



ORIGINAL CONTRIBUTIONS

Effects of Air Pollution on Daily Clinic Visits for Lower Respiratory Tract Illness

Jing-Shiang Hwang¹ and Chang-Chuan Chan²

The authors used data obtained from clinic records and environmental monitoring stations in Taiwan during 1998 to estimate the association between air pollution and daily numbers of clinic visits for lower respiratory tract illness. A small-area design and hierarchical modeling were used for the analysis. Rates of daily clinic visits were associated with current-day concentrations of nitrogen dioxide, carbon monoxide, sulfur dioxide, and particulate matter less than or equal to 10 μm in aerometric diameter. People over age 65 years were the most susceptible, and estimated pollution effects decreased as the exposure time lag increased. The analysis also suggested that several community-specific variables, such as a community's population density and yearly air pollution levels, modified the effects of air pollution. In this paper, the authors demonstrate the use of a small-area design to assess acute health effects of air pollution. *Am J Epidemiol* 2002;155:1–10.

air pollution; ambulatory care facilities; Bayes theorem; respiratory tract infections; small-area analysis

Editor's note: An invited commentary on this article appears on page 11, and the authors' response appears on page 16.

Many epidemiologic studies have applied time-domain methods to demonstrate significant associations between air pollution and various health effects, including mortality, hospital admissions, emergency room visits, and general-practice consultations for respiratory symptoms, in single geographic areas (1–3). All of these types of studies share two common features in their data collection methods. First, the studies are mainly carried out in places with large populations so that sufficient numbers of daily events can be

collected for time-series analysis. Second, population exposures are usually represented by aggregating air pollution data obtained from several monitoring stations in geographic areas as large as 100 km^2 . Misclassification is often compounded in these studies, because spatial variation of individual exposure is typically not considered.

One solution to this problem is to create less heterogeneous exposures by clustering hospitals around a monitoring station as suggested by Burnett et al. (4) and Zidek et al. (5). Unfortunately, exposure attribution based on clustered hospitals remains a serious challenge, because some hospitals are located up to 200 km away from any monitoring stations.

Although known census clusters can provide us with exposure populations in smaller and more homogeneous regions that have well-defined boundaries, we may encounter another statistical modeling problem in the use of census clusters, because data on many important explanatory factors are either unmeasured or not available in all clusters. This leads to the necessity of more complicated methods of data analysis in order to compensate for the uncertainty associated with modeled values for explanatory variables (6). The other shortcoming of using census clusters is that census areas are not equivalent to clinic catchment areas. Furthermore, the census subdivision as a cluster becomes less than ideal when daily outcomes are sparse, which is often the case for serious illness in a small area (7).

Received for publication June 22, 2000, and accepted for publication May 31, 2001.

Abbreviations: PM_{10} , particulate matter $\leq 10 \mu\text{m}$ in aerometric diameter; SD, standard deviation.

¹Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.

²Institute of Occupational Medicine and Industrial Hygiene, College of Public Health, National Taiwan University, Taipei 100, Taiwan.

Reprint requests to Dr. Chang-Chuan Chan, Institute of Occupational Medicine and Industrial Hygiene, Room 1447, National Taiwan University, Jen-ai Road, Taipei 100, Taiwan (e-mail: ccchan@ha.mc.ntu.edu.tw).

In this study, we used a new design for clustering data and a Bayesian hierarchical modeling strategy to estimate effects of air pollution on daily clinic visits for lower respiratory tract illness across small study areas in Taiwan. Data on clinic visits were obtained from a database of the National Health Insurance Program of Taiwan. We used a classical method to estimate the population at risk so that the counts of daily clinic visits could be converted to daily rates of lower respiratory tract illness for each area. The temporal and spatial patterns of these multiple time series of rates were modeled in two phases (8). In the first phase, we used general regression models to model temporal patterns in order to obtain the relative change in clinic visit rates according to changes in levels of nitrogen dioxide, carbon monoxide, sulfur dioxide, ozone, and particulate matter $\leq 10 \mu\text{m}$ in aerometric diameter (PM_{10}) for 50 individual small communities. In the second phase, we combined the pollution-health relative rates across the 50 communities by using a Bayesian hierarchical model to improve the estimates of air pollution effects for each community and to obtain an overall estimate of air pollution effects in Taiwan.

MATERIALS AND METHODS

Study communities and environmental data

Our study communities included 50 townships and city districts in Taiwan where ambient air monitoring stations of the Taiwan Air Quality Monitoring Network are situated (figure 1). The air pollutants measured at the monitoring stations include hourly levels of sulfur dioxide, nitrogen dioxide, carbon monoxide, PM_{10} , and ozone, which are presumed to represent the spatial distribution of air pollution levels in a study community with an area less than 15 km^2 (9). Meteorologic measurements, such as wind direction, wind speed, temperature, dew point, and precipitation, are also taken at these stations. As population exposures in our statistical models, we used the 24-hour average for nitrogen dioxide, sulfur dioxide, and PM_{10} , the hourly maximum for ozone, and the maximum 8-hour running average for carbon monoxide. Daily maximum temperature and average dew point temperature were also used to adjust for meteorologic effects.

Population at risk

For our clinic-based data set, we defined the population at risk for a selected study community as all persons who would go to a clinic in the study community if they needed to make a medical visit. This is actually the service coverage of all of the clinics in the community. Therefore, an unbiased estimate of the population at risk would include some nonresident daytime workers who might visit a clinic in the community but would exclude residents who preferred to use medical resources outside the community.

