

**Review of the Proposed Weighting Scheme for Air Pollutants in the  
Greater Vancouver Regional District**

**Suresh H. Moolgavkar, M.D., Ph.D.  
Member, Fred Hutchinson Cancer Research Center  
Professor of Epidemiology, Adjunct Professor of Biostatistics and  
of Applied Mathematics, University of Washington**

**February 27, 2005**

This report was supported by a contract from CPPI to Dr. Moolgavkar and should not be construed as reflecting the views either of the Fred Hutchinson Cancer Research Center or of the University of Washington.

## **Executive Summary**

I review and critique a weighting scheme that has been proposed to estimate the monetary benefits of reductions in air pollution in the Greater Vancouver Regional District. This health-based air emission weighting approach is expressed in the following equation:  $\text{Impact Weighted Emissions} = 25 \cdot \text{PM} + \text{NO}_x + \text{HC} + \text{CO}/7 + 3 \cdot \text{SO}_2$ . The GVRD proposes to use this equation to prioritize emission reduction measures in two studies that are currently under development. These studies are: “Emission Reduction Options for Heavy Duty Diesel Vehicles in the Lower Fraser Valley” by Levelton Consultants Ltd. and “Reduction of Nonroad Diesel Emissions in the Lower Fraser Valley and the Rest of BC” by Genesis Engineering Inc.

Although I was unable to locate a specific rationale for the proposed scheme, it is clear that the scheme is based on assessing the health risks associated with air pollution and monetizing the health benefits that could be expected with reductions in air pollution. Whatever the rationale for the proposed scheme it is clear that it could not have taken into account the recent epidemiological literature on air pollution because the weighting scheme dates back to at least 2001. In particular, a software implementation problem was discovered in 2002, which cast into doubt the results of most time-series studies of air pollution and health. The weighting scheme could not have taken into account the revised analyses prompted by the discovery of this software problem.

I briefly review the recent epidemiological literature on air pollution. Because of the small risks associated with exposure to air pollution and the myriad confounders that must be controlled, interpretation of the results of epidemiological studies is problematic. In particular it is difficult, if not impossible, to single out any specific pollutant as being the principal bad actor. Additionally, concentration-response relationships cannot be reliably estimated. Thus, the current epidemiological literature cannot be used to support the proposed, or any, weighting scheme to estimate the benefits that might accrue from a reduction of air pollution levels.

## Introduction

I have been asked to evaluate and critique a weighted emissions factor approach that has been proposed to estimate the benefits of air pollution control measures in the Greater Vancouver area. My area of expertise is air pollution epidemiology and I will restrict my comments to the epidemiological evidence on the association between air pollution and health, and to whether or not the evidence supports the proposed, or any, weighting scheme. A recent report prepared for the British Columbia Lung Association by Bates et al. (2003) summarizes the epidemiological literature on air pollution and discusses methods for estimating the impact of air pollution on human health. The proposed weighted emissions factor approach predates this report by several years and, therefore, does not take its recommendations into consideration. However, the Bates report could be used to either support or revise the proposed approach. I have reviewed the Bates report and I believe that it greatly downplays the uncertainties in the epidemiological literature on the association between air pollution and health. It is my belief that we do not currently have the scientific information required to measure quantitatively the impact of air pollution control measures on human health.

In a memo dated November 2004, Genesis Engineering Inc. suggests a weighting scheme to estimate the economic benefits that would accrue from a reduction in specific pollutants. Explicitly, the proposed scheme is described by the equation,  $\text{Impact Weighted Emissions} = 25 \cdot \text{PM} + \text{NO}_x + \text{HC} + \text{CO}/7 + 3 \cdot \text{SO}_2$ . The rationale for this weighting scheme is not described, but a reference is provided to another document, a report entitled "Review of the AirCare On-Road (ACOR) Program, Final Report" prepared by G. W. Taylor Consulting in May 2002. This report again simply states the emission weighting equation given above without providing any justification, referring instead to a report prepared by the Sheltair Consulting Group in September 2001. I could find no rationale for the proposed weighting scheme in that report either, although it is presented there as if it were generally accepted. What is clear from these reports is that the weighting scheme is somehow derived from monetizing the estimated adverse health effects of some of the monitored components of air pollution.

From my brief review and unsuccessful attempts to discover the genesis of the proposed weighting scheme, one thing is clear: that it dates back to at least 2001. Thus, whatever its rationale, it is based on the epidemiological literature on air pollution prior to 2001. In May 2002, a potentially serious problem was discovered with the implementation of a software package widely used for analyses of most epidemiological studies of air pollution. This software implementation problem cast serious doubt on the analyses of most so-called time series studies of air pollution and human health conducted prior to May 2002. Thus, to the extent that the weighting scheme is based on time-series studies of air pollution, it could be based on analyses that are seriously flawed. In addition, many time-series studies were reanalyzed in response to the discovery of the software problem, and several important long-term studies of air pollution and mortality appeared after 2001. The results of the revised analyses of time-series studies and of the recent long-term studies were clearly not considered in the weighting scheme. I discuss both time-series and long-term studies of air pollution below.

Clearly the recommended weighting scheme is based on the belief that, of all the components of air pollution, the particulate matter (PM) fraction poses by far the greatest risk to human populations. Conclusions about the possible health effects of PM pollution come largely from epidemiological studies. Toxicological studies, while providing some interesting leads, have largely failed to elucidate mechanisms by which PM mass concentrations at the current low levels in North American cities could exert adverse cardiovascular and respiratory effects in human populations. Chamber studies likewise have, at best, produced equivocal results. The quantification of health benefits that would accrue from a reduction of ambient PM mass concentrations is derived entirely from the epidemiological studies. Because of the central role of epidemiology in the study of PM mass and human health I will discuss epidemiological studies of air pollution in some detail below. These studies certainly do not support the notion that PM mass is the most important component of air pollution insofar as human health is concerned. Methodological problems make the interpretation of epidemiological studies of multiple air pollutants difficult. Nevertheless, a number of studies report that PM mass is no more strongly associated with human health than are the gaseous pollutants. Still other studies suggest that the association of PM mass with human health is considerably weaker than the association of the gaseous pollutants with human health.

