RESEARCH REPORT

HEALTH EFFECTS INSTITUTE

Number 142
October 2009

PRESS VERSION

Air Pollution and Health: A European and North American Approach (APHENA)

Klea Katsouyanni and Jonathan M. Samet

Europe Investigators: H. Ross Anderson, Richard Atkinson, Alain Le Tertre, Sylvia Medina, Evangelia Samoli, and Giota Touloumi

Canada Investigators: Richard T. Burnett, Daniel Krewski, Timothy Ramsay

United States Investigators: Francesca Dominici, Roger D. Peng, Joel Schwartz, and Antonella Zanobetti

Includes a Commentary by the Institute's Health Review Committee

Air Pollution and Health: A European and North American Approach (APHENA)

Klea Katsouyanni, and Jonathan M. Samet

Europe Investigators: H. Ross Anderson, Richard Atkinson, Alain Le Tertre, Sylvia Medina, Evangelia Samoli, and Giota Touloumi

Canada Investigators: Richard T. Burnett, Daniel Krewski, Timothy Ramsay

United States Investigators: Francesca Dominici, Roger D. Peng, Joel Schwartz, and Antonella Zanobetti

with a Commentary by the HEI Health Review Committee

Research Report 142

Health Effects Institute

Boston, Massachusetts

Trusted Science • Cleaner Air • Better Health

Publishing history: The Web version of this document was posted at *www.healtheffects.org* in October 2009 and then finalized for print.

Citation for whole document:

Katsouyanni K, Samet J, Anderson HR, Atkinson R, Le Tertre A, Medina S, Samoli E, Touloumi G, Burnett RT, Krewski D, Ramsay T, Dominici F, Peng RD, Schwartz J, Zanobetti A. 2009. Air Pollution and Health: A European and North American Approach (APHENA). HEI Research Report 142. Health Effects Institute, Boston, MA.

When specifying a section of this report, cite it as a chapter of the whole document.

 Cameographics, Belfast, Me., Compositor. Printed by Recycled Paper Printing, Boston, Mass. Library of Congress Catalog Number for the HEI Report Series: WA 754 R432.

Cover paper: made with at least 55% recycled content, of which at least 30% is post-consumer waste; free of acid and elemental chlorine.Text paper: made with at l00% post-consumer waste recycled content; acid free; no chlorine used in processing. The book is printed with soy-based inks and is of permanent archival quality.

CONTENTS

ABOUT HEI

The Health Effects Institute is a nonprofit corporation chartered in 1980 as an independent research organization to provide high-quality, impartial, and relevant science on the effects of air pollution on health. To accomplish its mission, the institute

- Identifies the highest-priority areas for health effects research;
- Competitively funds and oversees research projects;
- Provides intensive independent review of HEI-supported studies and related research;
- Integrates HEI's research results with those of other institutions into broader evaluations; and
- Communicates the results of HEI research and analyses to public and private decision makers.

HEI receives half of its core funds from the U.S. Environmental Protection Agency and half from the worldwide motor vehicle industry. Frequently, other public and private organizations in the United States and around the world also support major projects or certain research programs. HEI has funded more than 280 research projects in North America, Europe, Asia, and Latin America, the results of which have informed decisions regarding carbon monoxide, air toxics, nitrogen oxides, diesel exhaust, ozone, particulate matter, and other pollutants. These results have appeared in the peer-reviewed literature and in more than 200 comprehensive reports published by HEI.

HEI's independent Board of Directors consists of leaders in science and policy who are committed to fostering the public–private partnership that is central to the organization. The Health Research Committee solicits input from HEI sponsors and other stakeholders and works with scientific staff to develop a Five-Year Strategic Plan, select research projects for funding, and oversee their conduct. The Health Review Committee, which has no role in selecting or overseeing studies, works with staff to evaluate and interpret the results of funded studies and related research.

All project results and accompanying comments by the Health Review Committee are widely disseminated through HEI's Web site (*www.healtheffects.org*), printed reports, newsletters, and other publications, annual conferences, and presentations to legislative bodies and public agencies.

ABOUT THIS REPORT

Research Report 142, *Air Pollution and Health: A European and North American Approach (APHENA),* presents a research project funded by the Health Effects Institute and conducted by Dr. Klea Katsouyanni of the Department of Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece, and Dr. Jonathan M. Samet of the Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, and their colleagues. This report contains three main sections.

The HEI Statement, prepared by staff at HEI, is a brief, nontechnical summary of the study and its findings; it also briefly describes the Health Review Committee's comments on the study.

