Statistical Issues in the Study of Air Pollution Involving Airborne Particulate Matter

Lawrence H. Cox

NRCSE

Technical Report Series

NRCSE-TRS No. 041

January 10, 2000

The **NRCSE** was established in 1996 through a cooperative agreement with the United States Environmental Protection Agency which provides the Center's primary funding.

STATISTICAL ISSUES IN THE STUDY OF AIR POLLUTION INVOLVING AIRBORNE PARTICULATE MATTER

Lawrence H. Cox
U.S. Environmental Protection Agency
National Exposure Research Laboratory (MD-75)
Research Triangle Park, NC 27711 USA
COX.LARRY@EPAMAIL.EPA.GOV

Abstract

Epidemiological research in the early 1990s focusing on health effects of airborne particulate matter pointed to a statistical association between increases in concentration of particles in ambient air and increases in daily nonaccidental mortality, particularly among the elderly. These results appear consistent across a range of U.S. cities. This and other scientific and policy information formed the basis on which the U.S. Environmental Protection Agency (USEPA) promulgated revised, stricter air quality standards for particulate matter in 1997. The schedule for implementing the revised standards is coincident with completion by USEPA of a second National Ambient Air Quality Standards (NAAQS) review for ambient particulate matter by July 2002 based on current science and policy information. Concurrently, the U.S. Congress directed USEPA to seek advice from the National Academy of Sciences, resulting in formation of the National Research Council Committee on Research Priorities for Airborne Particulate Matter. In its 1998 report, the Committee identified ten research priorities for setting regulatory standards for particulate matter. The tenth priority addressed statistical issues. To explore these issues, the National Research Center for Statistics and the Environment and the USEPA organized a 1998 interdisciplinary Workshop on Particulate Methodology. A primary objective of the Workshop was to initiate a statistical research program relevant to setting air quality standards for ambient particulate matter pollution to be completed during the second NAAQS review. This paper reports findings and recommendations from the Workshop and an agenda for statistical research motivated therefrom relevant to the study of air pollution involving particulate matter and to setting particulate matter air quality standards.

Keywords: air quality, particulate matter, NAAQS review, statistical science, epidemiology

1. INTRODUCTION

Airborne particulate matter refers to materials of varying size and composition present in ambient air in the form of solid particles or liquid droplets. Particulate matter and five airborne chemicals (carbon monoxide, sulfur dioxide, nitrogen dioxide, ozone, and lead) comprise the group of *criteria air pollutants* regulated by the United States Environmental Protection Agency (USEPA) under the Clean Air Act and Amendments (42 USC Sec. 7401 et seq.). Under the Act, USEPA promulgates specific National Ambient Air Quality Standards (NAAQS) for the criteria pollutants. The Act also requires USEPA to review each NAAQS every five years on a scientific and policy basis and to promulgate revisions as needed. For current regulatory purposes, particulate matter is classified on an individual "particle" basis, in terms of the particle's aerodynamic diameter, measured in micrometers.

USEPA, formed in 1970, began regulating particulate matter in 1971, based on the concentration (mass per unit volume of ambient air) of *total suspended particles* (TSP) in ambient air. The maximum size of particles collected by devices monitoring compliance with the first particulate matter regulatory standard was approximately 30 micrometers. Subsequently, a Federal Reference Method (FRM) was employed to provide standards for air quality monitoring devices used for regulatory purposes. The FRM for particulate matter is defined in terms of a *50% upper cut point*, i.e., the particle size at which 50% of the particles at this size will be collected by the monitoring device (thus, smaller/larger particles have a larger/smaller associated percentage). In 1987, USEPA changed the regulatory basis for the particulate matter standard from TSP to particulate matter with a 50% upper cut point of aerodynamic diameter 10 micrometers, denoted *PM_{10}*. Thus, PM_{10} mostly comprises particles smaller than 10 micrometers, but contains a decreasing percentage of larger particles, viz., from 50% at 10 micrometers down to 0% at approximately 20 micrometers. The 1987 Particulate Matter NAAQS required that at each monitoring site within a pollution control district and within any contiguous three calendar year period: (a) the three-year average of the annual maximum daily concentration measurements of PM_{10} at the site not exceed 150 micrograms per cubic meter (of ambient air), and (b) the three-year average of the daily PM_{10} concentration measurements at the site not exceed 50 micrograms per cubic meter.

Epidemiological studies over the past decade suggest the existence of a persistent statistical association between increases in daily ambient particulate concentration and increases in mortality (typically, nonaccidental mortality in the over-65 age group) in a variety of U.S. cities. PM_{10} is composed of *fine particles* and *thoracic coarse particles*. *$PM_{2.5}$*, particles with an upper 50% cut point of 2.5 micrometers, is used as an indicator of fine particles, even though the size range 1 - 2.5 micrometers contains some thoracic coarse particles. Because the Harvard Six Cities Study (Dockery et al. 1993), the first study to examine the relationship between particulate matter and mortality for both $PM_{2.5}$ and $PM_{15\text{-}2.5}$, found a better relationship with mortality for $PM_{2.5}$ than for $PM_{15\text{-}2.5}$, recent research has focused on $PM_{2.5}$ (more often than not, erroneously referred to as "fine particles"). In the same time frame, USEPA was conducting its periodic review of the Particulate Matter NAAQS (see: U.S. Environmental Protection Agency 1996), resulting in a small revision to the PM_{10} standard (basing the 3-year average on the annual 99^{th} percentile instead of the annual maximum) but most significantly introducing a new NAAQS regulating $PM_{2.5}$: (a) the three-year average of the annual 98^{th} percentile of $PM_{2.5}$ concentration measurements at any monitoring site should not exceed 50 micrograms per cubic meter of ambient air, and (b) the arithmetic mean (over one or multiple monitoring sites in the region) of site-specific three-year averages of daily $PM_{2.5}$ concentration measurements should not exceed 15 micrograms per cubic meter.