In summary, there is little guidance in the scientific literature to support any weighting scheme to estimate benefits that might accrue from reductions in air pollution.

### Epidemiological Studies of Air Pollution

The three major distinct, but inter-related, problems in interpretation of epidemiological studies of air pollution are the following. First, whichever study design is employed, myriad confounders must be adjusted. Proper adjustment is critical because the estimated risks are so small. Second, quite generally, effect estimates in epidemiological studies of air pollution are spatially rather heterogeneous, with estimates in one geographic region being quite different from those in another. Admittedly, consistency or heterogeneity of results is quite often in the eye of the beholder. The Bates report says, for example, that the results of studies are consistent. I look at the same body of literature and conclude that the results are far from consistent. Third, recent well-publicized examples indicate that observational epidemiological studies of small risks or benefits may yield seriously misleading results, probably because of residual confounding. This point is related to the first but is worth discussing separately.

There are few epidemiological studies of air pollution that measure exposure on an individual level by means of personal monitors. Most epidemiological studies of air pollution have no information on personal exposures and employ either an ecologic or semi-ecologic design. By far the most common ecologic studies of air pollution are time-series studies that investigate associations between daily fluctuations in air pollution levels and counts of adverse health effects, such as hospital admissions and deaths, in metropolitan areas. The focus of most, if not all, such studies in the past decade or so has been PM pollution. In these studies, the gaseous pollutants are considered mainly as

confounders of PM associations with the adverse health effects of interest. This focus on PM is driven largely, I believe, by the USEPA's regulatory agenda. In 1996, the USEPA proposed regulation of, and promulgated standards for, fine PM mass. Since then hundreds of time-series analyses of PM have appeared. A new Criteria Document for PM was approved by the Clean Air Scientific Advisory Committee (CASAC) of the USEPA in 2004, and new standards are expected to be promulgated in 2005. This unfortunate obsession with PM has come at the expense of a holistic approach to air pollution. Perhaps the single most influential time-series study of PM is the Health Effects Institute (HEI) sponsored National Mortality Morbidity and Air Pollution Study (NMMAPS) conducted by researchers at Johns Hopkins and Harvard Universities (Samet et al, 1995, 2000; Dominici et al., 2003).

The most widely cited semi-ecologic studies of air pollution are the Harvard Six Cities (Dockery et al., 1993) and the ACS II (Pope et al., 1995) studies. Recently published updates (Krewski et al., 2000; Pope et al., 2002) to these studies provide the best available information on the possible health effects of long-term exposure to air pollution.

### Time-Series Studies

With respect to confounding in time-series studies there are three major issues that must be addressed. First, can adequate adjustments be made for temporal trends in the health effect of interest due, for example, to temporal trends in the structure of the population or to episodic viral infections? Second, can the association of pollutants be teased apart from the effects of climate and weather? Third, can adequate statistical adjustments be made so that the associations of individual components of the air pollution mixture with adverse effects on human health can be reliably estimated? With respect to the third point, it is generally acknowledged that because the individual criteria pollutants are highly correlated, the teasing apart of their effects is difficult. Moreover, only a small fraction of all pollutants is routinely monitored so that each monitored pollutant is simply an index of the general pollution mixture.

In 2002, the generalized additive models (GAM) module in S-plus, the most commonly used software package for analyses of time-series studies, was found to yield misleading results when used with the default convergence criteria, casting doubt on the reported results of most time-series studies of air pollution. The revised analyses necessitated by the GAM convergence problems in S-plus, undertaken under the auspices of the Health Effects Institute (HEI), clearly indicate that methods used for controlling temporal trends and weather can have profound effects on the results of time-series analyses of air pollution data, as has been noted by the HEI expert panel (2003). To make matters even more difficult, there appears to be no objective statistical test to determine whether these factors have been adequately controlled in any analysis. The HEI Expert Panel for the reanalyses stated, "*Ritov and Bickel (1990) have shown, however, that for any continuous variable, no strictly data-based (i.e., statistical) method can exist by which to choose a sufficient number of degrees of freedom to insure that the amount of residual confounding due to that variable is small. This means that no matter what*

*statistical method one uses to select the degrees of freedom, it is always logically possible that even if the true effect of pollution is null, the estimated effect is far from null due to confounding bias.*” The expert panel goes on to say, “*Neither the appropriate degree of control for time, nor the appropriate specification of the effects of weather, has been determined for time-series analyses*”. In other words, it is impossible to adjust temporal trends or weather without accurate information from external sources regarding the appropriate degrees of freedom to be used. Such information simply does not exist. No conclusions can be drawn from time-series studies, particularly with respect to causality or concentration response relationships, unless the results are robust to extensive sensitivity analyses. Most time-series studies in the literature have undertaken only limited sensitivity analyses, if at all.

