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 necessary. 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 SO2 and NO2 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 necessary 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 misclassification 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, SO2 and NO2 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, 45 to 46, 46 to 47, 47 to 48, 48 to 49, 49 to 50, 50 to 51, 51 to 52, 52 to 53, 53 to 54, 54 to 55, 55 to 56, 56 to 57, 57 to 58, 58 to 59, 59 to 60, 60 to 61, 61 to 62, 62 to 63, 63 to 64, 64 to 65, 65 to 66, 66 to 67, 67 to 68, 68 to 69, 69 to 70, 70 to 71, 71 to 72, 72 to 73, 73 to 74, 74 to 75, 75 to 76, 76 to 77, 77 to 78, and 79; time segments were divided into 6 three-year periods: 1972 to 1974, 1975 to 1977, 1978 to 1980, 1981 to 1983, 1984 to 1986, and 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

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 Toshiro 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@iphth.go.jp

Figure 1. Map of the Tokyo metropolitan area, 23 wards. Two wards, Chuo and Chiyoda (shaded area), in the center of Tokyo, were excluded from the analysis.
Air excluded Tokyo metropolitan ward-specific because these areas has increased ness of Table 1. Three-year by three-year contingency structure for cohort analysis.¶

| Cohort No. | Age | 1972–1974 | 1975–1977 | 1978–1980 | 1981–1983 | 1984–1986 | 1987–1988 |
|------------|-----|-----------|-----------|-----------|-----------|-----------|-----------|
| 1          | 40–42 | ∗         |          |           |           |           |           |
| 2          | 43–45 |           | ∗         |          |           |           |           |
| 3          | 46–48 |           |           | ∗         |           |           |           |
| 4          | 49–51 |           |           |           | ∗         |           |           |
| 5          | 52–54 |           |           |           |           | ∗         |           |
| 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    |           |           |           |           |           |           |

*We ignored the inflow and outflow of population. ¶ Each symbol represents one cohort, carried over time.

Figure 3. The ward-specific time trend of NO₂ concentration for the years 1974 through 1988. The relative rank among wards has not largely changed.

排除了两个区域，Chuoh和Chiyoda，位于东京市区中部，因为这两个区域是典型的住宅区。该区域的人口规模与周边区域相比相对较小，而且这些区域的空气质量比周边区域好，有利于提高土地价格。

**Air Pollutants**

东京都的地区特定平均浓度表，二氧化硫和二氧化氮的日平均浓度，1974年到1976年，由Makino (11)提供，被用作该区域的暴露浓度表。图2中，我们用该区域的暴露浓度表和23个区域的浓度变化情况，研究了这些区域的暴露浓度表和23个区域的浓度变化情况。

图2中，我们用该区域的暴露浓度表和23个区域的浓度变化情况，研究了这些区域的暴露浓度表和23个区域的浓度变化情况。这些估计值被认为是暴露浓度表的近似比例。

这些估计值被认为是暴露浓度表的近似比例。在1972年，虽然很难显示可靠的二氧化硫和二氧化氮暴露浓度表，但是技术上，这个区域的暴露浓度表和23个区域的浓度变化情况没有显著变化。

**Other Data**

无数据的区域特定烟草消费情况在东京地区可用。我们使用一个不总是需要调整的吸烟率。在实践中，空气污染浓度和二氧化硫的增加比例在提升空气污染浓度。一些未发表的数据表明，在这些区域的二氧化硫和烟草使用习惯在中年女性中，1975年，呼吸系统症状在东京都地区。

**Hypothesis and Statistical Model**

主要假设在我们的研究中是，如果长期和一般水平的空气污染对肺癌有影响，那么肺癌发病率会逐步增加。

图3展示了 expose concentration of NO₂ for the years 1974 through 1988. The relative rank among wards has not largely changed.

Figure 2. Relation of ward-specific mean concentration of NO₂ and SO₂ (pphm, parts per hundred million) per day during 1974 through 1976, estimated by Makino (11).

 prevailing Poisson regression model for the time trend of mortality λ_j(t):

$$\log \lambda_j(t) = \alpha_j + \gamma_j(t) + \beta_j t$$  [1]

where 
- \(\lambda\): 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 \(\alpha_j\) and \(\gamma_j\) are explained by many factors such as cigarette smoking, socioeconomic factors, air pollution, and many other confounding variables. Second, the variation of slope \(\beta_j\) can be explained mainly by air pollution. Thus,
our model further postulates the simple dose-response structure for the slope $\beta_j$:

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

where $x_j$ indicates the exposure level of NO$_2$ or SO$_2$ or their product term SO$_2 \times$ NO$_2$ in the $j$th 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 $\chi^2$ 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 $\sigma^2 (>1)$ such that

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

is estimated by Pearson $\chi^2$ goodness-of-fit statistics divided 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

$$\text{slope (rate of increase per year)} = 0.313 \times \text{NO}_2 \text{(pphm)} - 0.880$$

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 trend.

Figure 4. The ward-specific time trend of age-adjusted 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 midle levels of NO$_2$. (C) The 7 wards with low levels of NO$_2$. 

Volume 102, Supplement 8, November 1994
trend of lung cancer mortality. The association between SO₂ 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 $\beta_j$ (logarithmic scale, different from the above slope) against the NO₂ concentration in Figure 6. Association between NO₂ 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, NO₂, SO₂, and the product NO₂×SO₂ were used independently as the exposure measure $x_j$ in Equation 2. When NO₂ was used as the exposure measure, a significant association with the slope $\beta_j$ of time trend of lung cancer was detected ($\theta=0.059 \pm 0.023$ (standard error), one-tailed $p=0.0055$). However, when SO₂ 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 NO₂, the product NO₂×SO₂ was also significantly associated with the slope.

On the other hand, no significant association with SO₂ or NO₂ was found for the slope estimates of ischemic heart disease and cerebrovascular disease.

Discussion

In this analysis, we used the ward-specific NO₂ and SO₂ 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:

$$\log y_j(t) = \alpha + \beta_j (t - \theta_j) + \epsilon_j$$

where $\epsilon_j$ is the true exposure level and its relation to $x_j$ might be

$$x_j = \lambda + \xi x_j + e_j, \quad e_j \sim N(0, \sigma^2)$$

Unfortunately, because information on the estimation error $\sigma^2$ is not provided, we cannot apply this model. However, the naive approach ignoring the measurement error usually underestimates $\theta$ 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 NO₂ or SO₂. 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, Shaddock 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. 1. 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).