The Investigators' Report, prepared by Drs. Katsouyanni and Samet and their colleagues, describes the scientific background, aims, methods, results, and conclusions of the study.

The Commentary is prepared by members of the Health Review Committee with the assistance of HEI staff; it places the study in a broader scientific context, points out its strengths and limitations, and discusses remaining uncertainties and implications of the study's findings for public health and future research.

This report has gone through HEI's rigorous review process. When an HEI-funded study is completed, the investigators submit a draft final report presenting the background and results of the study. This draft report is first examined by outside technical reviewers and a biostatistician. The report and the reviewers' comments are then evaluated by members of the Health Review Committee, an independent panel of distinguished scientists who have no involvement in selecting or overseeing HEI studies. During the review process, the investigators have an opportunity to exchange comments with the Review Committee and, as necessary, to revise their report. The Commentary reflects the information provided in the final version of the report.

HEI STATEMENT

Synopsis of Research Report 142

Air Pollution and Health: A European and North American Approach

BACKGROUND

Over the past decade, scientists seeking to understand the role that air pollution might play in population health effects have relied heavily on epidemiologic studies known as time-series studies. Time-series studies use information on daily concentrations of air pollutants and daily measures of health impact (numbers of deaths or of admissions to hospitals), initially at the level of a single city. However, the wide range of methods used to assemble and analyze data from individual cities has made their findings difficult to interpret and has led to progressive efforts to combine information across multiple cities and ultimately, across geographic regions. The goal of these larger analyses has been to develop more reliable estimates of the potential acute effects of air pollution on human health, to provide a common basis for comparison of risks across geographic areas, and to increase the ability to discriminate between health effects that may truly be related to air pollution and those that may be attributable to other factors. Ultimately, the goal is to improve the scientific basis for decisions about whether and how to regulate air pollution.

To explore these issues, HEI sponsored a unique collaboration among investigators from Europe, the United States, and Canada, led by co-principal investigators Klea Katsouyanni, University of Athens and Jonathan Samet, then at the Johns Hopkins Bloomberg School of Public Health, and overseen by an external Scientific Oversight Group. The resulting project was called Air Pollution and Health: A European and North American Approach or APHENA. With access to data from three geographic areas, APHENA offered a much larger and diverse data set with which to address methodological as well as scientific issues about the relationships between PM_{10}, ozone, and mortality and morbidity that were the subject of lively debates at the time the project was first conceived in 1999.

The investigators undertook a rigorous examination of time-series methods used to model the relationship between daily PM_{10} (particulate matter with an aerodynamic less than 10 micrometers) and ozone concentrations and daily mortality and hospital admissions. They sought to develop a standardized approach to the analysis of time series data at the city and regional level, to assess the consistency between relative rates of mortality and hospital admissions across Europe and North America when estimated using a common analytic protocol, and to explore possible explanations for any remaining variation in the results that analytic differences could not explain.

APPROACH

The APHENA project was designed to take advantage of the largest databases available at the time. These had been developed by the three groups of investigators for earlier studies: 1) the *Air Pollution and Health: A European Approach* Phase 2 (APHEA2) study involving 32 cities; 2) the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), conducted in the 90 largest U.S. cities; and 3) multicity research on the health effects of air pollution in 12 Canadian cities.

Each database included air pollution monitoring data for particulate matter and ozone, health outcome data in the form of daily mortality for all ages, for persons younger than 75 years, and for persons 75 years or older (from all nonaccidental causes [all-cause]), cardiovascular disease, or respiratory disease) and daily hospital admissions for persons 65 years or older (for cardiovascular and respiratory disease). Other database variables used for APHENA

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by Dr. Klea Katsouyanni at University of Athens Medical School, Athens, Greece and by Dr. Jonathan M Samet at John Hopkins Bloomberg School of Public Health, Baltimore, MD and their colleagues. Research Report 142 contains both the detailed Investigators' Report and a Commentary on the study prepared by the Institute's Health Review Committee.

included weather data and a number of socioeconomic and other variables known or suspected to influence mortality or hospital admissions. (These latter variables were considered potential *effect modifiers,* factors that can modify the main effect of air pollution on health and that may differ between study areas).

The decision to rely on the preexisting time-series databases had the advantage of lower costs and a more rapid start to the methodological work of the project. Also, the original published analyses of these datasets could serve as a baseline against which to explore the impact of the methodological choices made in APHENA. However, it had the disadvantage that inherent differences between the way air pollution or outcome data were collected by government agencies in different countries remained as a potential source of uncertainty. Furthermore, restrictions placed by various government agencies on the use of the data further limited evaluation of the datasets because the APHENA investigators could not create a central repository for the time-series data. This limitation precluded a full exchange of data sets, so quality assurance evaluations were based on an exchange of data from six cities.