Large time-series studies conducted in Canada do not support the weighted emissions factor approach being recommended for the Greater Vancouver area. For example, Burnett et al. (2000) studied daily non-accidental mortality in 8 Canadian cities (Montreal, Ottawa, Toronto, Windsor, Winnipeg, Edmonton, Calgary, Vancouver). The novel features of the study are the attempt to investigate in detail the associations between chemical components of the particulate mass and mortality, and the attempt to take the total mortality attributable to air pollution and partition it between PM and gaseous fractions of air pollution. As is usual in studies of this type the generalized additive model was the analytical tool used. To minimize the effect of correlated covariates, the actual analyses were carried out using principal components. Results for individual pollutants were then obtained by applying a transformation. This aspect of the methodology appears to be new and requires evaluation. Regarding the air pollution mix the authors concluded, “*Size fractionated particulate mass explained 28% of the total health effect of the mixture, with the remaining effects accounted for by the gases.*” Thus, the most important conclusion of this study is that the gaseous components of air pollution dominate the association with mortality. In reanalyses of this study to address the GAM convergence issues, Burnett & Goldberg (2003) evaluated the sensitivity of the PM results to various degrees of smoothing and concluded, “*Particulate associations with mortality were highly sensitive to the type of smoother and the amount of smoothing.*” However, Burnett & Goldberg did not repeat their original analyses to assess the contribution of individual components of air pollution to mortality, although the qualitative results regarding the importance of gases are not likely to change. Thus, this important study of the largest metropolitan areas in Canada suggests strongly that the primacy accorded PM in the proposed weighting scheme is simply not supported by the data. Similar results have been reported in many other time-series studies conducted in North America (e.g., Moolgavkar, 2000, 2003).

## NMMAAPS

NMMAAPS was an ambitious effort, funded by the Health Effects Institute (HEI) and carried out by investigators at Johns Hopkins University (JHU) and Harvard University, to conduct comprehensive time-series analyses using a unified approach, of the association between PM<sub>10</sub> and mortality in the 90 largest metropolitan areas in the US, and between PM<sub>10</sub> and hospital admissions in a small subset of these. As such it

plays a prominent role in regulatory decisions and in shaping scientific opinion regarding the effect of air pollution on human health. The focus of NMMAPS is PM<sub>10</sub>. The association of other criteria pollutants with mortality and hospital admissions is not systematically considered; rather the other criteria pollutants are viewed as possible confounders of the PM associations. The individual estimates of risk obtained in city-specific analyses in the first stage are combined in a second stage using a Bayesian procedure to arrive at a single 'mean' estimate of risk. This approach ensures that identical models and lag structures are used for analyses and that confounding factors are treated in the same way. This approach raises its own problems, however. For example, is it appropriate to treat temperature and relative humidity in the same way in cities as disparate as Los Angeles and Chicago? The investigators concluded that, averaged over all cities, a 10 µg/m<sup>3</sup> increase in average PM<sub>10</sub> concentration on any given day resulted in a 'mean' increase in deaths of 0.4% on the next day.

After NMMAPS had been completed, the investigators at Johns Hopkins University discovered and reported the GAM convergence problem in S-plus, referred to above. They also undertook reanalyses of the NMMAPS data, with more stringent convergence criteria for GAM and also by using generalized linear models (GLM). The results of these reanalyses indicated generally that the estimates of PM associations were attenuated and the standard errors of these estimates became larger leading to a decrease in the significance of the PM coefficient. Hierarchical Bayes analyses of the new outputs estimated a 'mean' increase in mortality at a one-day lag of about 0.28% for a 10 µg/m<sup>3</sup> increase in PM<sub>10</sub> concentration with the more stringent convergence criteria for GAM and a 0.22% increase in mortality using GLM.

A careful look at the city-specific estimates before they are combined into a single 'mean' estimate of the PM effect on mortality reveals the following. First, as noted above, the estimates of PM associations with mortality are generally smaller in the revised analyses, with the GLM analyses yielding the smallest coefficients. The standard errors in the revised analyses are generally larger than in the original analyses. Second, many of the city-specific analyses yielding positive associations between PM and mortality in the original analyses show small negative associations in the revised analyses. In 11 cities that showed positive and statistically significant associations in the original analyses, these associations become insignificant in the revised analyses. By contrast, no cities with statistically insignificant associations in the original analyses show significant coefficients in the revised analyses. Third, only two cities, New York and Oakland, show positive and statistically significant associations between PM and mortality in the revised analyses, with one city, Little Rock, showing a significant negative association.

Although a formal statistical test fails to reject homogeneity, the estimated coefficients in the 90 cities appear to be quite heterogeneous, with individual city-specific coefficients ranging between -3.5 and 3. Tests of homogeneity have low power, however. It is of interest to note that in the original GAM analyses, the same test clearly rejected homogeneity of coefficients across the 90 cities. What changed in the revised analyses? For one thing the coefficients shrank towards zero and for another the standard errors

became larger. Thus, the result of the test for homogeneity in the revised analyses is a reflection of our ignorance: the uncertainty in the individual estimates is so large that the test cannot detect departures from homogeneity. In fact, on *a priori* biological grounds it is hard to see how the coefficients could fail to be heterogeneous in view of the differences in the air pollution mix and climatic conditions across the country. With the exception of New York, none of the largest metropolitan areas in the US show statistically significant associations between PM<sub>10</sub> and mortality despite 8 years of data. While most cities have monitoring data on PM<sub>10</sub> for only about a sixth of the days, some large cities, such as Chicago, have monitoring data on PM<sub>10</sub> on a daily basis with actual readings available on more than 80% of the days. Thus significance is clearly not an issue of power alone, since one would expect adequate power in the Chicago time-series to detect an effect if there is one. In the face of such heterogeneity, I believe it is inappropriate to combine city-specific coefficients to arrive at a single 'mean' estimate of risk.