Clinic visits

Since its inauguration in 1995, the Taiwanese National Health Insurance Program has covered medical services for approximately 96 percent of a population of 23 million

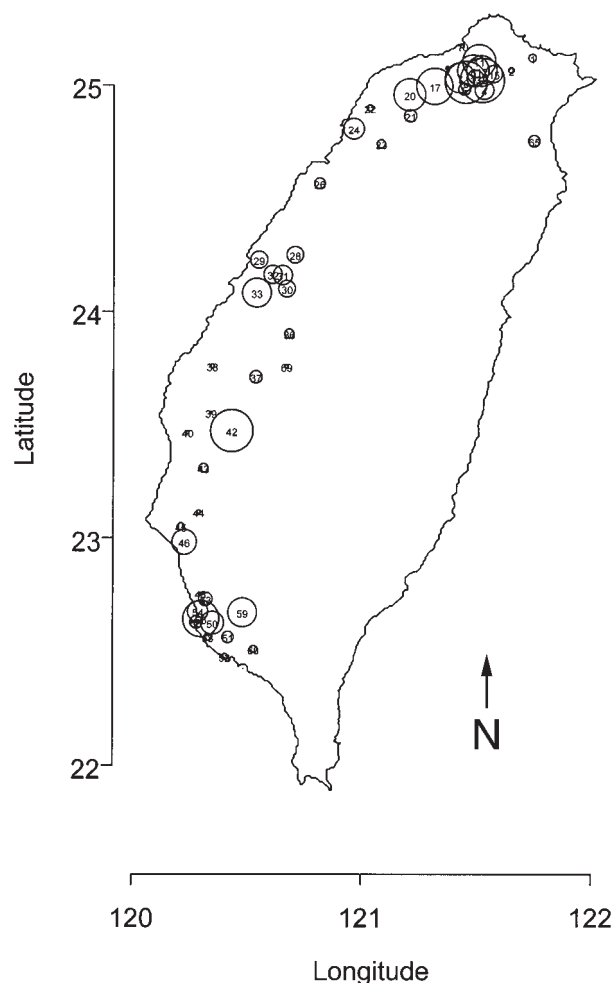


FIGURE 1. Map of 50 townships and city districts used in a study of air pollution effects on clinic visits for lower respiratory tract illness, Taiwan, 1998. The map shows the monitoring station numbers of the Taiwan Air Quality Monitoring Network. The dimensions of the circles are proportional to the population at risk in each community.

through a network of 16,122 contracted medical institutions. Computerized records of daily clinic visits are available for each contracted medical institution. The records contain data on clinics, townships, dates of visits, patient identification, gender, and birthday and the cause of each visit according to the *International Classification of Diseases*, Ninth Revision. In this study, we used 1-year records from 1998 from the 50 study communities for further data analysis. We included clinic visits due to lower respiratory tract illness (*International Classification of Diseases*, Ninth Revision, codes 466 and 480–486), such as acute bronchitis, acute bronchiolitis, and pneumonia, as outcomes in the models to estimate pollution effects. In contrast, we used the records of clinic visits due to all diseases to estimate the population at risk for the study communities in 1998. Clinic records available for 45 communities in 1997 were used to empirically verify our approach to population estimation. We further

classified the estimated population into three age groups—children (ages 0–14 years), adults (ages 15–64 years), and the elderly (ages ≥ 65 years)—in order to estimate age-specific pollution effects.

Population estimation

The service coverage of all of the clinics in one community, i.e., the population at risk, can be estimated by adding the number of people who visited a clinic in 1998 to the expected number of people who were not in the clinic records for 1998 but might visit a clinic in the community in the future. Here the problem is similar to that of estimating the number of unseen species in ecologic studies, using only the number of individuals captured during a fixed time interval. Under the assumption of stability across time and independent captures, the estimator developed for unseen species can be directly applied to estimate the population at risk (10, 11). An individual's number of clinic visits in a community during 1 year, x , is analogous to a species' having x members captured during one unit of time. For the species problem, the x members are assumed to be unrelated, while one person's x clinic visits are generally correlated. However, this equivalence assumption may still be approximately satisfied if we count only the first visit when there are consecutive visits with the same diagnosis category in a short time period.

Let n_x be the number of people having exactly x clinic visits in a community during 1998. Then

$$\sum_{x=1}^{x_0} n_x$$

is the total number of different people who have made at least one clinic visit in that community in 1998. We need to estimate the number of people who made no clinic visits in 1998 but would do so if they were later sick, n_0 , in order to complete the estimation of population at risk in 1998. We assume that all n_0 people will eventually get sick and visit one of the clinics in the coming t years. Then the expected number of these people who were not recorded in the 1998 database but might visit a clinic in the subsequent t years, denoted $\Delta(t)$, is very close to n_0 . A combination of available data from observed n_x with $x \geq 1$ is used to estimate $\Delta(t)$. In this study, we adopted the approach proposed by Efron and Thisted (11) to estimate $\Delta(t)$, which is

$$\hat{\Delta}(t) = \sum_{x=1}^{x_0} h_x n_x$$

with

$$h_x = (-1)^{x+1} t^x \Pr(B \geq x),$$

where B is a binomially distributed random variable with number x_0 and probability $1/(1+t)$. Efron and Thisted also showed in their paper (11) that the choice of $x_0 = 9$ could achieve reasonable estimation with a variance of

$$\text{Var}(\hat{\Delta}(t)) = \sum_{x=1}^{x_0} h_x^2 n_x.$$

Ideally, one should choose an appropriate t value to obtain a less biased population estimate without excess uncertainty. Our choice of $t = 5$ to estimate the population at risk was based on the observation that patients' medical-care-seeking behavior was stable under the National Health Insurance Program and that limited changes had taken place in the demographic profile of the study communities in the previous 6 years.

The validity of this population estimator was evaluated empirically using the database for 1997. We estimated the number of people not recorded in the 1997 database who appeared in the database in 1998. We found that our population estimator was quite accurate, because the mean absolute value of the relative difference between estimated additional subjects, $\hat{\Delta}(t = 1)$, and actually observed new patients in 1998 was less than 2 percent across study communities.

Phase I modeling

The reason we converted daily visit counts to rates in this study is that it is relatively easier to implement a linear regression model with time-series error on continuous rates as compared with a Poisson regression model for time series of counts (12). In the first phase, time series of daily clinic visit rates for each subpopulation were modeled separately. The rates of clinic visits, on a logarithmic scale, were modeled as a linear function of time-varying potentially confounding variables and pollution levels on the current day and on the immediately preceding days. Our models were general linear regressions with seasonal autoregressive moving-average residual processes (13). The regression terms were chosen through extensive exploratory data analyses. Dummy variables were used to account for day-of-the-week (SUN, MON, and SAT) and special-holiday (SH) fluctuations in the series. Average temperature and dew point temperature for the previous 3 days were considered as confounding variables in the models. We added another two temperature variables in the models to allow changes in the slope for temperature effects when the temperature was above 32°C. Two more dummy variables were used to indicate the extremely high rates of clinic visits during the winter season between January and February (WIN) and the extremely low rates of clinic visits during the summer season between July and August (SUM). With all of these variables being adjusted, we added individual air pollutants separately to the models in order to estimate individual air pollutants' effects on daily clinic visits.

