 1988. This population segment was targeted because we felt that middle-aged women tend to spend more time in their home neighborhoods every day than men, who may spend many hours at work in other wards.

Annual population data were obtained from the Annual Health Report of Tokyo (*10*). Causes of death examined are lung cancer (ICD9=162) as the target disease and ischemic heart disease (ICD9 = 410–414) and cerebrovascular disease as control diseases (ICD9=430–438).

Mortality and population data in each of 23 wards were tabulated in two-way, age-by-period contingency tables with unequal person-years at risk in each cell to perform a cohort analysis. Age was divided into 14 three-year segments: 40 to 42, 43 to 45,...,76 to 78, and 79; time segments were divided into 6 three-year periods: 1972 to 1974, 1975 to 1977,...,1987 to 1988. This structure is illustrated in Table 1. The diagonals of the table (from upper left to lower right) in Figure 2 define approximate birth cohorts with 5-year intervals. A total of nine birth cohorts are obtained and used in this study.

In the analysis, we ignored the inflow and outflow of population. We also

Figure 1. Map of the Tokyo metropolitan area, 23 wards. Two wards, Chuoh and Chiyoda (shaded area), in the center of Tokyo, were excluded from the analysis.

This paper was presented at the 4th Japan-US Biostatistics Conference on the Study of Human Cancer held 9–11 November 1992 in Tokyo, Japan.

This work was supported in part by a research grant on "Study on the Health Effects of Air Pollution, 1991" from the Environmental Pollution and Health Section, Medical Welfare Division, Bureau of Public health, Tokyo Metropolitan Government.

Address correspondence to Toshio Tango, Division of Theoretical Epidemiology, Department of Epidemiology, The Institute of Public Health, 4-6-1 Shirokanedai, Minato-ku, Tokyo, Japan. Telephone 03 3441 7111. Fax 03 3446 7164. E-mail tango@iphlth.go.jp

Table 1. Three-year by three-year contingency structure for cohort analysis.[a]

Cohort No.	Age	1972–1974 1	1975–1977 2	1978–1980 3	1981–1983 4	1984–1986 5	1987–1988 6
1	40–42	●[b]					
2	43–45	◆	●				
3	46–48		◆	●			
4	49–51			◆	●		
5	52–54				◆	●	Cohort No. 1
6	55–57					◆	●
7	58–60						◆
8	61–63						•
9	64–66	▲					•
10	67–69		▲				•
11	70–72			▲			•
12	73–75				▲		•
13	76–78					▲	•
14	79						▲
							Cohort No. 9

[a]We ignored the inflow and outflow of population. [b]Each symbol represents one cohort, carried over time.

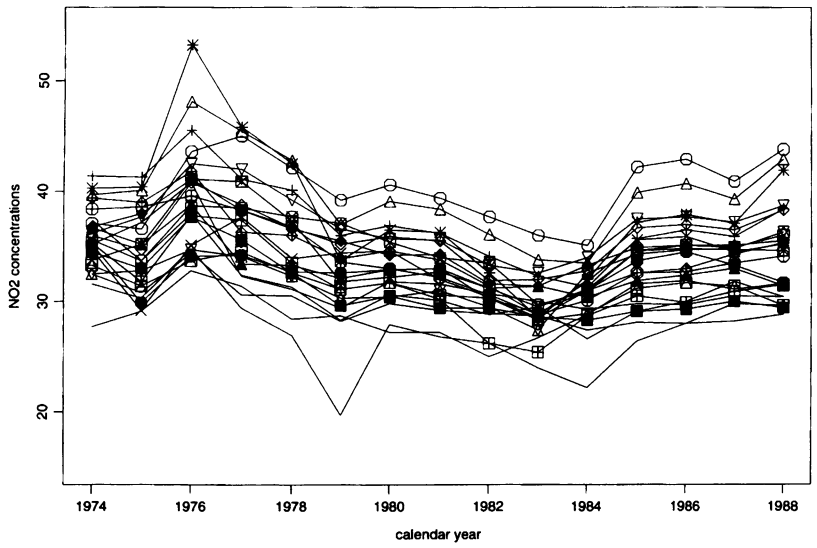

Figure 3. The ward-specific time trend of NO_2 concentration for the years 1974 through 1988. The relative rank among wards has not largely changed.

excluded two wards, Chuoh and Chiyoda, located in the central part of Tokyo, because these two wards are typical business districts. The population size in these areas is small compared with other wards, and the outflow rate from these wards to urban areas outside the 23 targeted wards has increased partly due to rapid increase of land prices in recent years.