USEPA promulgated revised air quality standards for particulate matter on July 18, 1997 (see: U.S. Federal Register 1997). Concurrently, U.S. President Clinton directed USEPA to complete the next Particulate Matter NAAQS review within the five-year statutory period required by the Clean Air Act, viz., by July 2002, and incorporated the PM NAAQS within the implementation plan for the revised particulate matter standards (Clinton 1997). Full implementation of the revised standards requires establishing $PM_{2.5}$ monitoring networks, determining whether or not a location is in or out of compliance with the standard(s), and, if out, developing a State Implementation Plan (SIP) to achieve compliance. This inevitably requires considerable effort and time, during which the second NAAQS review can extend and keep up to date the scientific basis for revising the standards.

Also at this time, Congress directed USEPA to request the National Research Council (NRC) to conduct a study to identify research priorities relevant to setting regulatory standards for ambient particulate matter. NRC responded by forming the Committee on Research Priorities for Airborne Particulate Matter, which quickly produced its first of four planned reports, **Research Priorities for Airborne Particulate Matter: I. Immediate Priorities and a Long-Range Research Portfolio**, listed in the references as National Academy of Sciences (1998) but henceforth referred to as "the NRC Report".

The revised standards set in motion a range of science and policy discussions, as well as several lawsuits. Recently, several of these lawsuits were adjudicated (see: U.S. Court of Appeals 1999). These legal issues are beyond the scope of this paper.

Evident from their form, environmental standards rely upon several statistical concepts–sampling, measurement, estimation and prediction–and at the same time standards raise important statistical issues–bias and representativeness, variability, uncertainty and precision. Particulate matter standards in particular evolved in large measure from epidemiological and other scientific studies representing a plethora of statistical designs and models. Any list of research priorities or needs involving particulate matter standards should involve statistical research. The NRC Report recommended statistical research under its Research Topic 10, described in the next section.

Soon after the release of the NRC Report, the National Research Center for Statistics and the Environment (NRCSE) and the USEPA Office of Research and Development decided to organize a workshop focused on Topic 10. With cosponsorship from the National Institute of Statistical Sciences and the Health Effects Institute, the NRCSE/USEPA Particulate Methodology Workshop was held October 19-22, 1998 at the University of Washington, Seattle, WA. The purpose of the Workshop was to bring together an interdisciplinary group of statistical and other scientists in pursuit of the following objectives: to illuminate statistical issues articulated or raised by Topic 10, to identify priority statistical research bearing upon these issues, and to organize interdisciplinary research projects on these topics, targeted for completion prior to the end of the second Particulate Matter NAAQS review. Judging from timely publication of this *Environmetrics* special issue, it would appear that Workshop objectives were met.

The Workshop was organized around formal presentations, discussion, and poster presentations. Presentations covered measurement, atmospheric transport, and modeling of particulate matter; understanding and developing models of particulate matter exposure and health effects; particle transformation; source apportionment; regulatory issues; and, statistical research questions and findings stemming from particulate matter studies. The Workshop program is available on the World-Wide Web at address http:www.nrcse.washington.edu/events/pm-workshop.html.

The core of the Workshop was the deliberations of working groups. Topics for six working groups were defined to cover the Topic 10 questions. A seventh, on case-crossover studies (Navidi et al. 1999), developed spontaneously and is not reported here (cf. the contribution by Lumley and Levy 2000). The purpose of this paper is to present the six working group conclusions and recommendations and opportunities for statistical research relevant to particulate matter. Section 2 summarizes Item 10 of the NRC Report. Section 3 presents the working group conclusions and recommendations, based on summaries prepared by workgroup reporters. Statistical methodology featured prominently in the Workshop, but references thereto were not provided and likewise are not provided here. Selected statistical research opportunities identified at the Workshop are discussed in Section 4. Section 5 contains concluding comments. Summaries of working group discussions and a public discussion board on particulate matter statistical methodology are available on the world-wide web at http://www.nrcse.washington.edu/discus/default.htm

2. RESEARCH TOPIC 10 OF THE NRC REPORT

The NRC Report identified ten high priority research areas for airborne particulate matter. Opportunities for contributions from statistical science can be found in several of the ten areas. Research Topic 10: Analysis and Measurement, deals most comprehensively and almost exclusively with statistical issues. Workshop working groups were organized around topics designed to include all research questions listed under Topic 10. These questions are summarized in the following excerpts from **Research Priorities for Airborne Particulate Matter**. Copyright 1998 by the National Academy of Sciences. Courtesy of the National Academy Press, Washington, DC. Questions posed in Research Topic 10 of the NRC Report have been numbered for ease of reference.

The first set of questions deals with statistical analysis (NRC Report, pp. 92- 93):

To what extent does the choice of statistical methods in the analysis of data from epidemiological studies influence estimates of health risks from exposures to particulate matter? Can existing methods be improved?