The Bates report suggests that the decrease in the estimated PM effect on mortality from 0.4% in the original analyses to 0.22% in the revised analyses is inconsequential. Surely, this rather substantial change in the estimated coefficient for PM mortality effects should have a significant impact on any quantitative estimates of the benefits to be had from declines in ambient PM mass concentration? More importantly, the change in estimate illustrates a general principle, that estimated air pollution effects are sensitive to choice of model. I discuss the influence of model choice in more detail below. I point out here, however, that even the revised NMMAPS estimate of 0.22% is driven largely by the two cities (New York, Oakland) with statistically significant coefficients. When these cities are excluded from the analyses, the 'mean' estimate is 0.12% (Moolgavkar, in press).

The Bates report suggests also that the regional Bayes 'mean' for the Northwest US derived from NMMAPS is the most appropriate coefficient to use for the mortality impact of PM in the Frasier Valley. However, a careful review of the Bayes estimates reported by NMMAPS raises concerns about the interpretation of the regional means. Thus, the city-specific empirical Bayes estimates range between 0.188 and 0.271, while the Bayes estimates for 7 regions in the US range between 0.180 and 0.409. How could the range of regional effects be larger than the range of individual city effects? To make my point clear, the regional mean for the Northeast is reported to be 0.409. There are 15 cities in this region. The empirical Bayes estimates for these cities range between 0.223 and 0.271, and one would expect the mean in the region comprised by those cities to lie in that range. It turns out that, whereas the reported city-specific Bayes estimates were derived from a two-stage Bayes analysis, the regional means were estimated from a separate three-stage Bayes analysis. It is thus technically possible for the Bayesian regional means to be outside the range of empirical Bayes estimates for the individual cities. But what is the correct interpretation of an analysis in which the regional mean is so much higher than any of the city-specific means?

In summary, a number of time-series studies of air pollution continue to show associations between indices of air pollution and adverse effects on human health. There

is, however, considerable heterogeneity in the specific components of air pollution associated with health effects in different locations, and in the magnitude of the effects. Additionally, given the difficulties of adequately adjusting weather and temporal trends in time-series studies, these studies do not provide reliable estimates of the magnitude of the associations of individual components of air pollution with human health. Thus the time-series studies cannot be used to support the proposed, or any, weighting scheme to estimate the benefits of reductions in air pollution.

### Long-Term Studies

Whereas in time-series studies inferences are based on differences in ambient levels of pollution from day to day in a city, in long-term studies these inferences are based on differences in pollutant levels between cities. The Bates report suggests that the results of long-term studies reflect the effects both of the chronic exposure to high background levels of air pollution and of the acute excursions in daily levels of pollutants. In reality, of course, the results of these studies show only that the hazard function for mortality is correlated with pollution levels in cities. This correlation could either be a direct result of higher pollution levels leading to higher mortality or due to residual confounding by lifestyle and other factors, which are important determinants of mortality, and which are also correlated with air pollution. Many more confounders must be controlled in long-term studies than in time-series studies. For example, smoking, which is not a confounder in time-series studies, is an important confounder in long-term studies. Because air pollution is measured on a citywide level, any factor or agent that is associated with mortality and varies from city to city, such as lifestyle, is a confounder in the long-term studies. Control of confounding can be difficult with long-term studies. Adjustment for socio-economic factors is particularly difficult and there is little assurance residual confounding bias can be eliminated.

In the Harvard Six Cities Study (Dockery et al., 1993), a random sample of over 8,000 adults was selected from six cities in Northeastern and Midwestern United States. Proportional hazards regression modeling was used for analyses. Relative risks for mortality for residence in a particular city were estimated after adjustment for cigarette smoking, education and body mass index. In the city with highest level of pollution as measured by levels of  $PM_{2.5}$ , Steubenville, Ohio, the adjusted death rate was 26% higher than in the city with the lowest pollution, Portage, Wisconsin. If the six cities are ranked in order of adjusted death rates from lowest to highest, and if this ordering is compared with the ordering imposed by various indices of air pollution, the agreement is quite good, particularly if  $PM_{2.5}$  is used as an index of air pollution. Thus in this study there was good correlation between levels of fine particulate pollution and death rates, after adjustment for some important confounders measured on the individual level.

The investigators made no attempts, however, to adjust for potential city-level confounders, such as weather. In fairness, with only six cities in this study it would have been difficult, if not impossible, to make adjustments for ecologic confounders. Although the results of this study are consistent with an effect of air pollution, particularly fine particulate pollution, on death rates, they are equally consistent with the hypothesis that

part or all of the difference in death rates can be attributed to uncontrolled city-level, i.e., ecologic, confounders. For example, Lipfert (1995) has shown good correlation between the mortality rate ratios in the Six Cities Study and the percentage of people with a sedentary life-style in five of the six cities. Relevant data were not available in the sixth city (Topeka, KS). Moreover, the Six Cities Study used pollution levels only at a single point in time as their estimate of air pollution over the entire period of the study. The investigators made no attempt to consider the changing pollution profiles over time.

With only six cities in the study the investigators could not consider joint pollutant analyses. Thus, even if some fraction of the difference in death rates among the six cities is attributable to air pollution, it is not at all clear that it can be attributed to fine particulate pollution.

The ACS II Study (Pope et al., 1995) was a much larger long-term study, involving 151 cities and more than 100,000 individuals. The design of the ACS II Study was similar to the design of the Six Cities Study. The study was undertaken specifically to test the major hypothesis raised by the Six Cities Study – that fine particulate mass was associated with mortality. With 151 cities in the data base there was a real opportunity to adjust ecologic confounders, particularly copollutants. However, the investigators did not do so. Even if the ACS II Study was designed specifically to address the hypothesis raised by the Six Cities Study, there was little reason not to address the obvious deficiencies of that study. The relative risk estimated in the ACS II study was somewhat lower than that reported in the Six Cities Study.

