

## Invited Commentary: Air Pollution and Health—What Can We Learn from a Hierarchical Approach?

Francesca Dominici

The potential for air pollution to cause excess deaths at high concentrations was established in the mid-20th century by a series of air pollution “disasters” in the United States and Europe (1–3) that caused striking increases in mortality. By the early 1990’s, time-series studies, each conducted at a single location (4–7), showed that air pollution levels, even at much lower concentrations, were associated with increased rates of mortality and morbidity in cities in the United States, Europe, and other developed regions. At present, although these relative rates are small (an increase in mortality or morbidity of a few percentage points over a realistic exposure range), the burden of disease attributable to air pollution may be substantial, considering the very large populations exposed to air pollution and the large numbers of persons to whom the relative rates of mortality or morbidity apply.

In the past, critics of single-site studies questioned the validity of the data used and the statistical techniques applied to them. The critics noted inconsistencies in findings among studies and even in the same city upon independent reanalysis (5, 6). They questioned the choice of particular cities and asked whether models had been selected that gave estimates of effect that were biased upwards. These criticisms have since been addressed by the use of multisite studies (8, 9) in which site-specific data on air pollution and health are assembled under a common framework. Hierarchical models, which combine information across locations, have provided a statistical approach for analyzing multisite studies (10).

The work by Hwang and Chan (11), published in this issue of the *Journal*, is one of the latest contributions on this topic. Their study illustrates the utility of using hierarchical models to analyze data on the relation between air pollution concentrations and clinic visits for treatment of lower respiratory tract illness. Hwang and Chan analyzed such data (as well as data on temperature and dew point levels) for 50 sites in Taiwan in 1998. Here I discuss the advantages of

using hierarchical models to analyze multisite time-series data on air pollution and health, provide perspective on the results of Hwang and Chan (11), and address the problem of publication bias in meta-analyses.

### HIERARCHICAL MODELS

Hierarchical models provide an appropriate approach for summarizing and integrating the findings of research studies in a particular area (12–15). Hierarchical models have been familiar to statisticians for four decades. Recently, because of the development of computational tools that facilitate their implementation (16, 17), hierarchical models have been widely applied in many disciplines and have been used to address environmental research questions.

The use of Bayesian hierarchical models to analyze multisite time-series data in relation to air pollution and health provides an appropriate approach for combining evidence across studies, quantifying the sources of variability, and identifying effect modification. For example, Hwang and Chan (11) assume a two-stage hierarchical model with the following structure.

#### Stage I: site

Given a time series of daily mortality counts at a given site, the association between air pollution and health within that site is described using a regression model, which takes into account potentially confounding factors such as trend, season, and climate. Among the output of the stage I analysis are the point estimate ( $\hat{\beta}^s$ ) and the statistical variance ( $v^s$ ) of the estimated mortality rate associated with each air pollutant at each site.

#### Stage II: the Taiwan region

Data from the 50 sites of the Taiwan Ambient Air Quality Network are combined by using a linear regression model, where the outcome variable is the true relative mortality rate associated with air pollution indexes within each site and the explanatory variables ( $X_j^s$ ) are the site-specific characteristics (population density, yearly averages of the pollutants and of temperature) that may modify the relative rate. Formally,

$$\beta^s = \alpha_0 + \sum_{j=1}^p \alpha_j X_j^s + \text{error}$$

Received for publication June 27, 2001, and accepted for publication August 15, 2001.

Abbreviations: NMMAPS, National Morbidity, Mortality, and Air Pollution Study;  $PM_{10}$ , particulate matter  $\leq 10 \mu\text{m}$  in diameter.

From the Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Room E4138, Baltimore, MD 21205 (e-mail: fdominic@jhsp.h.edu). (Reprint requests to Dr. Francesca Dominici at this address).

If the predictors  $X_j^s$  are centered about their mean values, the intercept ( $\alpha_0$ ) can be interpreted as the pooled coefficient for a hypothetical site with mean predictors. The regression parameters ( $\alpha_j$ ) measure the change in the true relative rate of mortality associated with a 1-unit change in the corresponding site-specific variable.