SA1. What methods of removing the influence of long-term trends from parallel data on daily particulate-matter concentrations and population morbidity and mortality are most appropriate?
SA2. Should time-series of health and/or environmental data be filtered before analysis? What filters are most appropriate for this type of analysis?
SA3. What is the nature of the autocorrelation function in time-series studies? How should the autocorrelation be taken into account in the analysis of time series data?
SA4. How can the critical timing of exposure (e.g., frequency or duration) for particulate-matter-related morbidity and mortality be determined?

SA5. How can the unique health effects of particulate matter and its biologically important constituents be determined in the presence of exposure to multiple copollutants?
SA6. Are existing epidemiological data adequate to identify the most relevant timing characteristics of exposure?
SA7. Is residual confounding a concern for particulate-matter epidemiological studies?
SA8. What types of exposure-response models are most appropriate to describe the observed relationships between mortality and morbidity and exposure to ambient particulate matter?
SA9. How can key covariates, including potential confounders and modifying factors, be best incorporated into risk models used to describe the effects of particulate matter on population health?
SA10. How can the effects of long and short-term exposures to particulate matter on mortality, including reduction in life expectancy, best be estimated?
SA11. Could the positive associations between particulate matter and adverse health outcomes, as observed in time-series studies, be false positives resulting from multiple statistical tests using various regression models?

The second set of questions deal with measurement error (NRC Report, p. 96):

What is the effect of measurement error and misclassification on estimates of the association between air pollution and health?

ME1. How large are the various components of measurement error for each independent environmental variable (e.g., pollutants or weather)?. How is the measurement error of one variable related to the measurement errors of other variables in the same model?
ME2. What are the statistical distribution and types (e.g., Berkson or classical) of measurement error?
ME3. What are the effects of measurement error on the estimated associations between particulate matter (or its size fractions and biologically important chemical constituents) and health?
ME4. Does the presence of differential measurement errors in other variables in the model influence the estimate of association between a specific environmental agent and health?
ME5. Is there any error or misclassification likely to be present in the outcome variable? Is that error likely to have any effects on the outcome of the statistical models used for the analysis?
ME6. Can methods of adjusting for the effects of exposure measurement error be used to mitigate the effect of exposure measurement error on the risk estimates?
ME7. Can spatial interpolation methods provide more accurate estimates of individual exposures to particulate air pollution?
ME8. How would the use of measures of personal exposure improve the estimates of the association between particulate matter and health?

3. CONCLUSIONS AND RECOMMENDATIONS OF THE WORKING GROUPS

Six working groups were formed to cover the 19 NRC questions itemized in the preceding section, but also to provide the opportunity for discussion beyond these issues.

3.1 Time Series Analysis (SA1, 2, 3, 5)

The issues framing this discussion were: methods for removing long term trends, filtering of time series, dealing with autocorrelation, and separating unique effects of particulate matter from co-pollutants.

The first two statistical questions from the NRC Report are related questions about effects at different time scales: influence of long term trends (SA1) and filtering (SA2). As long term time trends and seasonal variations in concentrations of particulate matter in ambient air are confounded with secular trends and seasonal variation in mortality and morbidity, it is important to remove these sources of variation from the analysis. The method used to remove them is unlikely to have a large influence on the results. Available methods include: classical time series high-pass filters; smoothing variables against time and taking the residuals; including a smooth function of time in a regression model; and, transforming the regression model into the frequency domain where different time scales can be explicitly selected or removed. Models that filter only the predictor variables or that include a smooth function of time in the model are more transparently interpretable than those that filter the response. This is particularly true when the response is a count of rare events.

Seasonality is likely to often be an important issue. While the method used to remove longer term variation is unlikely to be important, the choice of which time scales to include and which to exclude may influence the results. Removing too little information exposes the analysis to confounding by season, removing too much reduces the power of the analysis and may exclude important health effects. However, it appears empirically that the results are often insensitive to the choice of time scale. More thorough, explicit modeling of seasonal effects might reduce difficulties associated with medium term information. Much of the seasonal variation in mortality and morbidity might be explained from meteorological and climatic information, patterns of infectious disease, and other better understood factors. This would result in models that might give very similar results but would rely less on purely statistical adjustments and more on explicit scientific knowledge. Another important aspect of seasonal modeling would be to examine differences in the associations between particulate matter and health in different seasons. Both the exposure to outdoor air and the composition of particulate air pollution change with the seasons. Stratifying the analysis by season is one possible approach but it is relatively inefficient.

The NRC Report asks about autocorrelation (SA3). A second important issue is overdispersion. After removing seasonal variations and long term trends and adjusting for other covariates such as temperature, autocorrelation is typically very weak. In analyses of mortality data even the residual overdispersion is often small, though in some morbidity studies it is substantial. There are a number of statistical techniques for adjusting standard error estimates to account adequately for overdispersion and autocorrelation of this magnitude. A remaining statistical question is the extent to which residual autocorrelation and overdispersion are an acceptable feature of time series models rather than an indication of inadequate modeling. Measuring and including more covariates will typically reduce autocorrelation, but it is neither practical nor desirable to include every possible covariate, and attention should be focused on variables that are important potential confounders.

SA5 is concerned with separating particulate matter (health) effects from those of other pollutants. In the context of a single time-series analysis it is extremely difficult to make reliable inferences about individual pollutants. Complicating factors in distinguishing between individual pollutant effects include high correlation between pollutants and the presence of substantial, possibly correlated measurement error. Still, useful information about classes of pollutants may still be gained. One approach is to use techniques such as factor analysis to group pollutants so that those in different classes are relatively uncorrelated. This allows some inferences to be made about those classes of pollutant.