The interpretation of the results reported in the Six Cities and ACS II studies is difficult because these studies took no account of ecologic covariates, including pollutants other than fine PM and sulfates. In order to address these concerns, the HEI funded thorough reanalyses of the data. The results of these reanalyses were published in 2000 (Krewski et al., 2000). The reanalyses were conducted in two phases. In the first phase of replication of the original studies, the HEI investigators audited the data carefully and analyzed them using the methods used by the original investigators. Since the original papers were published by careful and reputable investigators, it is not surprising that the HEI phase 1 reanalyses reported results that were essentially identical to those obtained by the original investigators. In a recent follow-up of the Harvard Six Cities Study, however, Villeneuve et al. (2002) used Poisson regression models to investigate the association between time-dependent measures of PM and mortality. The most interesting result of these analyses was the considerably lower RR among the elderly (age greater than 60) than among those less than 60. If the elderly are particularly susceptible to air pollution as has often been argued and as suggested in the Bates report – and indeed higher relative risks have been reported for the elderly in many time-series studies - why is the RR among the elderly in this study substantially lower than the RR among those under 60?

Of equal or greater interest are the sensitivity analyses (phase 2) conducted by the HEI investigators on the ACS II data (sensitivity analyses were not possible on the six cities data because of the small size of the study). This phase

explicitly considered a number of ecologic confounders, including copollutants. The most important findings of the sensitivity analyses were, 1) effect modification by level of education; 2) attenuation of the particle effect when spatial correlation was considered; and most importantly 3) the substantial attenuation of the PM associations when SO<sub>2</sub> was considered simultaneously in two-pollutant analyses. Remarking on the spatial models used in the reanalyses, Smith et al. (2001) say in a report submitted to EPA, *“In spite of the incomplete nature of the spatial analysis, it did have a significant impact on the results....If such a substantial change is possible through only a one-parameter addition to the model, it can only be speculated what would happen with more realistic spatial models.”*

When sulfur dioxide was considered along with PM in the model for all-cause mortality, the coefficient for sulfates was reduced from 0.17 to 0.05 and that for fine particles was reduced from 0.18 to 0.03, and both became statistically insignificant. It is also of interest to note as Smith et al. point out, that consideration of spatial correlations attenuated the PM coefficients to a much greater extent than the coefficients for SO<sub>2</sub>, *“For example, in an analysis including both sulfate particles and SO<sub>2</sub> (Krewski et al. (2000), pp. 210-211), the RR for sulfate dropped from 1.20 to 1.08 (95% CI: 0.91 to 1.28) though that for SO<sub>2</sub> was less affected (RR from 1.35 to 1.31; CI 1.12 to 1.50).”*

In a more recent study of the ACS II cohort that doubles the follow-up time and triples the number of deaths, Pope et al (2002) report significant associations between fine particles and oxides of sulfur with all-cause, cardiovascular and lung cancer mortality. The risks reported in this paper appear to be quite a bit smaller than those in the previous Krewski reanalyses, although a direct comparison is difficult because it is not clear that identical models were used in the two sets of analyses. The risks for all-cause and cardiovascular mortality are lower than those reported in the original 1995 analysis, whereas the risk for lung cancer mortality is considerably higher. Surprisingly, despite the findings in the Krewski analyses that SO<sub>2</sub> was the pollutant most strongly associated with mortality, no joint pollutant analyses were carried out. Thus this paper leaves unaddressed the question of which of these two pollutant classes is more strongly associated with mortality, which is probably the single most important question raised by the Krewski reanalyses of the ACS II Study. In contrast to the earlier Krewski et al. reanalysis of ACS II, this analysis reports no attenuation of the PM coefficient with spatial adjustment. The reason for this discrepancy is not clear.

Sulfur dioxide is known to be scrubbed out in the upper airways and the consensus of scientific opinion is that it could not be directly responsible for the observed mortality in the ACS II cohort. Thus the SO<sub>2</sub> association may reflect either an association with a mix of pollutants for which the gas is a surrogate, or it may reflect residual confounding. If the reported positive associations between air pollution and mortality in these reanalyses are due to uncontrolled confounding, what are the possible confounders? Two strong candidates are changing smoking habits and changing life-style factors. We know that there have been profound changes in life-style and smoking habits over the period of this study. People generally are eating better, exercising more, and smoking less. These life-style changes are more likely to be adopted by the more affluent, better-

educated communities, which are also exposed to less air pollution. Thus, the association between either PM or SO<sub>2</sub> and mortality may simply reflect the impact of changing lifestyle factors, including changes in smoking habits, on mortality. In particular, smoking is such a strong risk factor for mortality that controlling changing habits well enough to assure absence of residual confounding would be extremely difficult. The reanalyses of the original ACS study attempted to account for changing smoking habits. I would like to note here, however, that, in addition to the inherent difficulties of incorporating a detailed enough history of this strong risk factor, the Cox proportional hazards model, which was the main analytic tool used, is not particularly well-suited to analyses of epidemiological data with time-varying covariates. Specifically, in most implementations of the Cox model the hazard function behaves like a step function and changes instantaneously when covariates change. A glance at the hazard functions among ex-smokers clearly shows that the risk does not decrease in this fashion.