The sources of variation in the estimation of the health effects of air pollution are specified by the levels of the hierarchical model. Under a two-stage hierarchical model, the difference between the estimated site-specific relative rate ( $\hat{\beta}^s$ ) and the true pooled relative rate ( $\alpha_0$ ) can be broken down as

$$(\hat{\beta}^s - \alpha_0) = (\hat{\beta}^s - \beta^s) + (\beta^s - \alpha_0)$$

The variation of  $\hat{\beta}^s$  about  $\beta^s$  is described by the within-site variance ( $v^s$ ), which depends on the number of days with available air pollution data and on the predictive power of the site-specific regression model. The variation of  $\beta^s$  about  $\alpha_0$  is described by the between-site variance ( $\tau^2$ ), which measures the heterogeneity of the air pollution effects across cities. The specification of a Bayesian hierarchical model is completed with the selection of the prior distributions for the parameters at the highest level of the hierarchy. If there is no desire to incorporate prior information into the analysis, then vague prior distributions are the default choice. However, it is important to complete the Bayesian analysis by investigating the sensitivity of the substantial findings to the choice of the prior distributions.

Posterior distributions of the pooled estimate ( $\alpha_0$ ), of the between-site variance ( $\tau^2$ ), and of the second-stage regression parameters ( $\alpha_j$ ) provide an appropriate summary of the site-specific relative rates of mortality; a full characterization of the heterogeneity of the air pollution effects across the 50 locations; and the identification of site-specific characteristics that modify the association between air pollution and health. The two-stage hierarchical approach used by Hwang and Chan (11) can be extended to include additional levels of the hierarchical models (for example, sites within geographic regions, geographic regions within nations, etc.) which lead to the estimation of additional sources of variability (within sites, between sites within regions, and between regions) and potential effect modifiers at the site or regional level (for example, see Dominici et al. (18)).

Complex hierarchical models can be fitted using simulation-based methods (17, 19) which provide samples from the posterior distributions of all parameters of interest. Several software packages for this process are now available (for example, see <http://www.mrc-bsu.cam.ac.uk/bugs/>). One of the appealing features of simulation-based approaches is that site ranking with respect to the magnitude of the health effects of air pollution is straightforward. For example, in the paper by Hwang and Chan (11), the posterior probability that a particular site is the worst location in terms of the health effects of air pollution can be estimated easily by determining the empirical frequency with which the relative rate of hospital admissions at that particular site is the largest.

## HETEROGENEITY AND EFFECT MODIFICATION

Hwang and Chan (11) used simulation-based methods to approximate the posterior distributions of all parameters of interest. They summarized their results by calculating Bayesian estimates (and 95 percent posterior intervals) of the site-specific and overall air pollution effects. Alternatively, a point estimate of the pooled effect can be obtained by assuming a random-effects model and by taking a weighted average of the site-specific estimates—as suggested by DerSimonian and Laird (20), for example. Under the weighted average approach for a random-effects model, the weights of the site-specific estimates are modified to take into account the variability between locations—for example, by including a point estimate of  $\tau^2$ . Unfortunately, in the Hwang and Chan paper (11), very little attention was given to the issue of heterogeneity, which can be appropriately assessed under a Bayesian approach. In fact, the inspection of posterior distributions of  $\tau^2$  provides a better characterization of the degree of heterogeneity of effects across sites than a point estimate of  $\tau^2$  and/or the classical  $\chi^2$  statistic for testing  $\tau^2 = 0$ .

Hwang and Chan (11) reported that rates of daily clinic visits were positively associated with current-day concentrations of nitrogen dioxide, carbon dioxide, sulfur dioxide, and particulate matter  $\leq 10 \mu\text{m}$  in diameter ( $\text{PM}_{10}$ ). Overall, they found that a 10-unit (ppb) increase in current-day nitrogen dioxide concentrations was associated with approximately a 5.8 percent increase (95 percent posterior interval: 4.9, 6.8) in daily clinic visits for respiratory illness (11). This is a much larger estimate than the 0.2 percent increase in mortality (95 percent posterior interval: -0.25, 0.7) reported in the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) (21). Hwang and Chan also found that a 10-unit ( $\mu\text{g}/\text{m}^3$ ) increase in current-day  $\text{PM}_{10}$  concentrations was associated with approximately a 0.84 percent increase (95 percent posterior interval: 0.35, 1.31) in daily clinic visits for respiratory illness (11). This pooled  $\text{PM}_{10}$  result is slightly lower than the morbidity results of the multisite NMMAPS, which reported a 1.4 percent increase (95 percent confidence interval: 0.34, 2.45) in hospital admissions for chronic obstructive pulmonary disease associated with current-day  $\text{PM}_{10}$  (21).