Another method is to separate the pollutants by source instead of by chemistry. The correlation between different sources may be substantially lower, allowing useful multiple source models to be fitted. Adjustment for the meteorological and other factors that cause some of the correlations may also allow some inference about differing health effects of different pollutants.

3.2 Assessment of Current Epidemiologic Studies (SA5, 6, 7, 11)

The issues framing this discussion were: Are current epidemiological studies adequate to identify timing characteristics? Is residual confounding a concern? Are multiple statistical tests a concern? Do studies adequately separate unique effects of particulate matter? The working group address three broad topics.

3.2.1 Timing Issues

A rather broad-ranging discussion was held concerning the possible aspects of timing that are of concern, including (but not limited to): onset and duration of effect in relation to acute exposure; onset and duration of exposure required to produce chronic health effects; and, extent and onset of life shortening in mortality studies. The group concluded:

* published analyses of time series are not yet adequate for quantifying the distribution of life shortening on a scale of years
* published analyses of prospective cohort studies are not yet adequate for identifying the timing and duration of premature mortality from long term exposure to particulate matter

Several approaches to use in reanalyses of the prospective cohort studies were recommended. These may allow evaluation of alternative time-dependent exposure metrics for long term studies:

* Harvard Six Cities Study: use every-second day fine and coarse particle data to construct long term indices for each subject
* AHSMOG (Abbey et al. 1991, 1995)and ACS (Pope et al. 1995)studies: use visibility or other surrogates collected on a daily basis to construct time-dependent exposure metrics roughly equivalent to fine particles

There may be more information in daily time series studies concerning longer term exposure effects than has yet been used. These may include use of longer term averages, if the series is not adjusted by use of smoothers, or Fourier regression models. Better and more appropriate statistical methods are needed to evaluate the effect of mortality displacement in daily time series studies. Published approaches, based on autocorrelation patterns or episode patterns, are believed to be inadequate. (Validated) models for the recruitment and depletion of pools of susceptible subjects would greatly help to understand the pattern of events in time series studies. These should include patterns of hospital admissions or symptoms in individual subjects, where feasible.

3.2.2 Multiple Pollutants

More appropriate statistical methods are needed to evaluate the health effects of airborne particulate matter present in multiple forms or with other air pollutants. Promising methods include:

* grouping or combining pollutants as defined by source profiles
* grouping or combining pollutants as defined by principal components or factor analysis

Also, particulate matter health effects need to be estimated under several conditions:

* particulate matter acting alone (no others)
* particulate matter acting in the presence of a variety of mixtures of other air pollutants
* particulate matter acting as a surrogate for a mixture of air pollutants

The working group discussed at length the following hypothesis: the demonstration of quantitatively similar particulate matter health effects in cities with different concentrations and mixtures of copollutants proves the existence of a particulate matter effect that is not substantially confounded with copollutants. Various participants expressed caution about comparing effects across cities with different baseline populations and disease incidence, where different mixtures of pollutants may produce different health effects or endpoints. A worthwhile approach may be to see how the estimated particulate matter effect varies with the mean level of other pollutants across cities or studies. Other hierarchical modeling approaches to this problem were discussed.

3.2.3 Multiple Testing

The working group expressed concern that published analyses of time series studies appear largely exploratory in nature, with numerous model specifications evaluated before achieving the published results. This is particularly troublesome with respect to finding the lag structure for the effect from air pollution or meteorological variables. Bayesian hierarchical model averaging was suggested as a possible solution.

3.3 Exposure-Response Models (SA4, 5, 8, 9)

The issues framing this discussion were: determination of critical timing of exposure, separating unique effects of particulate matter from co-pollutants, types of exposure-response models, and incorporation of key covariates into risk models.

Observed exposure response relationships for fine particles are probably due to combinations of pollutants and not solely to particulates. Discussion focused on how to understand which combination of air pollution components is responsible for observed exposure-response relationships.

Results from epidemiologic studies such as the Harvard Six Cities Study and the University of Southern California's Children's Health Study (CHS) (Peters et al. 1999) suggest that variation in exposure derives from community of residence, indicating that between community or "ecologic" comparisons may provide the most information about exposure-response relationships. Between-community comparison are highly susceptible to confounding, suggesting that careful measurement and assessment of potential confounding variables may become increasingly important.

At present, it is not clear which combinations of components of air pollution to study. Toxicology studies in laboratory animals have not demonstrated effects at low exposure levels similar to those observed in human populations. These studies of air pollution components have usually considered young healthy animals and have thus investigated effects other than those likely to be the major sources of mortality and morbidity observed in humans. Future toxicology studies need to focus on the likely types of effects occurring in people, and to use laboratory animals whose health has been compromised before exposure.

Results from other disciplines also do not provide any specific hypotheses to test in future studies. To effectively plan, carry out, and evaluate new toxicology, epidemiology, air pollution, and exposure studies of people, will require interaction and coordination among four scientific disciplines: toxicology, epidemiology, statistics, and atmospheric sciences. Interaction and coordination will allow leads from one of these fields to be rapidly incorporated in the others, resulting in studies that are most likely to increase our understanding. Scientists in each discipline can best accomplish this by pursuing parallel paths while staying in close contact with other fields.