The APHENA methods exploration followed the two-stage process typical of multicity studies, beginning with analysis of the data for individual cities. The APHENA investigators undertook careful analysis of: 1) the class of models to be used for analysis, including the data smoothing methods to control for potential confounding by seasonal or other temporal trends in the data; 2) the amount of control for these trends (that is, the amount of smoothing represented by the number of degrees of freedom associated with the smoothing method chosen); and 3) the suite of other variables to be included in the model.

On the basis of extensive sensitivity analyses and other evaluations, the investigators decided that they could not justify either fitting a single model for all cities or fitting separate models for each city — the most common approach at the time. They agreed instead on a common protocol involving a range of models and assumptions to be applied to all city-specific analyses of daily deaths and hospital admissions.

For the second stage of analysis, the APHENA investigators conducted a systematic comparison of the statistical approaches previously used by the NMMAPS and APHEA groups to pool the estimates across cities and to explore variations in the effect estimates for their regions. Although the HEI Scientific Oversight Group had encouraged selection of a single approach for the second stage, the investigators felt that their comparative analysis did not suggest a clear preference for one method over another, so each investigative group continued to apply its preferred method.

Having explored the sensitivity of the health effects estimates to analytic choices, the investigators then focused on trying to understand other factors that might explain the variation in the effect estimates between cities and regions. They evaluated potential modification of the effects of PM_{10} on all-cause mortality by variables common to the data sets for those cities that had daily PM data — 22 cities in Europe and 15 in the United States. They investigated potential modification of the ozone effect on mortality in all three regions, including Canada. Although the investigators presented pooled estimates of PM_{10} mortality effects for Europe and the United States and of ozone mortality effects across all three regions, too many differences existed among the databases to explore effect modification using the combined data sets.

RESULTS

The problem the APHENA investigators addressed in their substantial methodologic work was essentially one of model selection. They found that effect estimates remained fairly stable across a broad spectrum of model assumptions. In particular, they found that the amount of smoothing (numbers of degrees of freedom selected) for control of seasonal and temporal confounding was more important than the method of smoothing (for example, the use of natural or penalized splines in the models). This finding is important because it suggests that a relatively simple choice of method may usually be appropriate, but that investigators should explore several choices for the amount of smoothing. Coupled with the limitations arising from using large administrative databases that have been constructed for other purposes, a reasonable guideline for future investigators is to choose the simplest model that seems to capture the main variability in the data, and to explore in detail the sensitivity of the most scientifically relevant conclusions.

Using the standardized protocol developed from this work, the APHENA investigators largely confirmed the basic findings from previous independent analyses of the three data sets for both PM_{10} and ozone, including the much higher effect estimates in Canada relative to the other two regions. The investigators reported small, but positive and statistically significant associations between a 10 $\mu g/m^3$ increase in PM_{10} and all-cause and cardiovascular mortality in Canada, Europe, and the United States. Effects of PM_{10} on respiratory mortality were less consistent across region and model. Ozone showed a smaller, but generally positive association with all-cause and cardiovascular mortality in each of the three regions, but not with respiratory mortality. Quality control measures implemented during the project showed that the three teams produced essentially the same results when asked to analyze the same subset of data sets from the project.

Estimates of the effect of PM_{10} on hospital admissions for cardiovascular and respiratory disease varied across the three centers. Daily increases in PM_{10} were not associated with significantly increased risks of hospitalization for cardiovascular disease in Canada, but were in Europe and in the United States. The risks of admission for respiratory disease in the three regions were more uncertain, variable, and model dependent.

In contrast to estimates of the effect of ozone on cardiovascular mortality, the effects on cardiovascular hospital admissions were closer to zero and not statistically significant. The converse was observed with the respiratory effects; the effects of ozone on respiratory disease admissions were largely positive while respiratory mortality effects were close to zero. Effects on respiratory admissions were both higher and more uncertain in Canada than in Europe or the United States.

The exploration of effect modification was ultimately limited by the number of variables common to the data sets and the smaller number of cities (only those with daily pollutant data) included in the analyses. For PM_{10}, the most consistent evidence of effect modification was found for age and unemployment in Europe and the United States; a higher percentage of older people and a higher unemployment rate were each associated with a greater effect of PM_{10} on all-cause mortality in both regions. The investigators found no consistent patterns of effect modification for O_3 across the three regions.