In particular, for the a th age group in the i th community, we fitted the following linear regression model in order to obtain the estimated pollution coefficient, denoted $\hat{\beta}_{iah}$, and its standard error $\hat{\sigma}_{iah}$:

$$\begin{aligned} \log(y_{iat}) = & \gamma_{iah0} + \gamma_{iah1}\text{SUN} + \gamma_{iah2}\text{MON} + \gamma_{iah3}\text{SAT} \\ & + \gamma_{iah4}\text{SH} + \gamma_{iah5}\text{WIN} + \gamma_{iah6}\text{SUM} + \gamma_{iah7}\text{TG32}_{it} \\ & + \gamma_{iah8}\text{TL32}_{it} + \gamma_{iah9}\text{TP3}_{it} + \gamma_{iah10}\text{DEW}_{it} + \gamma_{iah11}\text{DP3}_{it} \end{aligned}$$

$$+ \beta_{iah} \text{POLL}_{i,t-h} + W_{iaht},$$

where y_{iat} is the observed clinic visit rate of the a th age group in the i th community on the t th day; TG32_{it} is the value of the maximum temperature on day t if the maximum temperature is greater than or equal to 32°C , and 0 otherwise; TL32_{it} is the value of the maximum temperature on day t if the maximum temperature is less than 32°C , and 0 otherwise; TP3_{it} and DP3_{it} are the average temperature and dew point temperature, respectively, for the 3 days prior to day t ; and DEW_{it} is the dew point temperature on day t . $\text{POLL}_{i,t-h}$ is the level of pollutant on day $t-h$, where t is the current day and h ranges from 0 to 2. A first-order autoregressive process models the error term W_{iaht} . Since the error series contained a periodic component, we further modeled the difference $W_{iaht} - W_{iaht-7}$ with another first-order autoregressive process.

The proposed model was first examined in several communities representing typical urban and rural regions in Taiwan. We also obtained a reasonably good fit with a mean $R^2 = 0.53$ in fitting the data of all subpopulations to the regression model with these same covariates. Ideally, we could explore the data to find the best models for each setting of the combination of five air pollutants, three time lags, and four age categories (including one with all ages combined) in all 50 locations, respectively. However, because of efficiency considerations and limited statistical gains in reducing estimation errors, we applied this single regression model to all subpopulations at all 50 locations in this phase.

We fitted 600 regression models to the time series of clinic visits in phase I: one for each choice of three time lags and four age categories in 50 study communities for each air pollutant separately. The health impact of each air pollutant is reported as the percentage increase in clinic visit rates that corresponded to a 10 percent increase in local air pollution levels. That is, the percentage change is expressed by

$$100\{\exp(0.1\bar{X}_{ih}\hat{\beta}_{iah}) - 1\},$$

where $\hat{\beta}_{iah}$ is the estimated pollution coefficient from the phase I model for community i , age group a , and lag h and $\bar{X}_{ih}$ is the corresponding average pollution level. The 95 percent confidence interval for the percentage change was constructed by replacing $\hat{\beta}_{iah}$ with

$$\hat{\beta}_{iah} \pm 2\hat{\sigma}_{iah},$$

where $\hat{\sigma}_{iah}$ is the standard error.

Phase II modeling

The second phase of the modeling strategy was to borrow information from neighboring communities to improve estimates of the pollution coefficients in individual locations and to obtain an overall pollution coefficient across multiple locations. The hierarchical modeling approach has been widely applied to the analysis of multilevel data (14–17). In this study, we used the same hierarchical modeling approach as that proposed by Dominici et al. (8) to estimate the rela-

tion between air pollution and daily clinic visits across 50 small study areas.

The modeling approach can be described in the following three stages. First, the 50 pollution coefficients for a fixed age group and time lag are estimated by the phase I regression model, denoted $\hat{\beta} = (\hat{\beta}_1 \dots \hat{\beta}_{50})'$. These are assumed to be normally distributed multivariate random variables with true pollution coefficients $\beta = (\beta_1 \dots \beta_{50})'$ as the mean vector and a covariance matrix $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1^2 \dots \hat{\sigma}_{50}^2)$, where $\hat{\sigma}_i$ is the estimate of the standard error of $\hat{\beta}_i$. Secondly, spatial variation among the 50 true pollution coefficients must be addressed. Treating the β_i 's as random, we model the degree of similarity of the true pollution coefficients between community i and community j , $h_{ij} = \text{Corr}(\beta_i, \beta_j)$, as a function of a Euclidean distance between the air monitoring stations for communities, denoted d_{ij} . We define $h_{ij} = \exp\{-d_{ij}/R\}$, where R is a positive parameter. Note that h_{ij} goes to 0 as the distance d_{ij} grows large. The spatial distribution of the vector of true pollution coefficients β is assumed to be normal with mean $Z\alpha$ and covariance matrix Ω , where $\alpha = (\alpha_1 \dots \alpha_p)'$ is a vector of random coefficients corresponding to covariate matrix $Z_{50 \times p}$ and nonsingular $\Omega = \tau^2 H$, where $H = (h_{ij})_{50 \times 50}$ and τ^2 is a common variance that applies to each β_i . For the current study, we construct

$$Z = [I, sPD, sT, sNO_2, sSO_2, sPM_{10}, sO_3, sCO],$$

where I is a vector of 1 for the overall mean and the seven covariates include the standardized score vectors of population densities (sPD), annual average temperatures (sT), and annual average concentrations of five air pollutants (nitrogen dioxide (sNO_2), sulfur dioxide (sSO_2), PM_{10} (sPM_{10}), ozone (sO_3), and carbon monoxide (sCO)) in these 50 communities. The standardized score for each covariate was computed by first subtracting the mean value across the 50 locations and then dividing by the standard deviation based on the 50 mean values. Since all predictors are centered about their respective means, the intercept α_1 can be interpreted as an overall pollution coefficient for any location with mean predictors. The other coefficients, α_2 , K , and α_p , reflect the effects of each location's population density, average temperature, and pollution levels on its local pollution coefficient. Based on empirical correlograms for the 50 estimated pollution coefficients, the range parameter R in h_{ij} of H is fixed at 5 km (18).