Research experience indicates that it is sometimes possible to separate risks by looking at the air pollution *sources*. This approach identifies a particular combination of air pollutants emitted by a source. Air quality at a particular time and location can then be modeled as a combination of the sources attenuated by distance from the sources and meteorological conditions. Measurements of air pollutants at sampler locations provide a validation of these models. Spatial-temporal patterns of mortality and morbidity can then be related to the sources. H. Ozkaynak has examined daily time-series data from Toronto using principal components analysis. Included were data on multiple pollution fields, and it was possible to separate out risks by sources, such as ozone, motor vehicles, and power plants. T. Larson related use of principal components analysis in recent studies of Spokane, WA to associate health risks with sources of air pollution. There are limitations to this type of approach: difficulty in plausible scientific understanding or interpretation of results, inaccuracies in emissions data, and difficulty in interpreting factor scores as sources. There is a concern about inclusion of meteorological variables in the model, since they are co-linear with many of other variables, e.g., concentration of a pollutant depends upon wind speed. Several members of the group felt that it may only be possible to relate health effects to sources of air pollution, or a combination of sources. Geographic regions of the country with unique combinations of sources of air pollutants are important to compare to other regions for differences in health effects. If there are different patterns in health effects between geographic regions, it may help to understand which sources is the cause.

In addition to looking at sources, components of air pollution, other than particles, should be examined in as careful manner as has been fine particles. Particles in air pollution were once dismissed as a nuisance dust. Many of the present analyses of particulate matter may not have looked carefully enough at some of the other components of air pollution, assuming that they did not contribute to the observed effects. It is also important to consider other cofactors, such as diet, exercise, and disease.

Concern was expressed about examination of nonlinear effects of air pollution. This may appear in several factors, such as, short term loading of the lung and lag structure in the models. Short term loading of the lung may occur because the usual clearance mechanisms for particles deposited on surfaces of the lung are slowed down or stopped. This would lead to a larger mass of deposited particles in the lung at one time than would normally be present which may alter other normal functions of the lung. Lag structure in air pollution models is complicated because each component of air pollution may have a different lag structure, for example ozone and particulate matter. It may not be possible to sort out all of the possible combinations of lags. Distributed lag models may be useful in this regard. In addition, both exposure and exposure-rate may be important in all of these questions.

Several major studies are under way that may help to understand exposure-response relationships, and questions raised here. A study funded by the Health Effects Institute will analyze 100 urban areas using the same methodology as the Samet-Zeger reanalysis of the Harvard Six-City Study (see: Dominici et al. 2000). A Canadian study will be able to link health records with death certificates, which is not possible in the United States. This will help to understand the role of various diseases.

3.4 Study Design Towards Estimation of Long and Short Term Effects of Exposure (SA10, ME8)

The issues framing this discussion were: How can the effects of long and short term exposures to particulate matter on mortality, including reduction in life expectancy, best be estimated? How would the use of measures of personal exposure improve the estimates of the association between particulate matter and health?

3.4.1 Estimation of Long and Short Term Effects

Estimation of the health effects due to particulate matter pollution requires that studies be designed that can capture the major sources of exposure variation. The study design requirements will depend upon whether the goal is to estimate long or short term exposure effects. Variation in exposure to ambient source particles will depend upon geographic location, weather patterns, behavior patterns, and penetration rates.

For long term exposure effects, it is most likely that sufficient exposure variation will only be realized in between-community studies. These studies, which will be observational, will have to be analyzed with considerable care since there are potential community-level confounding effects in such analyses.

For short term exposure effects, ecologic time series studies have seen extensive application for estimation of particulate matter health effects. While researchers have pointed to considerable consistency across studies in the estimated exposure effect, there has been limited work to determine the impact of model selection, residual confounding, or harvesting on the results. Other study designs, such as the case-crossover study, may contribute to estimating short term particulate effects specifically because they will be differentially impacted by potential biases.

3.4.2 Measures of Exposure

USEPA does not regulate personal exposure to particles per se but rather regulates the concentration of particles in ambient air. Thus the focus on personal exposure should be in the context of understanding personal exposure to pollutants from ambient sources. This will inherently limit the types of research questions of interest when focusing on personal exposure data.

Research experience suggests that personal exposures to air pollutants from indoor versus ambient sources typically are not correlated. Exceptions can occur when the indoor sources impact the ambient levels, as with wood burning. The association between ambient measurements at monitoring sites and personal exposure measurements (which are composed of both indoor and ambient source pollutants) varies considerably and depends on factors such as particulate matter penetration rates, particle deposition rates (which may vary with particle size), air exchange rates (which may vary seasonally), and individual behavior (also seasonally varying).

The Children's Health Study suggests that the largest source of variation in particulate matter exposure is due to a person's community of residence rather than a their location within the community or their individual behavior. Some caution should be used in generalizing this result however since this study was designed to maximize between-community variation and individual exposure measurements were obtained from a microenvironmental measurement and modeling approach rather than from actual personal exposure monitors. A personal exposure substudy in the CHS would help clarify this result.

There is a cost-driven limitation in the type of data that can be collected in a large-scale epidemiology study. Measuring personal exposure is very expensive. Personal exposure data will be of necessity limited to a few subjects with measurements over a limited time period. Optimum answers to many questions are unresolved such as how often measurements should be taken and on how many individuals, which metric is best (e.g. peak vs. average exposure), and which time interval is best for averaging. A significant benefit of personal exposure data for studies relevant to USEPA's mandate will be to collect them in conjunction with ambient site data. This allows modeling health effects due to personal exposure to ambient pollutants including that which has penetrated indoors. Priority should be given to studies that will provide sufficient data to allow modeling of personal and ambient exposure data for both long and short term health effect studies. Modeling issues will differ depending upon the term of the effect.