DISCUSSION

APHENA was an ambitious project undertaken by a highly qualified team of investigators from Europe, the United States, and Canada. It made substantial contributions to how multicity time-series studies should be designed and conducted. In particular, the APHENA investigators' careful development and application of a common analytic approach to city-level analyses was an important advance over other meta-analytical approaches, in common use prior to this study, which relied on published city-specific results. Furthermore, it demonstrated the importance of a well-reasoned strategy of sensitivity analyses, both to support model development and to provide a transparent presentation of the role of model choices on the estimates of health effects.

The investigators successfully demonstrated that methodological differences between the centers' approaches to the regional-level analyses were not the reason for variability observed in previous study findings of the three geographic areas. Although the decision to allow each center to continue to use separate methods was not the original plan, the HEI Review Committee found it reassuring that the APHENA investigators' extensive comparison of the APHEA and NMMAPS approaches indicated little reason to choose between the two methods. However, the Committee found that the information provided in the report was not sufficient to support an independent assessment of that conclusion.

The findings of small, but significant effects of PM_{10} and ozone on daily death and, to varying degrees, on hospital admissions are important. They corroborate earlier findings on the health effects of daily air pollution and, coupled with the systematic analytic approaches and quality control measures used in the studies, demonstrate that the effect estimates can not be attributed solely to the vagaries of model choice.

The hope that the common analytic strategy might help reveal some of the other potential contributors to variations observed within and between the three regions was largely unfulfilled. Few new insights were possible given the limited number of potential effect modifiers that were common to the databases for the three regions, and also given the restriction of the analyses to the smaller number of cities with daily PM_{10} and ozone data within regions. Some of the more puzzling differences

between regions therefore remained largely unexplained — in particular the much higher effect estimates for PM_{10} and ozone in Canada relative to Europe and the United States.

The APHENA study demonstrated the substantial challenges that face efforts to standardize and integrate data from different countries. Overcoming government agency reluctance or other impediments to establishing centralized databases might help, but some challenges remain beyond the control of investigators. Basic underlying disparities in the existing databases with respect to air pollutant monitoring methods and frequency, mortality and hospitalization records, and sociodemographic data are very difficult to fix retrospectively.

The authors suggest that periodic pooling of data, as in APHENA, should be considered both to explore methodological questions and to assess the progress of air pollution controls in reducing health impacts. The HEI Review Committee believes this recommendation should be evaluated cautiously. Studies like APHENA that use a well-reasoned, common analytic strategy, may offer the best approach for comparing and combining data across regions or countries. However, APHENA illustrated how the limitations of using existing data sets can impact the ability to make clear comparisons and to explain the health effects of air pollution — sometimes just as much as the technical details of model selection. For these limitations to be overcome in any future collaboration across international boundaries, thought needs to be given to substantially greater coordination and harmonization of the air pollution monitoring, health outcome, and covariate data collected in different countries. The costs of undertaking such exercises would need to be weighed against the expected advances in our understanding of air pollution and health effects.

INVESTIGATORS' REPORT

Air Pollution and Health: A European and North American Approach (APHENA)

Principal Investigaors: Klea Katsouyanni and Jonathan M. Samet

Europe Investigators: H. Ross Anderson, Richard Atkinson, Alain Le Tertre, Sylvia Medina, Evangelia Samoli, and Giota Touloumi

Canada Investigators: Richard T. Burnett, Daniel Krewski, Timothy Ramsay

United States Investigators: Francesca Dominici, Roger D. Peng, Joel Schwartz, and Antonella Zanobetti

Principal Investigators: *Department of Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece (K.K.), and Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (J.S.);* ***Europe:*** *Department of Community Health Sciences, St. George's, University of London, United Kingdom (H.R.A., R.A.), Environmental Health Department, Institut de Veille Sanitaire, Paris, France (A.L., S.M.), and Department of Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece (E.S., G.T.);* ***Canada:*** *Department of Epidemiology and Community Health, University of Ottawa, Ottawa, Ontario (R.T.B., D.K., T.R.), and Health Canada, Ottawa, Ontario (R.T.B.);* ***United States:*** *Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (F.D., R.D.P.), and Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (J.S., A.Z.)*

ABSTRACT

INTRODUCTION

This report provides the methodology and findings from the project: *Air Pollution and Health: a European and North American Approach* (APHENA*). The principal purpose of the project was to provide an understanding of the degree of consistency among findings of multicity time-series studies on the effects of air pollution on mortality and hospitalization in several North American and European cities. The project included parallel and combined analyses of existing data. The investigators sought to understand how methodological differences might contribute to variation in effect estimates from different studies, to characterize the extent of heterogeneity in effect estimates, and to evaluate determinants of heterogeneity. The APHENA project was based on data collected by three groups of investigators for three earlier studies: (1) *Air Pollution and Health: A European Approach* (APHEA), which comprised two multicity projects in Europe. (Phase 1 [APHEA1] involving 15 cities, and Phase 2 [APHEA2] involving 32 cities); (2) the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), conducted in the 90 largest U.S. cities; and (3) multicity research on the health effects of air pollution in 12 Canadian cities.