Thirdly, we must complete the hierarchical structure with a proper prior model for α and τ^2 in order to perform a Bayesian analysis. For mathematical convenience, we use conjugate priors, normal prior $\alpha \sim N(\eta, C)$ and inverse gamma prior $\tau^2 \sim \text{IG}(a, b)$, in our model (8, 19). The hyperparameters in our model, η , C , a , and b , are chosen to reflect no information on α and τ^2 . The Bayesian inference is based on the posterior distribution of β , α , and τ^2 given the phase I estimates $\hat{\beta}$, $\hat{\Sigma}$, and the specified hyperparameters. Samples from these posteriors can be obtained from the Markov chain Monte Carlo algorithm (8).

TABLE 1. Population densities, air pollution levels, weather data, and age-stratified rates of clinic visits for lower respiratory tract illness in 50 study communities in Taiwan, 1998

	Population density (persons/ km ²)	Pollutant level					Weather		Daily clinic visit rates (visits/1,000)			
		Nitrogen dioxide (ppb*)	Sulfur dioxide (ppb)	PM ₁₀ * (μg/m ³)	Ozone (ppb)	Carbon monoxide (ppm*)	Temperature (°C)	Dew point (°C)	Children (ages 0–14 years)	Adults (ages 15–64 years)	Elderly (ages ≥65 years)	All ages
Mean	7,123	23.6	5.4	58.9	54.2	1.00	28.1	19.6	2.39	0.88	1.02	1.34
Standard deviation	8,024	5.4	3.0	14.0	10.2	0.30	1.4	1.1	0.77	0.38	0.57	0.50
Minimum	252	13.0	1.5	33.3	38.9	0.51	24.6	17.3	1.24	0.30	0.26	0.58
Maximum	28,203	34.1	16.9	83.1	78.3	1.71	30.6	22.4	5.59	2.15	3.11	3.31

* ppb, parts per billion; PM₁₀, particulate matter with a diameter ≤ 10 μm; ppm, parts per million.

RESULTS

As table 1 shows, we found significant heterogeneity for environmental data and outcomes among the 50 study communities. Population density ranged from 250 persons per km² to 28,000 persons per km² for the study communities, and the population at risk ranged from 19,000 to 278,000. The averages of daily average nitrogen dioxide, sulfur dioxide, PM₁₀, and carbon monoxide levels and daily maximum ozone levels measured from the 50 monitoring stations in 1998 were 23.6 ppb (standard deviation (SD) 5.4), 5.4 ppb (SD 3.0), 58.9 μg/m³ (SD 14.0), 1.0 ppm (SD 0.3), and 54.2 ppb (SD 10.2), respectively. Yearly average temperature

ranged from 24.6°C to 30.6°C. The average of daily rates of clinic visits due to lower respiratory tract illness among the 50 communities was 1.34 per 1,000 (SD 0.5). The daily rates of clinic visits also varied among the three age groups, with the highest rates being seen for children (2.39/1,000) and the lowest rates for adults (0.88/1,000).

The results of phase I showed that the variation in clinic visits was statistically significantly related to variation in nitrogen dioxide, carbon monoxide, sulfur dioxide, and PM₁₀ exposures. By contrast, there was no significant pollution effect for ozone exposures. For illustration and simplicity, in this paper we present only the results of nitrogen dioxide with 0- to 1-day lags for all ages combined. As figure 2 shows, the

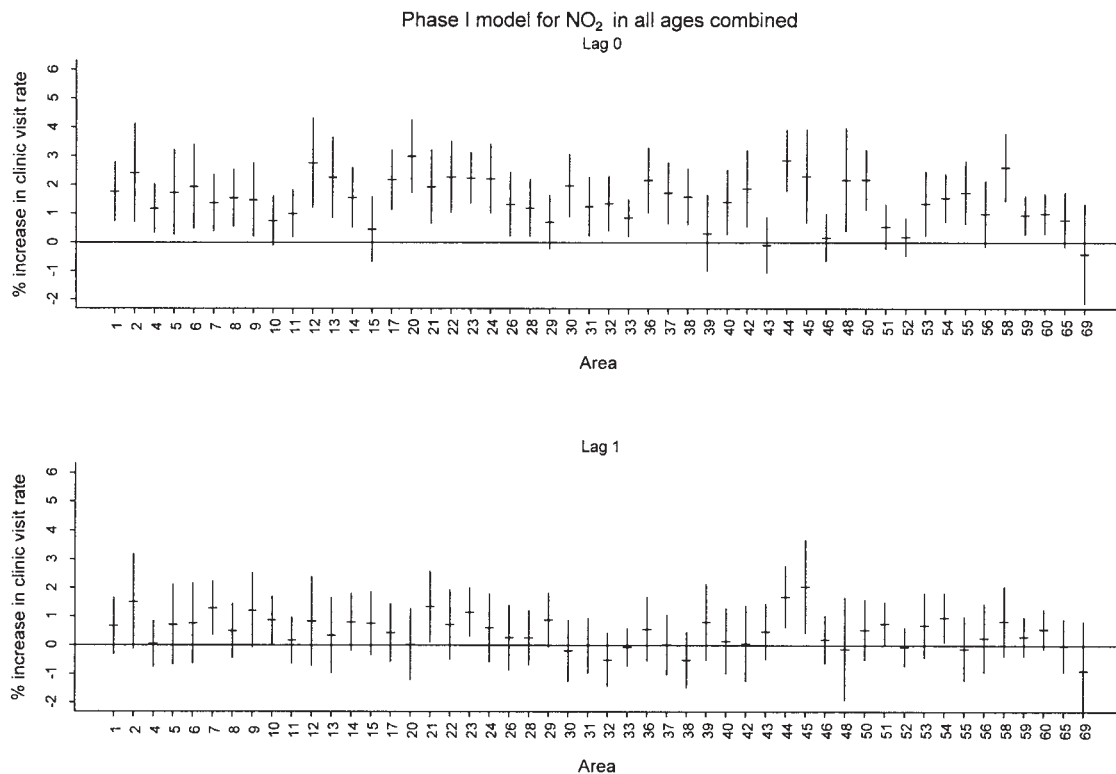


FIGURE 2. Effects of a 10% increase in nitrogen dioxide (NO₂) exposure with a 0- to 1-day lag on percentage change in rates of clinic visits for lower respiratory tract illness in 50 Taiwanese communities, as estimated by linear regression models (phase I; see text), 1998. For area, see figure 1. Mean values (—) and 95% confidence intervals (vertical bars) are indicated for each location.

percentage changes in clinic visits were significantly associated with nitrogen dioxide exposures at the current day (lag 0) but less significantly associated at the 1-day lag among these 50 communities. The plots shown in figure 2 also indicate significant intracommunity and intercommunity variability in the estimated percentage changes in clinic visit rates across the 50 communities.