With respect to effect modification, Hwang and Chan (11) found that an individual pollution coefficient for nitrogen dioxide was modified by a yearly average  $\text{PM}_{10}$  in a direction implying lower shorter-term (acute) effects of nitrogen dioxide at greater longer-term (chronic) levels of  $\text{PM}_{10}$ . Similarly the effects of sulfur dioxide and carbon monoxide exposure were also associated negatively with a community's annual  $\text{PM}_{10}$  concentrations.

In recent papers (18, 22, 23), researchers have explored and discussed effect modification of the health effects of air pollution but have focused mainly on the identification of modifiers of the short-term effects of  $\text{PM}_{10}$  instead of modifiers of the short-term effects of nitrogen dioxide, as Hwang and Chan did (11). In the NMMAPS, Dominici et al. (18) found that short-term  $\text{PM}_{10}$  effects are modified by long-term average  $\text{PM}_{10}$  level, indicating greater short-term (acute)  $\text{PM}_{10}$  effects at lower long-term (chronic) levels of  $\text{PM}_{10}$ . Negative associations between short-term effects of

air pollution (nitrogen dioxide and  $PM_{10}$ ) and long-term average  $PM_{10}$  level might indicate that the pool of susceptible individuals is smaller in cities with higher average  $PM_{10}$  concentrations. If this were the case, the short-term effects of air pollution on mortality would be lower in cities with a large particulate matter average than in cities with a lower particulate matter average and a relatively larger pool of susceptible individuals. On the other hand, Katsouyanni et al. (22), within the APHEA study [Air Pollution and Health, European Approach], found that long-term average nitrogen dioxide concentration was an effect modifier of the short-term effect of  $PM_{10}$ , but in the opposite direction: The higher the average nitrogen dioxide level, the larger the particle effect. These findings suggest that the acute effects of  $PM_{10}$  on mortality might be greater in locations with a greater long-term average of air pollution originating from vehicle exhaust (nitrogen dioxide is considered an indicator of traffic-derived pollution) as compared with pollution from other sources.

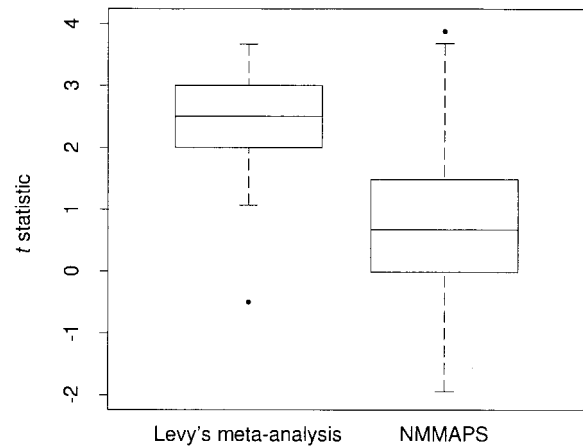
In these multisite studies, there has not been a consistent pattern of effect modification, probably because of difficult methodological problems. These difficulties include the following: 1) the sites have sociodemographic characteristics which vary spatially within that site, leading to serious concerns about ecologic bias (19); 2) the predictors included in the second stage of the model represent a crude proxy for the desired site-specific characteristics, and they are very highly correlated; and 3) the number of sites is limited, which restricts the number of potential effect modifiers that can be investigated jointly.

## PUBLICATION BIAS

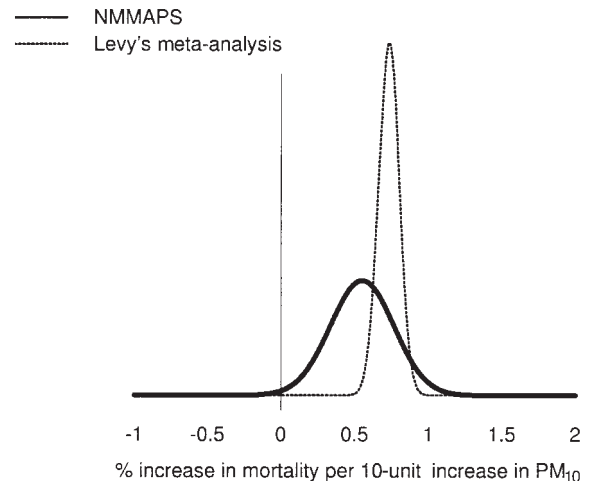
One advantage of using the multisite design to collect time-series data on air pollution and health is that it is less prone to the phenomenon of publication bias (24, 25), which might strongly affect results obtained from meta-analyses of published studies. To illustrate this point, I compared findings between the NMMAPS (21) and a recent meta-analysis of published studies of  $PM_{10}$  and mortality by Levy et al. (23). Figures 1 and 2 show the results of this comparison.