METHODS

The project involved the initial development of analytic approaches for first-stage and second-stage analyses of the time-series data and the subsequent application of the resulting methods to the time-series data. With regard to the first-stage analysis, the various investigative groups had used conceptually similar approaches to the key issues of controlling for temporal confounding and temperature; however, specific methods differed. Consequently, the investigators needed to establish a standard protocol, but one that would be linked to prior approaches.

Based on exploratory analyses and simulation studies, a first-stage analysis protocol was developed that used generalized linear models (GLM) with either penalized splines (PS) or natural splines (NS) to adjust for seasonality, with 3, 8, or 12 degrees of freedom (*df*) per year and

This Investigators' Report is one part of Health Effects Institute Research Report 142, which also includes a Commentary by the Health Review Committee and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. Jonathan M. Samet, University of Southern California, 1441 Eastlake Ave., Room 4436, MC 9175, Los Angeles, CA 90089.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83234701 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

* A list of abbreviations and other terms appears at the end of the Investigators' Report.

also the number of degrees of freedom chosen by minimizing the partial autocorrelation function (PACF) of the model's residuals. For hospitalization data, the approach for model specification followed that used for mortality, accounting for seasonal patterns, but also, for weekend and vacation effects, and for epidemics of respiratory disease. The data were also analyzed to detect potential thresholds in the concentration–response relationships.

The second-stage analysis used pooling approaches and assessed potential effect modification by sociodemographic characteristics and indicators of the pollution mixture across study regions. Specific quality control exercises were also undertaken. Risks were estimated for two pollutants: particulate matter ≤ 10 μm in aerodynamic diameter (PM_{10}) and ozone (O_3).

RESULTS

The first-stage analysis yielded estimates that were relatively robust to the underlying smoothing approach and to the number of degrees of freedom. The first-stage APHENA results generally replicated the previous independent analyses performed by the three groups of investigators. PM_{10} effects on mortality risk estimates from the APHEA2 and NMMAPS databases were quite close, while estimates from the Canadian studies were substantially higher. For hospitalization, results were more variable without discernable patterns of variation among the three data sets. PM_{10} effect-modification patterns, explored only for cities with daily pollution data (i.e., 22 in Europe and 15 in the U.S.), were not entirely consistent across centers. Thus, the levels of pollutants modified the effects differently in Europe than in the United States. Climatic variables were important only in Europe. In both Europe and the United States, a higher proportion of older persons in the study population was associated with increased PM_{10} risk estimates, as was a higher rate of unemployment — the sole indicator of socioeconomic status uniformly available across the data sets.

APHENA study results on the effects of O_3 on mortality were less comprehensive than for PM_{10} because the studies from the three regions varied in whether they analyzed data for the full year or only for the summer months. The effects tended to be larger for summer in Europe and the United States. In the United States they were lower when controlled for PM_{10}. The estimated effect of O_3 varied by degrees of freedom and across the three geographic regions. The effects of O_3 on mortality were larger in Canada, and there was little consistent indication of effect modification in any location.

CONCLUSIONS

APHENA has shown that mortality findings obtained with the new standardized analysis were generally comparable to those obtained in the earlier studies, and that they were relatively robust to the data analysis method used. For PM_{10}, the effect–modification patterns observed were not entirely consistent between Europe and the United States. For O_3, there was no indication of strong effect modification in any of the three data sets.

INTRODUCTION

This report provides the methodology and findings from the APHENA project. The principal purpose of the project was to provide an understanding of the extent of coherence among findings of multicity time-series studies carried out in Europe and North America. APHENA was developed after discussions that began in the late 1990s. The project originated at a time when major findings of multicity analyses, including APHEA and NMMAPS, were being reported and used as part of the evidence base for setting ambient air quality standards and formulating guidelines for particulate matter (PM) and O_3 concentrations in ambient air. As discussions about the purpose of APHENA continued, primary scientific objectives were identified: (1) to characterize the extent of heterogeneity of the effect of PM on mortality and hospitalization rates, and (2) to explore factors contributing to the observed heterogeneity. These objectives were put into the Mission Statement prepared by the APHENA investigators: "The elements of the project have an overall goal of getting a deeper understanding of the way in which the different modeling approaches applied affect the effect estimates and of providing an understanding of the extent of heterogeneity in estimates of the effect of air pollution in time-series studies carried out in North America and Europe and of the determinants of heterogeneity."