Based largely on the Six Cities and the ACS II studies and their updates the Bates report (table 3, page 37) includes lung cancer as one of the known health consequences of exposure to PM pollution. An effect of air pollution, and particularly PM pollution from combustion sources, on lung cancer is certainly biologically plausible. In an early publication Stevens and I (Stevens & Moolgavkar, 1984) have shown that the rates of lung cancer among non-smokers declined in England and Wales following the passage of the Clean Air Act. The interpretation of these results as causal effects of air pollution on lung cancer is questionable, however. In both the Six Cities and the ACS II studies, the reported risk of lung cancer associated with PM pollution is far too high as a back-of-the-envelope calculation shows (Moolgavkar, in press). In the ACS II reanalyses, Krewski et al. (2000) report *statistically significant elevated risks of cancers other than lung cancer* associated with exposure to particulates, and in many of the analyses the risks for cancers other than lung cancer are equal to or larger than the risk for lung cancer. This unexpected finding suggests that the reported association of fine PM with lung cancer be interpreted with caution.

The Washington University-EPRI Veterans Study (Lipfert et al., 2000) is another large long-term study of air pollution and all-cause mortality, but is not considered in the Bates report. The cohort consists of approximately 50,000 U.S. veterans who were diagnosed with hypertension in the mid 1970s. The cohort had an average age of about 51 at recruitment, is all male, and is about 65% white and 35% nonwhite. In addition to air pollution variables based on county of residence, which were considered in some detail, limited information on individual level covariates, such as smoking was also available and is included in the analyses. In contrast to the original Six Cities and ACS II studies, all available pollutants were considered equally in the analyses. As in the Harvard Six Cities and the ACS II studies, the basic analytic tool was Cox proportional hazards regression. Four different exposure and three different mortality periods were considered yielding a total of 12 distinct exposure and mortality period combinations. Among the pollutants, the strongest associations are reported for NO<sub>2</sub> and peak ozone. Of these two pollutants, ozone showed the stronger association with mortality, although there was an indication of a threshold at about 0.14 ppm. No significant PM effects are reported with any exposure metric used (TSP, PM<sub>10</sub>, sulfates, PM<sub>2.5</sub>). The authors point out, however,

*“It must be recognized that all potentially harmful pollutant species are not measured routinely and thus cannot be included in epidemiology studies of this type. For this reason, those pollutants that are included should be considered as indices of the overall urban pollution mix. Further the nature of this mix has changed significantly during the period evaluated in this study.”*

In summary, interpretation of results of long-term studies of air pollution and mortality is not straight forward. If the results of the largest study, the ACS II study including its reanalyses, are taken at face value, the pollutant most strongly associated with mortality is not fine PM but SO<sub>2</sub>. There are, moreover, questions regarding the adequacy of the simple spatial adjustments made in the Krewski et al. reanalyses. The reanalysis of the Six Cities Study by Villeneuve et al. (2002) reports smaller risks for the elderly, which is contrary to expectation. The lung cancer risks in the Six Cities and ACS II analyses appear to be too large, and the association of fine PM with cancers other than lung cancer in the Krewski reanalyses raises issues regarding the interpretation of the lung cancer associations. Finally, the Veterans Study reports no association between PM pollution as measured by multiple metrics and mortality in a susceptible population of hypertensive individuals.

#### Concentration-Response Relationships

Information on concentration-response relationships is required if one wishes to estimate benefits that would accrue from a reduction in air pollution. The assumptions regarding concentration-response that were used to derive the proposed weighting scheme were not available to me. The Bates report focuses on mortality associated with the PM component of air pollution saying, *“One reason for the Panel’s preference for selecting PM as the principal air pollutant is to avoid potential double counting of effects that are in fact redundant among several co-pollutants. Another reason is that the PM literature as a whole (whether for single- or multi-pollutant analyses) is more convincing in the Panel’s opinion than that relating the gaseous pollutants with mortality and morbidity outcomes.”*

The Bates report argues that linearity has been demonstrated for PM associated health effects and recommends the use of coefficients (table 5, page 51) derived from specific time-series and long-term studies to estimate the effects of short- and long-term effects of PM on mortality. Although arguments, summarized in the Bates report, have been made for linearity of the concentration-response function, I believe that in many data sets, the concentration-response functions are highly non-linear. In fact, the shapes of the concentration-response functions in many analyses are highly implausible from a biological perspective. An example is provided by a recent analysis in Cook County (Moolgavkar, 2003), where there appears to be a leveling off or actual decline in risk with increasing concentration of PM<sub>10</sub> above about 50 µg/m<sup>3</sup>. To directly investigate this suggestion, I analyzed total mortality by lag time (i.e. that day after the pollutant measurement at which mortality was compared with the PM concentration) in Cook County restricted to the days on which concentrations of PM<sub>10</sub> exceeded 50 µg/m<sup>3</sup>. There were approximately 610 days (excluding the days on which readings were not available)

over the entire period on which concentrations exceeded  $50 \mu\text{g}/\text{m}^3$ . The results of these analyses are presented in table 1 below. As suggested by the full analyses, the coefficient for each lag from 0 to 5 days is either negative or quite small and insignificant.

**Table 1:** Results of GAM analyses of full-year total non-accidental mortality in Cook County restricted to the days on which concentrations of  $\text{PM}_{10}$  exceeded  $50 \mu\text{g}/\text{m}^3$ . For lags between 0 and 5 days the estimated percent changes ( $\log \text{RR} \times 100$ ) in daily deaths associated with a  $10\mu\text{g}/\text{m}^3$  change in  $\text{PM}_{10}$  concentration is shown. The t-statistics are shown in parentheses. t-statistics with absolute values larger than 1.96 are significant at the 0.05 level.

| Lag 0       | Lag 1         | Lag 2       | Lag 3       | Lag 4        | Lag 5         |
|-------------|---------------|-------------|-------------|--------------|---------------|
| 0.04 (0.18) | -1.15 (-0.45) | 0.01 (0.02) | 0.15 (0.61) | -0.1 (-0.39) | -0.27 (-1.13) |

Because of the strong non-linearities in the concentration-response relationship evident in many analyses, the estimated coefficient from a linear fit cannot be interpreted as a measure of the increase in some adverse health effect to be expected for a unit increase in PM.