Figure 1 shows boxplots of site-specific estimates of  $PM_{10}$  effects divided by their standard errors ( $t$  statistics =  $\hat{\beta}^s/\sqrt{v^s}$ ) included in the meta-analysis of 19 US locations (23) and in the NMMAPS multisite study of 90 US locations (21). The  $t$  statistics in the meta-analysis are larger than the  $t$  statistics obtained in the NMMAPS—indicating the possible “file drawer” phenomenon, common in meta-analyses of published studies, in which studies finding no significant relation between  $PM_{10}$  and mortality may not have been published.

Figure 2 shows the posterior distribution of the pooled effect of  $PM_{10}$  on mortality ( $\alpha_0$ ) obtained under the meta-analysis (dotted line) and under the NMMAPS (solid line). Substantial overlap between the two curves indicates that the results obtained under the NMMAPS (21) and the results obtained under the Levy et al. meta-analysis (23) are relatively consistent. However, the posterior distribution of the pooled effect has a larger mean, and it is much more con-



**FIGURE 1.** Boxplots of site-specific estimates of the effects of particulate matter  $\leq 10 \mu\text{m}$  in diameter divided by their standard errors ( $t$  statistics =  $\hat{\beta}^s/\sqrt{v^s}$ ) from a meta-analysis of 19 US locations by Levy et al. (23) and the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), a study of air pollution in 90 US locations (21). The horizontal lines within the boxes represent the median value, and the edges of the boxes represent the interquartile range. Dashed lines denote 95% confidence intervals, and bullets represent outliers.



**FIGURE 2.** Comparison between the pooled effects of particulate matter  $\leq 10 \mu\text{m}$  in diameter ( $PM_{10}$ ) ( $\mu\text{g}/\text{m}^3$ ) on mortality obtained in a meta-analysis of 19 US locations by Levy et al. (23) and in the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), a study of air pollution in 90 US locations (21). The solid line represents the posterior distribution of the pooled effect of  $PM_{10}$  on mortality obtained in the NMMAPS. The dotted line represents the normal approximation to the estimated pooled effect of  $PM_{10}$  on mortality obtained in Levy et al.'s meta-analysis (23).

centrated under the Levy et al. meta-analysis (23) than under the NMMAPS (21). For example, the posterior probability that the pooled  $PM_{10}$  effect estimated with the meta-analysis is larger than the pooled  $PM_{10}$  estimated under the

NMMAPS is close to 80 percent. This indicates that meta-analyses of published studies that selectively favor findings with significant effects might seriously overestimate the pooled effect and underestimate the corresponding statistical uncertainty. However, this difference might also be explained by methodological differences between the NMMAPS and the time-series studies used in the meta-analysis.

## CONCLUSIONS

The paper by Hwang and Chan (11) in this issue of the *Journal* is an important contribution to the estimation of the health effects of air pollution from multisite time-series data using a Bayesian hierarchical model. For guidance in policy development, exposure-risk relations must be described with sufficient precision (26); this precision can be gained by pooling the large amounts of publicly available data on mortality, morbidity, air pollution, and potential confounders and modifiers.

Multisite studies and Bayesian hierarchical models provide a unified framework for 1) estimating individual pollutant effects for particular sites, pooled effects, and components of variation; 2) investigating effect modification; and, more generally, 3) producing more credible results by appropriately taking into account all sources of uncertainty and by overcoming the problem of publication bias.

Heterogeneity is potentially a key issue for the policy implications of epidemiologic studies of air pollution and health, because, in the presence of substantial heterogeneity across sites, the overall effect may have less public health relevance than site-specific estimates, and the characterization of effect modification becomes of primary scientific and public health interest (27). Clearly, interpretation of findings on effect modification is a difficult component of these analyses. The methods are weakened by gaps in the publicly available data on air pollution, mortality, and site-specific characteristics and by the inherent limitations of these data.

A promising approach to the investigation of effect modification might rely on the integration of results from small-area studies and global studies (that is, studies in which information from several multisite studies is combined). These two approaches are complementary with respect to their advantages and limitations. Small-area studies collect data in more homogeneous and restricted geographic areas. Therefore, they are less prone to ecologic bias, and the potential modifiers are more precisely measured. However, results of small-area studies are less generalizable to other locations. Global studies aim to combine information on air pollution and health within and between multisite studies from different countries and regions (the United States, Europe, Canada, and Asia), and therefore they are likely to provide a more complete characterization of the heterogeneity. However, they are still sensitive to the problem of ecologic bias, and additional work in these areas is necessary.