3.5 <u>Study Design and the Effects of Measurement Error in Health Effects Modeling</u> (ME1-ME6)

This discussion was framed by a plethora of issues relating to measurement error (broadly defined), including ME1-ME6 of the NRC Report. Two additional questions the working group regarded as paramount were discussed first: Are we using the right metrics? How do we meaningfully integrate exposure over history for chronic health effect assessment?

3.5.1 The two additional questions

There are two ways of viewing "measurement" error:

* the right variable is being measured, but measured badly
* the wrong variable is being measured (maybe very precisely)

Exposure is often measured in terms of averages, but other factors such as extreme values and variability may also be relevant. Not enough is understood as to the underlying biological mechanisms. Any metric for exposure will have an associated dose-response; it is preferable to choose the metric with the greatest sensitivity. The choice of appropriate metric might differ depending on whether individual or aggregate level data are being considered. More research is needed as to the impact of measurement error under different metrics.

The second question (on integration of exposure over history) was regarded as a special case of the first. There was discussion of how to weight different historical time periods. One concern is that different ages might be more or less susceptible to the effects of exposure. Residential mobility is a challenge in attempting to develop a measure of long term exposure.

It would appear that first exposure is modeled and second an attempt is made to link exposure to health effects. However, the properties of measurement error cannot be determined without first specifying the model being used and the data being used to fit the model. The structure of errors is well understood in the context of linear additive models, but in the nonlinear setting these issues are not as clearly resolved.

3.5.2 How bad is the problem?

As described in the preceding section, the distribution of measurement error depends on the (modeling) context. The true situation is a mixture of Berkson, classical and other types, often context dependent.

There is no general answer to ME2. However, although measurement error does lead to attenuation of the regression coefficient in a simple linear regression model, measurement error does not necessarily lead to attenuation in general.

3.5.3 Does it matter?

These issues are being addressed in empirical studies, such as that reported in a poster presentation of A. Marcus & C. Chapman at the workshop, "Estimating health effects of fine particles in epidemiology studies". However, more work needs to be done at a fundamental level on the characterization of the conditions under which differential measurement error leads to erroneous inference.

3.5.4 Can the problem be corrected?

Models define the relationships and the relevant attributes to be measured. Thus the measurement error problem transcends the classical paradigm implied by ME1-6. That fact must be recognized since it then appropriately expands the domain of enquiry to include such fundamental questions as: "Are we measuring the right attribute?" in addition to the more familiar question: "Are we measuring it precisely?"

3.6 Space-Time Modeling and Estimation Methods for More Accurate Estimates of Individual Exposures to Particulate Air Pollution (ME7)

This working group was to address just one question (ME7) from the NRC Report: Can spatial interpolation methods provide more accurate estimates of individual exposures to particulate air pollution? Spatial interpolation was regarded as an area with considerable potential in investigating particulate matter pollution, somewhat beyond its treatment in the NRC Report. Group discussion was organized under three related questions.

3.6.1 What are the potential roles and issues in space-time modeling of ambient particulate matter?

The primary role of space-time modeling is to characterize the particulate matter space-time field over a given geographic area. With future speciated particulate matter data this is likely to involve trying to characterize a multivariate spatio-temporal field. The group noted that fields resolved or estimated at an hourly level are probably most important for some considerations of health effects due to strong diurnal patterns and the use of hourly characterizations in exposure models such as pNEM (McCurdy 1995). It was also noted that ambient concentrations are, and are likely to remain the focus of air quality regulations even in view of the uncertainties in the relationships between ambient concentrations and human exposures.

A critical issue that has not been much addressed in the analyses of spatio-temporal fields of particulate matter and related atmospheric pollutant fields is the number and siting of the ambient monitors. The specific criteria for siting individual monitors should be incorporated in modeling the spatio-temporal field. Sites are never located according to any regular or probability based sampling design. In particular, some sites are supposedly located where experts believe the pollutant concentrations will be greatest (although in practice this appears to vary greatly by state and region.). To date, no spatio-temporal modeling or analyses have tried to take account of this type of prior information. In addition, the density of monitors must be taken into account in assessing how useful a spatial interpolation calculation may be. The adequacy of a given monitoring network will depend on the smoothness of the particulate matter space-time field, and this smoothness will depend on the temporal and temporal scales for which data are being considered. Experience, reported by J. Zidek, has shown that coping with the spatio-temporal correlation structure of hourly PM_{10} data can be challenging.

Preliminary evidence suggests that $PM_{2.5}$ fields are more spatially homogeneous than PM_{10} fields and certain other pollutant fields, such as ozone. Presumably, this will make the spatial interpolation easier. However, we cannot be as certain about the degree of spatial smoothness or variation in the speciated components of a $PM_{2.5}$ field, and, in any event, variation in topology and sources will affect variation in the $PM_{2.5}$ field.

3.6.2 What are the research needs relative to space-time modeling of particulate matter?

Space-time models will be needed for investigating the speciated components of particulate matter, both separately and in combination, especially if the components with the most serious health effects are identified. Again, spatial homogeneity should not be assumed a priori. Because speciated monitoring data will not be available for some time, to address variation in the components of particulate matter it may be useful to combine empirical size fraction measurement data with deterministic models that explicitly address the physical and chemical composition of ambient particulate matter.