The results of five single-pollutant models in phase II analysis for an all-ages population with current-day exposure are shown in table 2. The 95 percent posterior support intervals of the estimated overall pollution coefficient (α_1) showed that overall clinic visits were significantly related to nitrogen dioxide, carbon monoxide, sulfur dioxide, and PM₁₀ exposures but not to ozone. The coefficient of α_1 can be used to calculate the percentage increase in clinic visits. For example, there was a $100\{\exp(2.4 \times 0.00593) - 1\}$, i.e., 1.4 percent, increase for an α_1 of 5.93×10^{-3} and a 2.4 ppb increase in nitrogen dioxide levels. Furthermore, an individual community's pollution coefficient for nitrogen dioxide was adjusted by -1.34×10^{-3} (α_6) times the PM₁₀ difference (in standard deviation units) between its mean and the overall mean. The effects of sulfur dioxide and carbon monoxide exposures were also negatively adjusted by the community's annual PM₁₀ concentrations. The effects of sulfur dioxide exposures were positively affected by the community's annual carbon monoxide concentrations. The effects of carbon monoxide exposures were negatively affected by the community's annual averages of maximum 1-hour ozone concentrations. The community's population density slightly affected the effects of carbon monoxide and sulfur dioxide exposures. However,

yearly averages of the community's temperature, nitrogen dioxide levels, and sulfur dioxide levels had no significant influence on the acute effects of the five pollutants in the phase II models.

Comparing the same settings in figures 3 and 2, we found that phase II hierarchical models showed smaller intracommunity and intercommunity variations than the regression models of phase I after spatial correlation and community covariates were considered in the models. As a result, the estimates of percentage changes in phase I models were also pulled toward the overall mean in phase II models. As figure 4 illustrates, smaller percentage changes in clinic visits were also significantly associated with current-day sulfur dioxide, PM₁₀, and carbon monoxide exposures for most study communities.

The overall effects of pollution on clinic visits due to lower respiratory tract illness were calculated by using national average concentrations as baseline exposures and percentages above the national average level as contrast exposures in the hierarchical models. Table 3 shows the estimated percentage change in the clinic visit rate for every 10 percent increase in pollution levels, with 0- to 2-day lags for four age categories and five pollutants. Nitrogen dioxide had the greatest estimated percentage increases in daily clinic visit rates for a 10 percent increase in pollution levels. The pollution effects were always greatest for current-day exposures and decreased significantly as exposure time lags increased for nitrogen dioxide, carbon monoxide, sulfur dioxide, and PM₁₀. The magnitudes of pollution effects appeared to increase with age, with the elderly being the most susceptible.

TABLE 2. Posterior mean values for overall true pollution coefficients and coefficients of community covariates obtained using Bayesian hierarchical models for an all-ages population with current-day exposures to air pollution, Taiwan, 1998

Variable	Pollutant* considered in the model				
	Nitrogen dioxide	Sulfur dioxide	PM ₁₀ †	Ozone	Carbon monoxide
Overall pollution coefficient‡	5.93 (4.98, 6.87)§	10.01 (7.41, 12.58)	0.84 (0.35, 1.31)	0.08 (-0.42, 0.57)	78.83 (59.55, 98.26)
Community covariate¶					
Population density	-0.93 (-2.9, 0.96)	-4.37 (-9.19, 0.62)	-0.24 (-1.09, 0.62)	-0.12 (-1.02, 0.78)	-26.09 (-55.73, 3.9)
Temperature	0.96 (-1.32, 3.26)	-0.15 (-5.41, 5.07)	0.07 (-0.74, 0.89)	0.06 (-0.82, 0.93)	28.92 (-13.5, 71.08)
Nitrogen dioxide	0.11 (-3.03, 3.23)	-1.40 (-8.03, 5.14)	0.03 (-1.32, 1.42)	-0.24 (-1.74, 1.24)	-5.32 (-60.24, 49.51)
Sulfur dioxide	0.35 (-0.97, 1.65)	-1.84 (-4.47, 0.76)	0.14 (-0.47, 0.77)	0.05 (-0.55, 0.64)	0.43 (-21.01, 21.36)
PM ₁₀	-1.34 (-2.66, -0.01)	-3.14 (-5.93, -0.3)	-0.28 (-0.83, 0.28)	-0.16 (-0.75, 0.41)	-23.54 (-47.63, 0.83)
Ozone	-0.80 (-1.76, 0.15)	1.42 (-0.39, 3.21)	0.00 (-0.43, 0.43)	0.09 (-0.37, 0.56)	-21.36 (-39.41, -3.04)
Carbon monoxide	-0.15 (-1.12, 0.82)	6.36 (4.24, 8.5)	0.23 (-0.21, 0.68)	0.13 (-0.33, 0.61)	10.83 (-4.82, 26.6)

* Estimated coefficient multiplied by 10^3 .

† PM₁₀, particulate matter with a diameter $\leq 10 \mu\text{m}$.

‡ Estimated overall pollution coefficient across the 50 communities.

§ Numbers in parentheses, 95% posterior support interval.

¶ The standardized score for each covariate was obtained by first subtracting the mean across the 50 locations and then dividing by the standard deviation.