#### The Critical Importance of Model Choice

Issues of confounding of air pollutant associations by temporal trends, weather, and co-pollutants can be more generally discussed under the rubric of model choice. The Bates report does not acknowledge that the uncertainties in the estimates of PM effects (or those of any other pollutant) are almost certainly understated by consideration of the statistical uncertainty computed under the fitted model alone. Much more uncertainty derives from the lack of information regarding the choice of appropriate models for adjusting confounding by other covariates, and the choice of appropriate lag structures. As Lumley and Sheppard (2003) point out, *“Estimation of very weak associations in the presence of measurement error and strong confounding is inherently challenging. In this situation, prudent epidemiologists should recognize that residual bias can dominate their results. Because the possible mechanisms of action and their latencies are uncertain, the biologically correct models are unknown. This model selection problem is exacerbated by the common practice of screening multiple analyses and then selectively reporting only a few important results.”*

Others have made similar comments regarding the critical importance of model choice, particularly when effect estimates are small. For example, in comments on the Criteria Document for PM submitted to the EPA in 2001, Smith, et al., say, *“From a statistical point of view, the common epidemiological practice of choosing variables (including lagged variables, co-pollutants, etc.) that maximize the resulting effect estimates is a dangerous approach to model selection, particularly when the effect estimates are close to 0 (i.e. RR close to 1). As has been demonstrated in Lumley and Sheppard (2000), the effect of choosing lags for PM in this fashion has a bias which is of the same order of magnitude as the relative risk being estimated.”*

More recently others have expressed similar concerns in the peer-reviewed literature. In a recent publication, Koop and Tole (2004) used Bayesian Model Averaging (BMA) to investigate daily mortality in Toronto and report that when model uncertainty is considered there is little evidence of a PM association with mortality. They conclude, *“The main empirical finding of the paper is that standard deviations for air pollution-mortality impacts become very large when model uncertainty is incorporated into the analysis. Indeed they become so large as to question the plausibility of the previously measured links between air pollution and mortality.”* A more complete discussion of model uncertainty can be found in a recent paper (Moolgavkar, in press).

### The Limits of Observational Epidemiology

With limited toxicological evidence of adverse health effects of PM, almost the entire case against PM rests on the epidemiology. A question of particular relevance then is how reliably observational epidemiology studies can detect small risks. Particularly when risks (or benefits) are small, residual confounding and unrecognized sources of bias can alter profoundly the conclusions drawn from observational epidemiological studies, as was noted by science writer Gary Taubes in an article, ‘Epidemiology Faces its Limits’ published in 1995. Based on interviews with eminent epidemiologists on both sides of the Atlantic, Taubes suggests that observational epidemiology studies are not capable of reliably detecting relative risks much smaller than 2. I emphasize here that air pollution epidemiology studies are attempting to estimate relative risks much smaller than 2, approximately 1.2 in the cohort (long-term) studies and 1.003 (sic) in the time-series studies.

The limitations of observational epidemiology studies are dramatically illustrated by the case of postmenopausal hormone replacement therapy (HRT). Based largely on observational epidemiology studies bolstered by ‘biological plausibility’ HRT became one of the most prescribed therapies in the United States, *“with a highly diversified portfolio of presumed benefits for post-menopausal women”* (Herrington & Howard, 2003), including a decreased risk of coronary heart disease. Recent randomized trials of HRT (a randomized trial is the only way to assure that residual confounding has been eliminated) showed that this therapy does not decrease the risk of coronary heart disease among postmenopausal women (Rossouw et al., 2002, Barrett-Connor, 2003); in fact, there is a suggestion that HRT may actually increase the risk. One possible explanation of the contradictory findings in the observational studies and the randomized trials is residual confounding by life style factors in the observational studies. A more complete discussion can be found in a recent paper (Moolgavkar, in press).

### Conclusions

Because of the small risks being estimated in epidemiological studies of air pollution, and the problems of model choice exemplified by the myriad confounders that need to be controlled in both time-series and long-term studies, I do not believe that it is currently possible to estimate the benefits that would accrue from a reduction in levels of air pollution in North American cities, and it is inappropriate to ascribe relative rankings

of risk to individual air pollutants. My discussion in this report has focused on epidemiological studies of PM because the literature over the last many years has focused on PM. While many studies continue to report associations between indices of air pollution and adverse effects on human health, even at the generally low levels of pollution in North American cities, the causality of these associations is open to question for the reasons I have briefly discussed in the body of this report. While prudence dictates that continuing efforts be made to lower ambient air pollution levels in North American cities, no quantitative estimates of resulting health benefits, if any, can currently be made. There is simply no justification in the literature for the particular weighting scheme that is being proposed to estimate the benefits of air pollution control measures in the Greater Vancouver area.

### References

Barrett-Connor E. Clinical Review 162 Cardiovascular Endocrinology 3. An epidemiologist looks at hormones and heart disease in women. *J Clin Endocrinol Metab* 88:4031-4042, 2003.

Bates DV, Koenig J, Brauer, M. Health and Air Quality 2002-Phase 1, Final Report, Prepared for the British Columbia Lung Association. May, 2003.

Burnett, R. T.; Brook, J.; Dann, T.; Delocla, C.; Philips, O.; Cakmak, S.; Vincent, R.; Goldberg, M. S.; Krewski, D. Association between particulate- and gas-phase components of urban air pollution and daily mortality in eight Canadian cities. In: Grant, L. D., ed. *PM2000: Particulate Matter and Health*. *Inhalation Toxicol.* 12: 15-39, 2000.