Air Pollutants

Tokyo metropolitan ward-specific mean exposure levels of SO_2 and NO_2 per day during 1974 to 1976, estimated by Makino (*11*), were used as the exposure measure in each of 23 wards (Figure 2). There are two major reasons why these estimates were used in this study. First, these estimates were considered to be approximately proportional to the cumulative exposure levels of these two pollutants up to 1972, although it is difficult to show reliable data since ward-specific data on SO_2 and NO_2 are incomplete before 1972 in Tokyo. Second, as shown in Figure 3, the relative rank of ward-specific concentrations of these two pollutants among 23 wards has not changed largely over the past 20 years. Therefore, in this study, specification of unknown latency between exposure and mortality is considered to be not so important.

Other Data

No data on the ward-specific tobacco consumption in Tokyo are available. We have

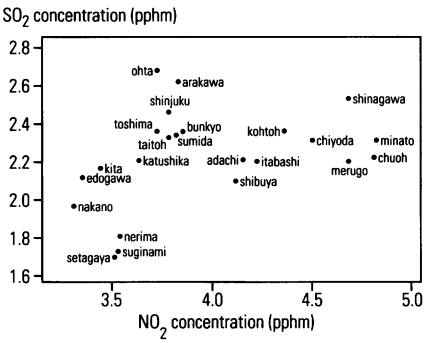

Figure 2. Relation of ward-specific mean concentration of NO_2 and SO_2 (pphm, parts per hundred million) per day during 1974 through 1976, estimated by Makino (*11*).

to use an approach that does not always need adjustment for cigarette smoking. In practice, it is unlikely that the exposure level of air pollutants such as SO_2 or NO_2 increases in proportion to the increase in tobacco consumption. Some unpublished data suggest no association between the levels of NO_2 and smoking habits in middle-aged women in the 1975 survey, and respiratory symptoms among residents in the Tokyo metropolitan area (K. Makimo, personal communication).

Hypothesis and Statistical Model

The main hypothesis in our approach is if the long-term and common-level air pollution had an effect on lung cancer, it would be expected that the death rate from lung cancer would gradually increase at a higher rate in the region with relatively high levels of SO_2 and NO_2 than in the region with low levels, and that this trend would not be expected for ischemic heart disease and cerebrovascular disease.

As a primary statistical model to test this hypothesis, we considered the following Poisson regression model for the time trend of mortality $\lambda_{ij}(t)$:

$$\mathrm{Log}\lambda_{ij}(t) = \alpha_i + \gamma_j + (\tau_i + \beta_j)t \qquad [1]$$

where

i: cohort (=1,...,9)
j: ward (=1,...,21, excluding Chuoh and Chiyoda wards)
t: period (=1,...,6)

This model makes two assumptions. First, the intercept terms $\alpha_i + \gamma_j$ are explained by many factors such as cigarette smoking, socioeconomic factors, air pollution, and many other confounding variables. Second, the variation of slope β_j can be explained mainly by air pollution. Thus,

our model further postulates the simple dose-response structure for the slope β_j:

$$\beta_j = \eta + \theta x_j \qquad [2]$$

where x_j indicates the exposure level of NO_2 or SO_2 or their product term $SO_2 \times NO_2$ in the jth ward. Therefore, our hypothesis of interest is

$$H_0 : \theta = 0$$

$$H_1 : \theta > 0$$

Goodness of fit of this model is assessed by the likelihood ratio χ^2 statistics or scaled deviance under the Poisson assumption for the number of deaths in each cell (i,j,k). Since there are many potential sources of variation in population-based data, it is very likely that the variance may be larger than the mean. To deal with this extra-Poisson variation, we use a quasilikelihood approach where over-dispersion parameter σ^2 (>1) such that

$$\mathrm{Var}(Y) = \sigma^2 E(Y)$$

is estimated by Pearson χ^2 goodness-of-fit statistics devided by degrees of freedom (*12*). Model fitting was carried out using Generalized Linear Interactive Modeling (GLIM)(*13*), and one-tailed p-value was used as the indicator of significance.