APHENA contributes evidence relevant to one of the key uncertainties in current understanding of the health effects of PM: identification of the chemical and physical characteristics of particles associated with toxicity (National Research Council and Committee on Research Priorities for Airborne Particulate Matter 2004). Current regulatory standards rely on measures of airborne particle mass to represent exposures to people, because the physical and chemical characteristics that determine toxicity are not yet known. Evidence on PM characteristics that determine health risks could be used to focus control strategies on the sources most relevant to the protection of public health. The APHENA project was undertaken with

the expectation that heterogeneity in the health effects of PM would be explored across the broad range of atmospheres represented by the APHENA cities. The data for APHENA came from all of the major cities in the United States and Canada and from many of the major cities in Europe. These cities represent a wide range of sources and meteorological conditions. Relative homogeneity across this diverse set of locations would weigh against differential toxicity by source, while the finding of heterogeneity could support more focused hypotheses.

Any evaluation of heterogeneity across studies needs to address the potential consequences of using differing analytic strategies. In recent years, there has been extensive exploration of the sensitivity of time-series models to various aspects of model specification and fitting, motivated in part by the finding that the original default assumptions in the S-Plus software, widely used for time-series analysis at the time, were not appropriate for analysing air pollution time-series data (Dominici et al. 2002b). In response to this finding, HEI coordinated the reanalysis and publication of results from key time-series studies using the more stringent convergence criteria of the S-Plus software for generalized additive models (GAMs) (Health Effects Institute 2003). The analyses included within that report explored the sensitivity of time-series results to additional aspects of model specification, including the model used for smoothing (NS, PS, smoothing spline [SS], or case-crossover matching), and the number of degrees of freedom used for smoothing. Overall, the sensitivity analyses showed that decisions made in model specification could have subtle or even substantial consequences that could change the interpretation of the model's results. The experience gained during the HEI reanalysis provided a further impetus for carrying out the APHENA project.

The APHENA project used time-series databases that had been developed for previous studies. The project's goals were both methodological and applied. In the first stage, a standard analytic methodology was developed for conducting individual city analyses of the time-series data used in the investigators' earlier reports. The regression estimates from the first stage were then used for second-stage analyses directed at characterizing the extent and potential sources of heterogeneity in estimates of the effects of air pollution on health.

The overall scientific objectives of the present study, based on the experience and past approaches of the APHENA investigators, include:

- Develop a common approach for first-stage analyses of mortality and admissions time-series data, and assess sensitivity of findings to critical elements of the model (using simulations and real data);
- Perform a comparative evaluation of different methods developed within the projects to identify and combine concentration–response curves;
- Compare alternative methods for addressing mortality displacement (the hypothesis that deaths associated with exposure represent simply a temporal shift in when deaths occur, not in the total number of deaths), leading to the eventual application of one or more approaches to time-series data from Europe and North America;
- Develop a database on potential effect modifiers, consistent across the three regions, such that their influence on estimates of health effects could be explored;
- Perform parallel and combined analyses of data on air pollution and mortality, as well as on air pollution and morbidity, that address issues of heterogeneity described above.

METHODS

OVERVIEW

The APHENA project brought together three groups of investigators that had independently developed methods for analysis of time-series data. The project had its origins in discussions initiated in the late 1990s when the main results of the studies from the three research centers were reported and questions about the comparability of methods and findings were raised. The concept for APHENA was further elaborated at a meeting in London in 2000, and a grant was subsequently submitted to and approved by the HEI. Funding began in April, 2003.

The APHENA investigators met to discuss the overall approach and to develop study policies. Then they used their combined experience and expertise to develop a common protocol for first- and second-stage analyses. The overall project proceeded under the direction of Drs. Katsouyanni and Samet. Teams with members from each of the component groups were formed to develop the first-stage models, the second-stage modeling approach, and a list of potential modifying factors common to the core databases. The study investigators worked together and with working group members through conference calls, e-mail exchanges, and periodic meetings with the entire team. A single database containing all of the time-series data could not be created because the various government agencies providing the source data imposed restrictions on the use of the data. A limited exchange of several single-city time-series data sets was allowed for quality assurance purposes.