With repeated application of hierarchical models to multisite data, current statistical methods offer an approach for tracking the health effects of air pollution over time as con-

trol measures are implemented. These techniques might contribute to the development of a global surveillance system for measuring the health effects of air pollution and for documenting improvements attributable to changes in the regulation of air purity.

## REFERENCES

1. Brimblecombe P. The big smoke. London, United Kingdom: Methuen Publishing Ltd, 1987.
2. Ciocco A, Thompson DJ. A follow-up of Donora ten years after: methodology and findings. *Am J Public Health* 1961; 51:155-64.
3. Logan WP, Glasg MD. Mortality in London fog incident, 1952. *Lancet* 1953;1:336-8.
4. Health effects of outdoor air pollution. Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society. *Am J Respir Crit Care Med* 1996; 153:3-50.
5. Lipfert F, Wyzga R. Air pollution and mortality: issues and uncertainty. *J Air Waste Manag Assoc* 1993;45:949-66.
6. Li Y, Roth H. Daily mortality analysis by using different regression models in Philadelphia county, 1973-1990. *Inhal Toxicol* 1995;7:45-58.
7. Schwartz J. Air pollution and daily mortality in Birmingham, Alabama. *Am J Epidemiol* 1995;137:1136-47.
8. Katsouyanni K, Toulomi G, Spix C, et al. 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. *BMJ* 1997;314:1658-163.
9. Samet JM, Zeger SL, Dominici F, et al. The National Morbidity, Mortality, and Air Pollution Study (HEI project no. 96-7): methods and methodological issues. Cambridge, MA: Health Effects Institute, 1999.
10. Dominici F, Samet JM, Zeger SL. Combining evidence on air pollution and daily mortality from the twenty largest US cities: a hierarchical modeling strategy (with discussion). *J R Stat Soc Ser A* 2000;163:263-302.
11. Hwang JS, Chan CC. Air pollution effects on daily clinic visits for lower respiratory illness. *Am J Epidemiol* 2002;155:000-00.
12. Lindley DV, Smith AF. Bayes estimates for the linear model (with discussion). *J R Stat Soc Ser B* 1972;34:1-41.
13. Morris CN, Normand S-L. Hierarchical models for combining information and for meta-analysis. In: Bernardo JM, Berger JO, Dawid AP, et al, eds. Bayesian statistics 4: proceedings of the Fourth Valencia International Meeting, April 15-20, 1991. Oxford, United Kingdom: Clarendon Press, 1992:321-44.
14. Gelman A, Carlin J, Stern H, et al. Bayesian data analysis. London, United Kingdom: Chapman and Hall Ltd, 1995.
15. Carlin BP, Louis TA. Bayes and empirical Bayes methods for data analysis. New York, NY: Chapman and Hall, Inc, 1996.
16. Thomas A, Spiegelhalter DJ, Gilks WR. BUGS: a program to perform Bayesian inference using Gibbs sampling. In: Bernardo JM, Berger JO, Dawid AP, et al, eds. Bayesian statistics 4: proceedings of the Fourth Valencia International Meeting, April 15-20, 1991. Oxford, United Kingdom: Clarendon Press, 1992:837-42.
17. Gilks WR, Richardson S, Spiegelhalter DJ, eds. Markov chain Monte Carlo in practice. London, United Kingdom: Chapman and Hall Ltd, 1996.
18. Dominici F, Daniels M, Zeger SL, et al. Air pollution and mortality: estimating regional and national dose-response relationships. (Technical report). Baltimore, MD: Johns Hopkins University, 2001.
19. Tierney L. Markov chains for exploring posterior distributions (with discussion). *Ann Stat* 1994;22:1701-62.
20. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986;7:177-88.

21. Samet JM, Zeger SL, Dominici F, et al. 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, 2000.
22. Katsouyanni K, Touloumi G, Samoli E, et al. Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. *Epidemiology* 2001;12:521–31.
23. Levy JJ, Hammitt JK, Spengler JD. Estimating the mortality impacts of particulate matter: what can be learned from between-study variability? *Environ Health Perspect* 2000;108:109–17.
24. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000;283:2008–12.
25. Macaskill P, Walter SD, Irwig L, et al. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 2001; 20:641–54.
26. Barnett V, O'Hagan A. Setting environmental standards: the statistical approach to handling uncertainty and variation. New York, NY: Chapman and Hall, Inc, 1997.
27. Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *Am J Epidemiol* 1999;150:469–75.