Results

Preliminary Analysis

Figure 4 shows the ward-specific time trend of age-adjusted death rate (standard population is from the 1985 national census) from lung cancer for females, aged 40 to 79, for the years 1972 to 1988. From this figure, we cannot observe any meaningful change among wards but we can see the whole trend of gradually increasing death rates. Therefore, a linear time trend analysis was performed for each of the ward-specific time trends and then the estimated slope of secular trend (rate of increase per year) was linearly regressed on NO_2, SO_2, and their product $NO_2 \times SO_2$, independently. In Figure 5, the estimated slope was plotted against NO_2. The regression line is

slope (rate of increase per year)=
$$0.313 \times NO_2 \text{ (pphm)} - 0.880$$

and one-tailed p-value is $p=0.062$, indicating that, even in this simple and rough analysis, NO_2 level is positively associated with the rate of increase in the secular

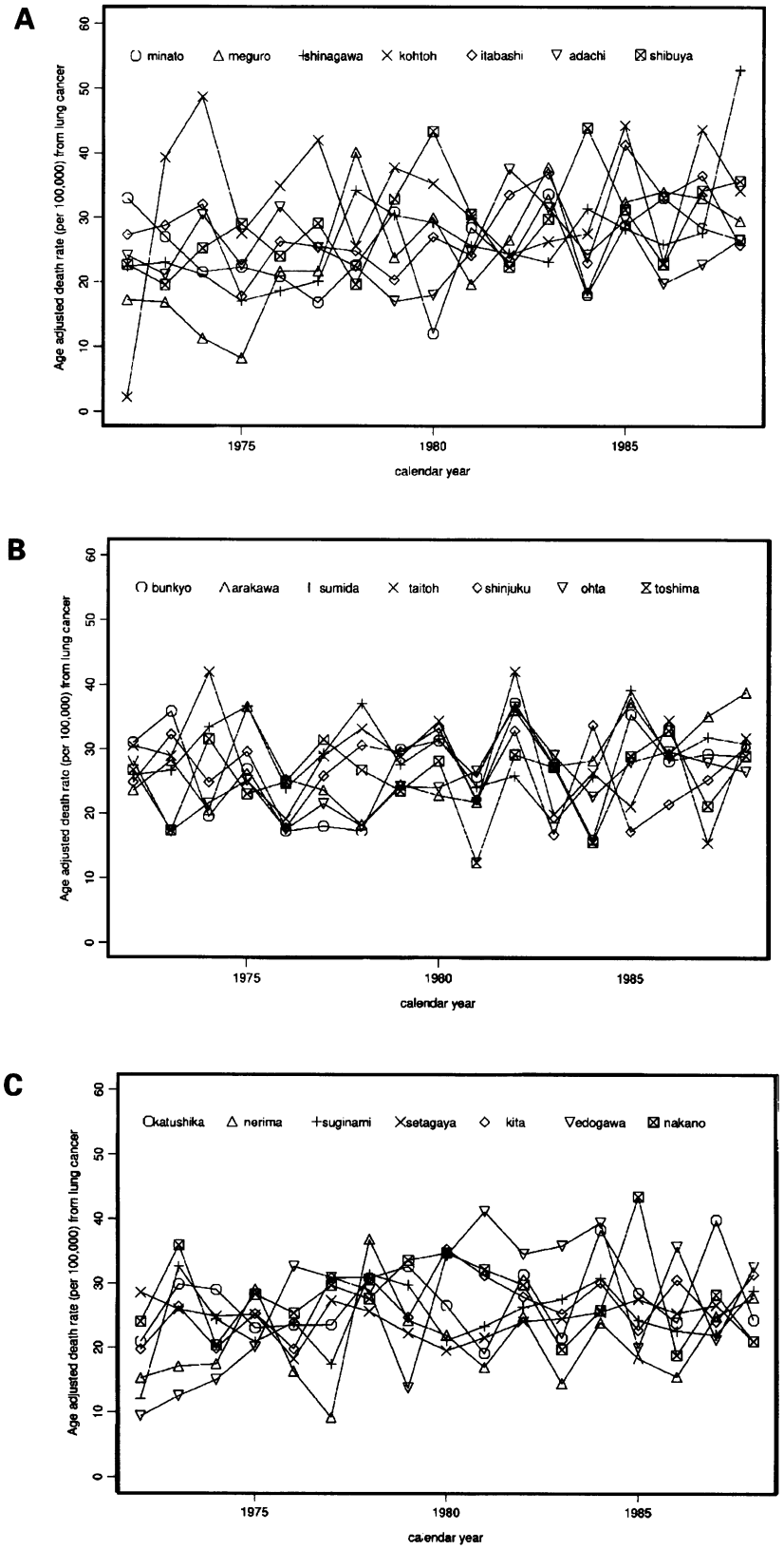

Figure 4. The ward-specific time trend of age-adjuted death rates from lung cancer, females aged 40 to 79, for 1972 through 1988. (*A*) The 7 wards with relatively high levels of NO_2. (*B*) The 7 wards with middel levels of NO_2. (*C*) The 7 wards with low levels of NO_2.