HEI appointed a Scientific Oversight Group to provide oversight throughout the project. Multiple joint meetings were held involving the investigators and this group. The Scientific Oversight Group offered extensive comments on the analytic protocol.

The development of the first-stage analytic approach included: (1) selection of the statistical model for the analysis; and (2) selection of the approach to control for temporal confounding factors and exploration of the sensitivity of findings to the degrees of freedom used for control of confounding. In the second stage, variables were identified that described potential determinants of heterogeneity common to all of the data sets.

DESCRIPTION OF DATABASES

In brief, the project used existing databases for the first-stage analyses, as planned when the study proposal was submitted, but following a protocol outlined in Appendix B. The databases had differing origins, and some aspects of the database development were not under the direct control of the APHENA investigators. The APHEA databases had been assembled by the collaborating investigators from the participating cities; the NMMAPS database was put together by the Johns Hopkins investigators from various national sources; and the Canadian databases were assembled by investigators working with Health Canada who had direct access to the necessary data. The starting point for the analytic databases was the existing time-series database. HEI commissioned a team that conducted a detailed audit procedure during a site visit. The audit team reviewed the construction of the databases, and verified the database construction (Appendix C).

NMMAPS

The development of the NMMAPS database is described in the HEI-funded Web site, internet-based Health and Air Pollution Surveillance System (iHAPSS) (*www.ihapss.jhsph.edu*) and also in the NMMAPS reports (Samet et al. 2000b,c). The Web site also provides descriptions of the data for each of the included cities. Of note, several errors in the NMMAPS data were found in a review conducted after publication of the first NMMAPS studies and were corrected. The APHENA project used the corrected NMMAPS data set for analysis.

The geographic location, identified by Federal information processing standards codes (FIPS code), provided the basis for merging the various data sets, as this identifier was common to the various databases. For NMMAPS, counties had been identified that corresponded to cities to the extent possible. The correspondence between city and county was evident for most cities. For others the investigators documented the decisions they made when assigning counties to cities.

Mortality Data Mortality data were obtained from the National Center for Health Statistics (*www.cdc.gov/nchs*) for years 1987 through 1996. The FIPS code again served as the basis for aggregation of the deaths. Nonresidents of the counties were excluded. Daily mortality counts were classified into three age categories (< 65 years; 65–74 years; ≥ 75 years). Accidental deaths (i.e., International Classification of Diseases (ICD)-9 ≥ 800) were excluded from the database. Details of the NMMAPS mortality file development, the cause of death groups and codes of the ICD 9th and 10th Revisions (World Health Organization [WHO] 1979, 1992) that were used, along with the three strata of ages to which all deaths were assigned, are available on the iHAPPS Web site (*www.ihapss.jhsph.edu/*).

Hospital Admissions Data Hospital admissions data were extracted from the Health Care Financing Administration (Medicare) billing records, which were obtained for years 1985 through 1994, by state and county of residence of the patient. The Medicare system provides hospital coverage for all U.S. citizens age 65 or older. The APHENA investigators created daily counts of hospital admissions for cardiovascular disease (ICD-9, 390–429), chronic obstructive pulmonary disease (COPD) (ICD-9: 490–496, except 493), pneumonia (ICD-9: 480–487), and all respiratory diseases, conditions, or infections (ICD-9:460–519).

Meteorologic Data Weather data were obtained from the National Climatic Data Center EarthInfo CD-ROM and included daily measurements of temperature, dew point temperature, and relative humidity. Initial data were obtained from EarthInfo on September 2, 1996. Further data were requested from the National Oceanic and Atmospheric Administration on January 10, 2002, and supplementary data were downloaded from the web on October 8, 9, and 10, 2002.

Air Pollution Data These data were downloaded from the U.S. Environmental Protection Agency's (U.S. EPA) Aerometric Information Retrieval System and AirData System, now called the Air Quality System (*www.epa.gov/air/data/index.html*). Data were first requested on November 19, 2001. These data were downloaded in unaggregated form. Downloaded data included measurements of all criteria pollutants, except lead: PM_{10}, O_3, sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO).