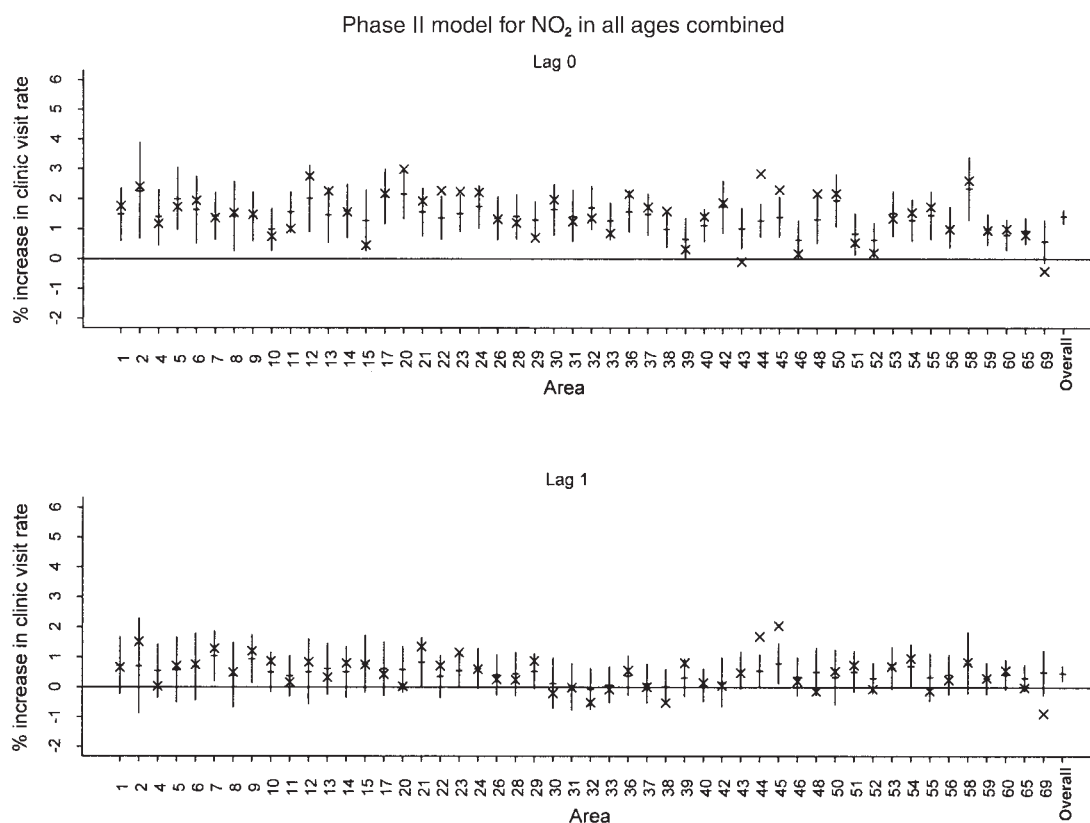


FIGURE 3. Effects of a 10% increase in nitrogen dioxide (NO₂) exposure with a 0- to 1-day lag on percentage change in rates of clinic visits for lower respiratory tract illness in 50 Taiwanese communities, as estimated by Bayesian hierarchical models including community-specific covariates (phase II; see text), 1998. For area, see figure 1. Posterior mean values (—) and 95% posterior support intervals (vertical bars) from phase II are indicated for each location. The × symbol indicates estimated mean percentage change from phase I.

DISCUSSION

Few epidemiologic studies have related numbers of clinic visits to ambient air pollution levels. The information obtained from this study can improve our understanding of health effects associated with air pollution. Our study found that nitrogen dioxide, carbon monoxide, sulfur dioxide, and PM₁₀ all had significant estimated effects on daily clinic visits due to lower respiratory tract illness in Taiwan. In contrast, daily clinic visit rates were not associated with maximum hourly ozone levels. These findings are similar to those of Hajat et al. (3), who reported that daily general-practice consultations for respiratory conditions were related to similar air pollutants and unrelated to ozone in London, United Kingdom. The estimated effects of nitrogen dioxide identified in this study are also consistent with our previous findings. In a previous study (20), we reported that nitrogen dioxide exposure was related to increased schoolchildren's absences due to respiratory illness in the subsequent 3 days. Our current analysis also suggests that people aged 65 years or more are more susceptible than other age groups to the effects of nitrogen dioxide, carbon monoxide, sulfur dioxide, and PM₁₀. In addition, the estimated effects of these air pollutants

decreased as exposure time lags increased. Accordingly, we recommend that more studies be conducted in order to explore the biologic plausibility of respiratory illnesses' being caused or exacerbated by specific air pollutants or by air pollution in general.

Although high collinearity among air pollutants prevented us from using multipollutant models directly, we did obtain better estimates of acute exposure effects by adjusting for chronic exposures. We did this by adding yearly averages of the five major pollutants to our Bayesian hierarchical models. For example, chronic PM₁₀ exposures slightly modified the estimated acute exposure effects of nitrogen dioxide and sulfur dioxide, even after each community's population density and temperature were adjusted for in the models. These findings indicate that in the future we must consider multiple pollutants and interactions between acute and chronic exposures simultaneously in order to estimate individual air pollutants' concentration-response relations.

A major strength of this study was the innovation of estimating the population at risk from the medical claims data of small private clinics and medium-sized medical institutions in Taiwan's National Health Insurance Program. The estimates of populations at risk allowed us to calculate rates