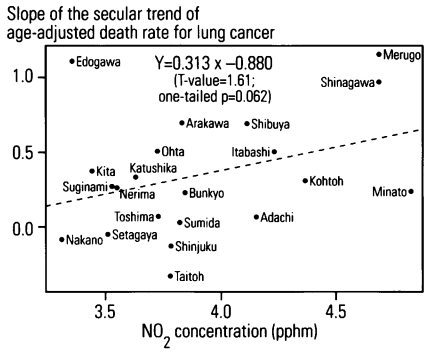

Figure 5. Association of the slope of the secular trend of age-adjusted death rate from lung cancer with NO_2 concentrations.

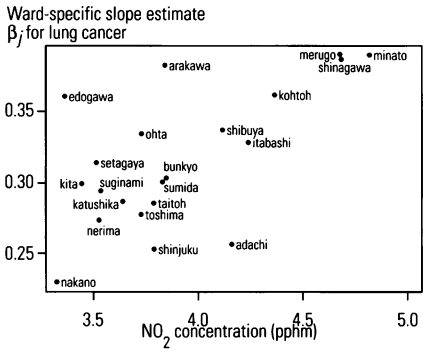

Figure 6. Association of the slope estimate β_j in the model (Equation 1) with NO_2 concentrations.

Table 2. Poisson regression analysis for the time trend of death rate, females aged 40 to 79, in Tokyo metropolitan area, 21 wards, 1972 to 1988. Estimates θ and goodness of fit of models (df = 1103).

Exposure measure	θ	SE	*t*-value	One-tailed *p*-value[a]	Scaled deviance	Pearson χ^2
Lung Cancer						
NO_2	0.059	0.023	2.54	0.0055	1226.0	1079
SO_2	0.050	0.033	1.51	0.0655	1230.2	1084
$NO_2 \times SO_2$	0.0137	0.0057	2.40	0.0081	1226.7	1081
Ischemic Heart Disease						
NO_2	−0.0120	0.0160	−0.75	0.773	1188.3	1114
SO_2	−0.0151	0.0232	−0.65	0.742	1188.4	1114
$NO_2 \times SO_2$	−0.0033	0.0040	−0.84	0.799	1188.1	1114
Cerebrovascular Disease						
NO_2	−0.0130	0.0096	−1.34	0.910	1223.6	1211
SO_2	0.0010	0.0138	0.10	0.376	1225.4	1213
$NO_2 \times SO_2$	−0.0017	0.0024	−0.71	0.762	1224.9	1212

[a] One-tailed *p*-value was calculated by the quasilikelihood approach (*12*), considering the over-dispersion.

trend of lung cancer mortality. The association between SO_2 levels and slope estimates was weaker.