Burnett, R. T.; Goldberg, M. S. Size-fractionated particulate mass and daily mortality in eight Canadian cities. In: Revised analyses of time-series studies of air pollution and health. Special report. Boston, MA: Health Effects Institute; pp. 85-90, 2003.

Dockery, D. W.; Pope, C. A., III; Xu, X.; Spengler, J. D.; Ware, J. H.; Fay, M. E.; Ferris, B. G., Jr.; Speizer, F. E. An association between air pollution and mortality in six U.S. cities. *N. Engl. J. Med.* 329: 1753-1759, 1993.

Dominici F, McDermott A, Daniels M, Zeger SL, Samet JM. Mortality among residents of 90 cities. In *Revised Analyses of the National Morbidity, Mortality, and Air Pollution Study, Part II. Health Effects Institute Special Report, Revised Analyses of Time-Series Studies of Air Pollution and Health*. Health Effects Institute, May 2003.

Health Effects Institute Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute, Cambridge, 2003.

Herrington DM & Howard TD. From presumed benefit to potential harm – hormone therapy and heart disease. *N Engl J Med*, 349:519-521, 2003.

Genesis Engineering Inc. Proposed Emissions Weighting, November 2004.

GW Taylor Consulting, Energy and Environmental Analysis, Inc., Rowan Williams Davies & Irwin West, Inc. Review of the AirCare On-Road (ACOR) Program, Final Report. May, 2002.

Koop G., Tole L. Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air? *J. Environ Econ and Management*, 47:30-54, 2004.

Krewski, D.; Burnett, R. T.; Goldberg, M. S.; Hoover, K.; Siemiatycki, J.; Jerrett, M.; Abrahamowicz, M.; White, W. H. Reanalysis of the Harvard Six Cities study and the American Cancer Society study of particulate air pollution and mortality. A special report of the Institute's Particle Epidemiology Reanalysis Project. Cambridge, MA: Health Effects Institute, 2000.

Lipfert FW. Estimating air pollution –mortality risks from cross-sectional studies: prospective vs. ecologic study designs. In, “Particulate Matter: Health and Regulatory Issues”, AWMA Symposium Proceedings , AWMA, Pittsburgh, PA 1995.

Lipfert, F. W.; Perry, H. M., Jr.; Miller, J. P.; Baty, J. D.; Wyzga, R. E.; Carmody, S. E. (2000) The Washington University-EPRI veterans' cohort mortality study: preliminary results. In: Grant, L. D., ed. PM2000: particulate matter and health. *Inhalation Toxicol.* 12: 41-73, 2000.

Lumley T, Sheppard L. Time series analyses of air pollution and health: straining at gnats and swallowing camels? *Epidemiology* 14:13-14, 2003.

Moolgavkar, S. H. Air Pollution and Mortality in Three U.S. Counties. *Environ. Health Perspect.* 108: 777-784, 2000.

Moolgavkar, S. H. Air pollution and daily deaths and hospital admissions in Los Angeles and Cook counties. In: Revised analyses of time-series studies of air pollution and health. Special report. Boston, MA: Health Effects Institute; pp. 183-198 2003. Available: <http://www.healtheffects.org/news.htm> [16 May, 2003].

Moolgavkar SH. Air pollution and daily mortality in two U.S. counties: season-specific analyses and exposure-response relationships. *Inhalation Toxicology*, 15:877-907, 2003.

Moolgavkar SH, A review and critique of the EPA's rationale for a fine particle standard, *Reg. Toxicol and Pharmacol*, in press.

Pope, C. A., III; Thun, M. J.; Namboodiri, M. M.; Dockery, D. W.; Evans, J. S.; Speizer, F. E.; Heath, C. W., Jr. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am. J. Respir. Crit. Care Med.* 151: 669-674, 1995.

Pope, C. A., III; Burnett, R. T.; Thun, M. J.; Calle, E. E.; Krewski, D.; Ito, K.; Thurston, G. D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA J. Am. Med. Assoc.* 287: 1132-1141, 2002.

Rossouw JE, Anderson GL, Prentice RL et al. Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*, 288:321-333, 2002.

Samet JM, Zeger SL, Berhane K. The association of mortality and particulate air pollution. In: *The Phase 1a Report of the Particle Epidemiology Evaluation Project*. Pp 3-104. Health Effects Institute, Cambridge, MA. 1995.

Samet JM, Zeger SL, Dominici F, et al. The national morbidity, mortality, and air pollution study, part II: Morbidity and mortality from air pollution in the United States. Research Report 94. Health Effects Institute, Cambridge, MA, 2000.

Smith R, Guttorp P, Sheppard L, Lumley T, Ishikawa N. Comments on the Criteria Document for Particulate Matter Air Pollution, NRSCE Technical Report Series #66, July 25, 2001. Available: <http://www.nrcse.washington.edu/research/reports.html>

Stevens RG, Moolgavkar SH: A cohort analysis of lung cancer and smoking in British males. *Am J Epidemiol* 119:624-641, 1984.

Taubes G. Epidemiology faces its limits. Special News Report. *Science*, 269:164-169, 1995.

The Sheltair Group Resource Consultants, Inc., Alchemy Consultants, Inc. Phase 2 Final Report: Harmonized Measures for Reducing Greenhouse Gases and Air Pollution in the LFV. September 2001.

Villeneuve PJ, Goldberg MS, Krewski D, Burnett RT, Chen Y. Fine particulate air pollution and all-cause mortality within the Harvard Six-Cities study: variations in risk by period of exposure. *Ann Epidemiol.* 12:568-576, 2002.