A second need is the identification and development of synergistic methods bringing together deterministic models such as Models-3, statistical space-time models, and mass-balance models. Three areas were noted where deterministic and statistical models must be brought together:

* use of empirical modeling data and space-time statistical models in the validation of deterministic model predictions
* use of deterministic model predictions to evaluate the adequacy of air quality monitoring networks
* formal combination of deterministic model predictions and space-time statistical models for improved spatio-temporal prediction

It will be important to investigate several issues of scaling. Issues of spatial scale are critical when attempting to attach measures of precision to spatial estimates. Methods need to be considered with regard for the relationship or consistency between aggregations of predictions and predictions at an aggregated level. The working group recommended priority investigation of space-time scales and levels of aggregation and prediction most relevant for health effects.

3.6.3 What are the existing opportunities and future data needs for space-time modeling of particulate matter?

A shortage of data at relevant scales was noted. In particular, greater densities of monitoring sites will be necessary in order to assess network designs and to provide an adequate basis for the development and assessment of both deterministic and statistical models (by, for example, cross-validation). Upcoming networks of speciation monitors were discussed, noting that an expected network of 25-30 monitoring sites in New Jersey will be one of the most valuable resources for model development. Also useful for model development in areas of the eastern U.S. should be the existing SO_4 monitoring network because SO_4 is a good surrogate for particulate matter in certain areas. Finally, the group noted that another important resource for investigating issues of spatial and temporal variation was the Riverside ("PTEAM") Study in California where PM_{10} and $PM_{2.5}$ monitors were located outside of about 160-170 individual homes.

4. OPPORTUNITIES FOR STATISTICAL RESEARCH

This section provides a sampling of statistical research opportunities relevant to the study of particulate matter air pollution. It is a composite of discussions and presentations from the Workshop, based on the observations of the author. Unlike the preceding section, this section is not based on consensus views.

Most of the 19 questions raised in Topic 10 of the NRC Report do not have simple, unambiguous answers. The issues raised regarding measurement error and its effects on statistical analysis and interpretation of epidemiological and other particulate matter studies are likely to be data dependent, e.g., measurement error in predictor variables may or may not attenuate estimated regression coefficients. However, the 19 questions articulate quite clearly questions that need to be considered in each and every instance by epidemiologic investigators and point out areas where focused statistical research would be beneficial.

Both bias, e.g., due to multiple testing, and measurement error are likely to be present in studies of particulate matter. Often, bias and measurement error are intertwined, e.g, bias introduced into estimated regression coefficients by measurement error in explanatory variables, and distinguishing between them may not be straightforward. An attractive approach to separating and evaluating these effects, presented at the Workshop by M. Clyde, is Bayesian model averaging. A fully Bayesian approach based on data from a single city can identify the most plausible models, given the data, and thereby shed light on multiplicities. Analysis across cities can help distinguish between universal and regional covariates and temporal lags. Dominici et al. (2000) report a hierarchical Bayesian approach to combining particulate matter epidemiology studies across 20 major U.S. cities.

It is unlikely that particulate matter alone, or particulate matter classified only according to size and measured only in terms of mass, uniquely causes health effects and increased mortality. Recent scientific discussion is beginning to focus on other characteristics and classifications of particulate matter, notably number of constituent particles and chemical composition (*particle speciation*). There is a vicious cycle in that sufficient data to identify the most promising hypotheses are not available, while at the same time efficient design and deployment of data collection programs demands a manageable number of well-defined hypotheses. There are easily 100 potential variables and covariates possible. This creates a compelling case for use of dimension reduction methods such as factor analysis and principal components analysis. The Workshop presentation of R. Henry involved application of singular value decomposition to particulate source apportionment. Such methods should be further explored towards detecting and understanding clear particulate pollution signals from myriad sources of noise and for generating scientific hypotheses, particularly with regard to multi-pollutant relationships and effects. In addition, statistical associations between meteorology, particulate size and concentration, and health effects should continue to be explored.

Spatio-temporal modeling can contribute across a range of particulate matter problems. The NRC Report focused on evaluating measurement error due to using single-site (or simple combinations of multi-site), central city ambient concentration measurements as a surrogate measure of exposure to ambient particles. Concentrations predicted from a spatial model would offer improvement (or validate standard approaches). Spatio-temporal models would provide dynamic ("real-time, real-space") solutions. Statistical methodology for spatio-temporal models is in its infancy, and more research in this area, using particulate matter data, is likely to be very beneficial.

Spatio-temporal models offer other opportunities. In conjunction with atmospheric pollution models, spatio-temporal models can be used to refine or validate estimation of model parameters and boundary values, and for comparison with model outputs. An important decision in a modeling context is whether to aggregate estimates or estimate aggregates. Statistical methods can be brought to bear to assess and reduce associated bias and error. Spatial models offer an appealing potential enhancement to modern source apportionment methods.

Regulatory standards are based on data sampling, measurement and estimation, each with associated biases and errors. Research into statistical properties of current and proposed standards, effects of uncertainty and variability on regulatory outcomes, and on statistical foundations for regulatory standards would be helpful. Empirical studies on effects of specific measurement errors on analytical outcomes, as recommended by the NRC Report, is needed. Rigorous examination from a time series perspective of the use of temporal lags in epidemiological studies and increasing awareness on the issue of multiple testing may help to reduce the publication of spurious or artifactual findings.