Fitting Poisson Regression Model

From the results of fitting the model (Equation 1) for lung cancer, we plotted the estimated slope β_j (logarithmic scale, different from the above slope) against the NO_2 concentration in Figure 6. Association between NO_2 and slope estimates is seen more clearly than the association shown in Figure 5. A summary of fitting Equation 1 combined with Equation 2 is shown in Table 2. In this analysis, NO_2, SO_2, and the product $NO_2 \times SO_2$ were used independently as the exposure measure x_j in Equation 2. When NO_2 was used as the exposure measure, a significant association with the slope β_j of time trend of lung cancer was detected ($\theta = 0.059 \pm 0.023$ (standard error), one-tailed $p = 0.0055$). However, when SO_2 was used as a exposure measure, association with the slope is less significant ($\theta = 0.050 \pm 0.033$, one-tailed $p = 0.0655$). Owing to NO_2, the product $NO_2 \times SO_2$ was also significantly associated with the slope.

On the other hand, no significant association with SO_2 or NO_2 was found for the slope estimates of ischemic heart disease and cerebrovascular disease.

Discussion

In this analysis, we used the ward-specific NO_2 and SO_2 levels estimated by Makino (*11*). His estimation is based on two-dimensional spline interpolation using daily data from 38 measurement stations scattered in the Tokyo metropolitan area. Therefore, measurement error or estimation error is not negligible and an errors-in-variables formulation for Equations 1 and 2 is required as follows:

$$\mathrm{Log}\lambda_{ij}(t) = \alpha_i + \gamma_j + (\tau_i + \eta + \theta z_j)t$$

where z_j is the true exposure level and its relation to x_j might be

$$x_j = \lambda + \xi z_j + \varepsilon_j, \quad \varepsilon_j \sim N(0, \sigma_E^2)$$

Unfortunately, because information on the estimation error σ_E^2 is not provided, we cannot apply this model. However, the naive approach ignoring the measurement error usually underestimates θ in absolute value and has reduced power for the test of $H_0: \theta = 0$ (*14*). Namely, our results are also attenuated and the true effect of air pollution on lung cancer might be larger than that estimated.

In our analysis, we did not adjust the result for cigarette smoking simply because data on tobacco consumption are not available. If we can obtain such data, we can use them for adjustment. But, practically speaking, it will be unlikely that the ward-specific tobacco consumption is strongly correlated with the ward-specific level of NO_2 or SO_2. Although our results are largely based on this unproved important assumption, we shall conclude that our analysis suggested the existence of the long-term effects of air pollution of common levels on lung cancer.

REFERENCES

1. Ministry of Health, United Kingdom. Mortality and Morbidity during the London Fog of December 1952. Reports on Public Health and Medical Subjects, 95. London:HMSO, 1954.
2. Gore AT, Shaddick CW. Atmospheric pollution and mortality in the county of London. Br J Prev Soc Med 123:104–113 (1958).
3. Martin AE, Bradley WH. Mortality, fog and atmospheric pollution: an investigation during the winter of 1958–59. Monthly Bull Min Health Public Health Lab Serv 19:56–72 (1960).
4. Martin AE. Mortality and morbidity statistics and air pollution. Proc R Soc Med 57:969–975 (1965).
5. Wynder EL, Hammond EC. A study of air pollution carcinogenesis. I. Analysis of epidemiological evidence. Cancer 15:79–92 (1962).
6. Doll R. Atmospheric pollution and lung cancer. Environ Health Perspect 22:23–31 (1978).

7. Vena JE. Air pollution as a risk factor in lung cancer. Amer J Epidemiol 116:42–56 (1982).
8. Speizer FE. The assessment of the epidemiological data relating lung cancer to air pollution. Environ Health Perspect 47:3–42 (1983).
9. Trichopoulos D, Hatzakis A, Wynder E, Katsouyanni K, Kalandidi A. Time trends of tobacco smoking, air pollution, and lung cancer in Athens. Environ Res 44:169–178 (1987).
10. Bureau of Public Health of Tokyo. Annual Health Report of Tokyo 1972–1988. Tokyo Metropolitan Government.
11. Makino K. Computer analysis for annual mean concentration distribution in air pollutants. Ann Rep Tokyo Metro Res Lab Public Health 29:306–310 (1978).
12. Wedderburn RWM. Quasi-likelihood functions, generalized linear models. Biometrika 61:439–447 (1974).
13. Baker RJ, Nelder JA. The GLIM system, Release 3: Generalized Linear Interactive Modeling. Oxford:Numerical Algorithm Group, 1978.
14. Caroll RJ. Covariance analysis in generalized linear measurement error models. Stat Med 8:1075–1093 (1989).