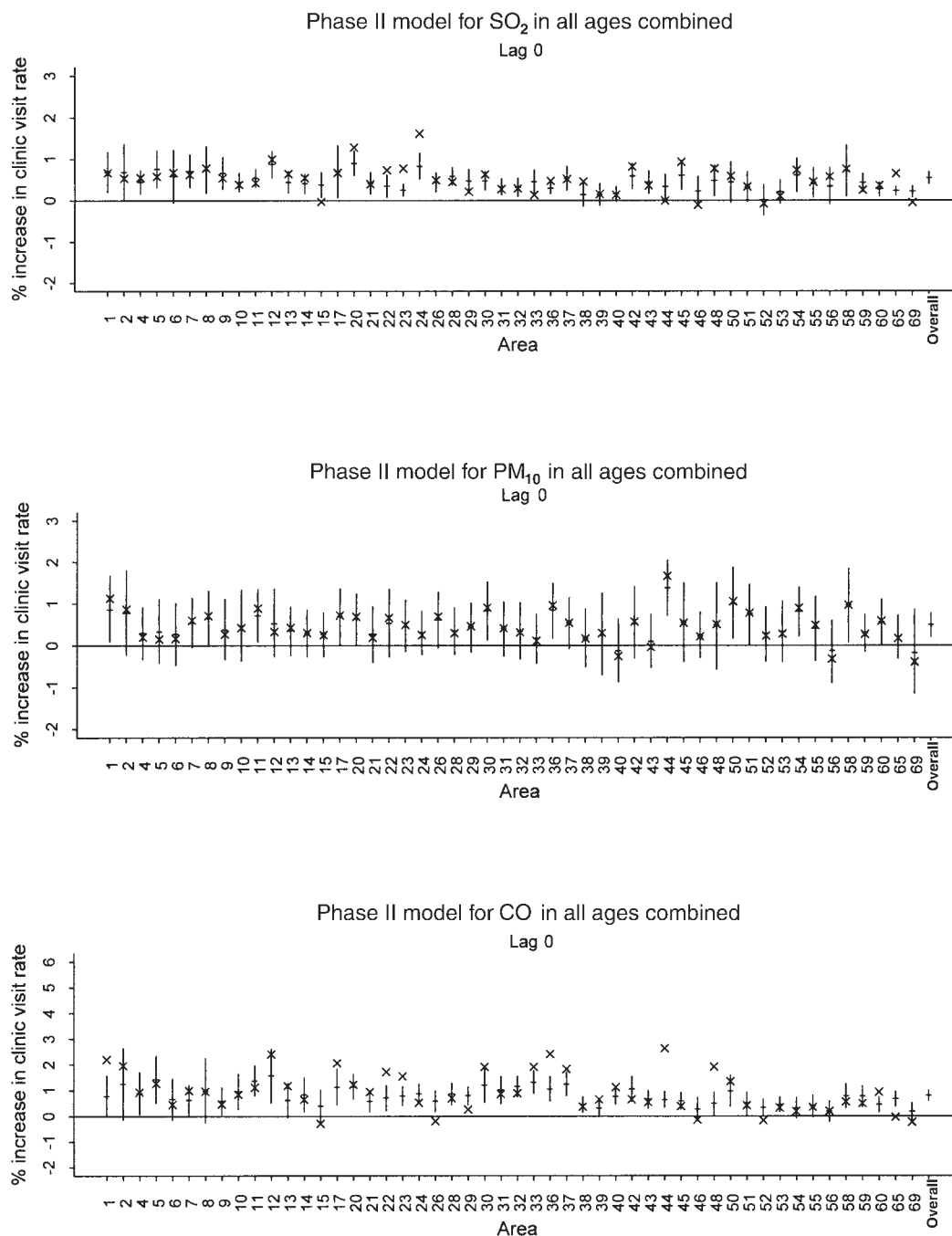


FIGURE 4. Effects of 10% increases in sulfur dioxide (SO₂), particulate matter $\leq 10 \mu\text{m}$ in aerometric diameter (PM₁₀), and carbon monoxide (CO) exposures at the current day on percentage change in rates of clinic visits for lower respiratory tract illness in 50 Taiwanese communities, as estimated by Bayesian hierarchical models including community-specific covariates (phase II; see text), 1998. For area, see figure 1. Posterior mean values (—) and 95% posterior support intervals (vertical bars) from phase II are indicated for each location. The × symbol indicates estimated mean percentage change from phase I.

of daily clinic visits across 50 small study areas. We believe that the accuracy of this method of estimating population at risk can be explained by the provision of comprehensive medical coverage to a majority of people in the National

Health Insurance Program; this allows residents to have easy access to widely available medical clinics with minimal surcharges for clinic visits in their communities. We also found no significant difference in the estimation of pol-

TABLE 3. Estimated percentage changes in rates of clinic visits due to lower respiratory illness for a 10% increase in air pollution levels with 0- to 2-day lags, by age group, Taiwan, 1998

Age group (years)	Lag (days)	Pollutant considered in the model									
		Nitrogen dioxide (2.4 ppb*)†		Sulfur dioxide (0.5 ppb)		PM ₁₀ * (5.9 µg/m³)		Ozone (5.4 ppb)		Carbon monoxide (0.1 ppm*)	
		%‡	95% CI*	%	95% CI	%	95% CI	%	95% CI	%	95% CI
Children (ages 0–14 years)	0	1.3	1, 1.6	0.5	0.3, 0.6	0.5	0.2, 0.8	0.0	–0.2, 0.3	0.7	0.5, 1.0
	1	0.6	0.3, 0.9	0.2	0.0, 0.3	0.0	–0.3, 0.3	0.0	–0.3, 0.3	0.1	–0.2, 0.3
	2	0.3	0.1, 0.6	0.2	0.1, 0.4	0.3	0.0, 0.6	0.1	–0.2, 0.4	0.2	–0.1, 0.4
Adults (ages 15–64 years)	0	1.5	1.3, 1.8	0.7	0.5, 0.8	0.6	0.2, 0.9	0.1	–0.2, 0.3	0.9	0.6, 1.1
	1	0.6	0.3, 0.9	0.2	0.0, 0.3	0.2	–0.1, 0.5	0.1	–0.2, 0.4	0.2	0.0, 0.5
	2	0.2	–0.1, 0.5	0.2	0.0, 0.3	0.3	0.0, 0.6	0.1	–0.2, 0.4	0.2	–0.1, 0.4
Elderly (ages ≥65 years)	0	1.8	1.4, 2.2	0.8	0.6, 1.1	0.8	0.4, 1.1	0.2	–0.2, 0.5	1.1	0.8, 1.5
	1	0.9	0.5, 1.3	0.3	0.1, 0.6	0.3	–0.1, 0.6	0.1	–0.2, 0.5	0.6	0.3, 1.0
	2	0.4	0.1, 0.8	0.3	0.1, 0.5	0.5	0.1, 0.8	0.1	–0.3, 0.4	0.4	0.1, 0.8
All ages	0	1.4	1.2, 1.6	0.5	0.4, 0.7	0.5	0.2, 0.8	0.0	–0.2, 0.3	0.8	0.6, 1.0
	1	0.5	0.2, 0.7	0.1	0.0, 0.2	0.0	–0.3, 0.3	0.0	–0.2, 0.3	0.1	–0.1, 0.3
	2	0.2	–0.1, 0.4	0.1	0.0, 0.3	0.3	0.0, 0.6	0.0	–0.2, 0.3	0.1	–0.1, 0.3

* ppb, parts per billion; PM₁₀, particulate matter with a diameter ≤ 10 µm; ppm, parts per million; CI, confidence interval.