5. CONCLUDING COMMENTS

The NRCSE/USEPA Workshop on Particulate Methodology raised meaningful issues regarding the role of statistical science in the study of particulate matter air pollution. Leaders from statistical and environment science shared their expertise and concerns and appeared to benefit from the interaction. Based on the Workshop, NRCSE launched a national, interdisciplinary research program on priority issues for particulate matter research, including spatio-temporal modeling of particulate matter, Bayesian model averaging, time series analysis, sensitivity analysis of particulate matter-mortality associations, and statistical aspects of particulate matter regulatory standards. Papers in this issue reflect a portion of that research. It is expected that other results from this research will be published in time for inclusion the 2002 Particulate Matter NAAQS review.

ACKNOWLEDGMENTS

The Particulate Matter Workshop benefitted enormously from the expertise and energy of its 50 participants. Special efforts were made by several individuals, some doing double or triple duty, including the following. The Planning Committee consisted of Gerald van Belle (Chair), Paul D. Sampson (Co-Chair), Merlise Clyde, Lawrence Cox, Peter Guttorp, Tim Larson, Thomas Lumley, Lianne Sheppard, Mary Lou Thompson, and Jim Zidek. Invaluable contributions were made by the speakers and discussants at the workshop. Others, joined by several just mentioned, served as workgroup chairs and reporters: Aaron Cohen, John Creason, Frank Gilliland, Bill Griffiths, Mark Kaiser, Drew Levy, Allan Marcus, Lianne Sheppard, and Rick White. Gerri Goedde provided logistical and on-site support. Thanks also go to William E. Wilson for illuminating discussions on particulate matter monitoring and regulation and to an anonymous referee for valuable suggestions.

Principal funding for the Particulate Matter Workshop was provided by the Office of Research and Development, U.S. Environmental Protection Agency, to the National Research Center for Statistics and the Environment, University of Washington, under cooperative agreement CR 825173-01-0. This article has been subjected to Agency review and approval for publication. The views contained herein are those of the author and are not intended to represent Agency policies or practices. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use.

REFERENCES

Abbey, D.E., P.K. Mills, F. Petersen, and W.L. Beeson (1991). Long-term ambient concentrations of total suspended particulates and oxidants as related to incidence of chronic disease in California Seventh-Day Adventists. *Environmental Health Perspectives* **94**, 43-50.

Abbey, D.E., B.L. Hwang, R.J. Burchette, T. Vancuren and P.K. Mills (1995). Estimated long-term ambient concentrations of PM_{10} and development of respiratory symptoms in a nonsmoking population. *Archives of Environmental Health* **50**, 139-152.

Clinton, W.J. (1997). Memorandum for the Administrator of the Environmental Protection Agency. Washington, DC: The White House, dated July 16, 1997. (Available: http://www.epa.gov/ttn/oarpg/naaqsfin/impfac.html).

Dominici, F., J.M. Samet and S.L. Zeger (2000). Combining evidence on air pollution and daily mortality from the 20 largest U.S. cities: a hierarchical modeling strategy. Journal of the Royal Statistical Society, Series A **163**, Part 2, to appear.

Dockery, D.W., C.A. Pope III, X. Xu, J.D. Spengler, J.H. Ware, M.E. Fay, B.G. Ferris, Jr. and F.E. Speizer (1993). An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine* **329**, 1753-1759.

Lumley, T. and Levy, D (2000). Bias in the case-crossover design: implications for studies of air pollution. *Environmetrics*, to appear.McCurdy, T.R. (1995). Estimating human exposure to selected motor vehicle pollutants using the NEM series of
models: Lessons to be learned. *Journal of Exposure Analysis and Environmental Epidemiology* **5**(4), 533-550.

National Academy of Sciences (1998). **Research Priorities for Airborne Particulate Matter: I: Immediate Priorities and a Long-Range Research Portfolio**. Committee on Research Priorities for Airborne Particulate Matter, National Research Council. Washington, DC: National Academy Press.

Navidi, W., D. Thomas, B. Langholz and D. Stram (1999). **Statistical Methods for Epidemiological Studies of the Health Effects of Air Pollution**. Cambridge, MA: Health Effects Institute. HEI Research Report 86.

Peters, J.M., E. Avol, W. Navidi and S.J. London (1999). A study of twelve Southern California communities with differing levels and types of air pollution. I. Prevalance of respiratory morbidity. *American Journal of Respiratory Critical Care Medicine* **159**(3), 760-767.

Pope, C.A. III, M.J. Thun, M.M. Namboodiri, D.W. Dockery, J.S. Evans, F.E. Speizer and C.W. Heath, Jr. (1995). Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *American Journal of Respiratory Critical Care Medicine* **151**, 669-674.

U.S. Court of Appeals (1999). Decision: American Trucking Associations, Inc. et al., Petitioners v. United States Environmental Protection Agency, Respondent. Washington, DC: U.S. Court of Appeals for the District of Columbia Circuit Court, Case No. 97-1440. Argued: December 17, 1998. Decided: May 14, 1999.

U.S. Environmental Protection Agency (1996). **Air Quality Criteria for Particulate Matter**. Research Triangle Park, NC: National Center for Environmental Assessment-RTP Office. EPA Report Numbers EPA/600/P-95/001 aF-cF.3v. Available from: NTIS, Springfield, VA: PB96-168224.

U.S. Federal Register (1997). 40 CFR Part 50: National Ambient Air Quality Standards for Particulate Matter; Final Rule. Washington, DC: *U.S. Federal Register, Vol. 62, No. 138, Part II: Environmental Protection Agency* (July 18, 1997).

Submitted: June 21, 1999
Revised: January 10, 2000