† Numbers in parentheses, 10% national average pollution concentration.

‡ Estimated posterior mean percentage change in the clinic visit rate.

lution effects when we estimated the population at risk using t values from 1 to 10, although the estimated populations did change.

We could have estimated air pollution effects by modeling the original daily counts of clinic visits as a Poisson process instead of using Gaussian linear processes. Because linear predictors of these two models are the same except for one constant term of population-at-risk on the log scale, we can expect almost identical estimated pollution coefficients by either approach. A minor difference between these two models is the assumed variance structure. The choice of the time-series approach in this study provided us with flexible model selection and diagnostic checking and simplified computation in comparison with the Poisson approach. Modeling rates would also allow one to interpret the effects of community-level variables on the intercepts as chronic effects on clinic visits. It is worthwhile to explore various risk factors associated with such chronic effects, including long-term air pollution exposures.

Normally a joint temporospatial model can fit the multiple time series of clinic visit rates simultaneously with relatively complicated calculations. In this paper, we fitted a linear regression model first for the time domain and then Bayesian hierarchical models for the space domain. Instead of using a best-fitted linear regression model for each of 3,000 settings in the first phase, we used only one single model for all settings in order to simplify our calculation. We suspect that the use of Bayesian hierarchical models in phase II with confounding adjustment and cross-community information input could actually offset loss of statistical power due to suboptimal linear regression models in phase I.

Furthermore, we found that the estimates of pollution effects were relatively robust, both to the choice of range parameter, R , for spatial correlation and to the choice of pri-

ors of α and τ^2 in Bayesian hierarchical models. By using R and alternative noninformative priors, we found similar effect estimates in a sensitivity analysis of our models.

Bayesian hierarchical models with community-specific covariates can smooth out heterogeneous estimates but still retain meaningful information on variations in pollution effects among communities. Accordingly, through a small-area design and Bayesian hierarchical models, one can efficiently obtain national risk estimates for air pollution effects and improve community-specific estimates for individual communities. We believe that such a design and modeling approach could be used more widely in environmental epidemiologic studies.

ACKNOWLEDGMENTS

This research was conducted through the sponsorship of a contract (NSC 89-EPA-Z-001-001) awarded by the National Science Council, Taiwan.

The authors acknowledge the support of the Taiwan Environmental Protection Administration and the Bureau of National Health Insurance in providing air pollution data and clinic visit records.

REFERENCES

1. Pope CA III, Dockery DW. Epidemiology of particle effects. In: Holgate ST, Samet JM, Maynard RL, et al, eds. Air pollution and health. London, United Kingdom: Academic Press,

- 1999;671–705.
2. Thurston GD, Ito K. Epidemiological studies of ozone exposure effects. In: Holgate ST, Samet JM, Maynard RL, et al, eds. *Air pollution and health*. London, United Kingdom: Academic Press, 1999;485–510.
3. Hajat H, Haines A, Goubet SA, et al. Association of air pollution with daily GP consultations for asthma and other lower respiratory conditions in London. *Thorax* 1999;54:597–605.
4. Burnett RT, Dales RE, Raizenne ME, et al. Effects of low ambient levels of ozone and sulfates on the frequency of respiratory admissions to Ontario hospitals. *Environ Res* 1994;65:172–94.
5. Zidek JV, White R, Le ND, et al. Imputing unmeasured explanatory variables in environmental epidemiology with application to health impact analysis of air pollution. *Environ Ecol Stat* 1998;5:99–115.
6. Cressie N, Kaiser MS, Daniels MJ, et al. Spatial analysis of particulate matter in an urban environment. In: Gómez-Hernández J, Soares A, Froidevaux R, eds. *geoENVII—geostatistics for environmental applications*. (Proceedings of the Second European Conference on Geostatistics for Environmental Applications, Valencia, Spain, November 18–20, 1998). Dordrecht, the Netherlands: Kluwer Academic Publishers, 1999;41–52.
7. Saez M, Tobias A, Munoz P, et al. A GEE moving average analysis of the relationship between air pollution and mortality for asthma in Barcelona, Spain. *Stat Med* 1999;18:2077–86.
8. Dominici F, Samet JM, Zeger SL. Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy (with discussion). *J R Stat Soc Ser B* 2000;163:263–302.
9. Chan CC, Hwang JS. Site representativeness of urban air monitoring stations. *J Air Waste Manag Assoc* 1996;46:755–60.
10. Chao A, Yip P, Lin HS. Estimating the number of species via a martingale estimating function. *Stat Sin* 1996;6:403–18.
11. Efron B, Thisted R. Estimating the number of unseen species: how many words did Shakespeare know? *Biometrika* 1976;63:433–47.
12. Zeger SL. Regression model for time series of counts. *Biometrika* 1988;75:621–30.
13. Box GE, Jenkins GM, Reinsel GC. *Time series analysis: forecasting and control*. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.
14. Gelfand AE, Hills SE, Racine-Poon A, et al. Illustration of Bayesian inference in normal data models using Gibbs sampling. *J Am Stat Assoc* 1990;85:972–85.
15. Lindley DV, Smith AF. Bayes estimates for the linear model (with discussion). *J R Stat Soc Ser B* 1972;34:1–41.
16. Gelfand AE, Smith AF. Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 1990;85:398–409.
17. Smith AF. A general Bayesian linear model. *J R Stat Soc Ser B* 1973;35:67–75.
18. Ripley BD. *Statistical inference for spatial processes*. Cambridge, United Kingdom: Cambridge University Press, 1988.
19. Hobert JP, Casella G. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J Am Stat Assoc* 1996;91:1461–73.
20. Hwang JS, Chen YJ, Wang JD, et al. Subject-domain approach to the study of air pollution effects on schoolchildren's absenteeism. *Am J Epidemiol* 2000;152:67–74.