The data on the gases SO_2, NO_2, and CO were available as hourly concentrations, while the data for PM_{10} were included as 24-hour averages. The data were screened to select population-based monitoring sites, and replicates were excluded. A location-specific average was calculated; if multiple measurements had been made, and a 365-day moving average had been calculated from the daily means (because the European and Canadian pollution data did not have a similar detrending), an analysis was carried out to document that there were no consequences for the analysis of the detrending. As in the original NMMAPS analyses, a 10% trimmed mean was then calculated for the pollution variables. Further details can be found in the NMMAPS reports (Samet et al. 2000b,c). The editing process for the resulting data set included visual screening of plots, as created by iHAPPS, for unusual values or time periods. Details about the decision process for creating and editing the air pollution data set are available on the iHAPPS Web site (*www.ihapss.jhsph.edu/*).

Census Bureau Data The 1990 census data were obtained with the initial development of the NMMAPS database. The 2000 summary file 3 was requested October 17, 2003.

APHEA

The APHEA contribution to APHENA derived from the APHEA2 project (Katsouyanni et al. 2001). Twenty-two research groups from 19 European countries and associated regions had participated in the APHEA2 project. They provided data from 32 cities: Athens (University of Athens Medical School); Erfurt (GSF Institute of Epidemiology); London and Birmingham (St. George's Hospital Medical School, University of London); Paris and Marseille (Insitut de Veille Sanitaire, Paris); Lyon (Universite Joseph Fourier, Grenoble); Barcelona, Bilbao, Valencia, and Madrid (IMIM program); Milan (A.S.L. Citta de Milano, Servizio di Epidemiologia); Rome (Osservatorio Epidemiologico Lazio); Turin (Pisa University); Netherlands (University of Groningen); Helsinki (National Public Health Institute); Dublin (R.C.D.H. Research & Education Institute); Tel-Aviv (Tel Aviv University); Istanbul (Istanbul Medical School); Basel, Geneva, and Zurich (Universitat Basel, Institute für Sozial und Praventivmedizin); Cracow, Lodz, Poznan, and Wroclaw (National Institute of Hygiene, Population Studies Laboratory); Prague (Charles University Medical Faculty); Teplice (OHS Teplice); Ljubljana (Institute of Public Health); Budapest (J. Fodor National Public Health Centre–National Institute of Environmental Health); Bucharest (Institute of Hygiene Public Health) and Stockholm (Umeå University). Because Istanbul provided data only after the end of the contractual period, it was excluded from the APHEA2 and APHENA analyses. Each research group prepared the data sets for their cities according to a standard approach and sent them to Athens (mortality series), or to London and Paris (admissions series), between October 1998 and June 2001.

Within APHEA2, the format of the files containing the detailed daily values for mortality, hospitalization, and air pollution data was predetermined. More specifically, all variable names were predefined and the same among the different cities. Missing values were noted by a blank field. Two files (one containing the outcome data and one the pollution data) from each city were sent to the three centers that analyzed the mortality series (Athens) and the admissions series (Paris for cardiovascular outcomes, London for respiratory outcomes) between October 1998 and June 2001. The original and final data files are available in the three centers that performed the analysis.

After the data collection step, a form was filled in for each city to document any missing variable names, percentages of missing values for each variable, and any comments about outliers, missing patterns, or peculiar values. Descriptive tables were circulated between the centers that performed the analysis and the individual cities for quality control purposes. The original data sets were corrected based on the results of this data review. More particularly, Bilbao provided missing values for humidity; London, Birmingham, Prague and Teplice provided updated pollution data; Barcelona was excluded from the SO_2 analysis due to problems in the pollutant's distribution (too many days with values of 10 $\mu g/m^3$); Valencia and Dublin provided corrected mortality data.

Mortality Data All cities (except Erfurt) provided data for the total daily number of deaths (excluding accidental deaths, i.e., ICD-9 > 800); the number of deaths from respiratory causes (ICD-9:460–519) and cardiovascular causes (ICD-9:390–459). Erfurt provided only all-cause mortality for all ages. Data for both sexes was provided for the following age groups: 15–64 years; 65–74 years, $\geq$ 75 years, and all ages. To the extent possible, the data were only provided for residents of the city who died in the city. Data covered at least three consecutive years between 1990 and 1997. The study period was longer than five years (1826 days) for most cities.

Hospital Admissions Data Barcelona, Birmingham, London, Milan, Netherlands, Paris, Rome, and Stockholm provided hospital admissions data. Admissions data were collected for all respiratory conditions (ICD-9:460–519)

from the following age groups: 0–14 years, 15–64 years, 65–74 years, ≥ 75 years, and all ages. Data were specifically coded for two respiratory subgroups: COPD (ICD-9:490–496 excluding 493), all ages only; asthma (ICD-9:493), 0–14 years, 15–64 years, and all ages.

Total cardiovascular admission data were collected from the following age groups: 15–64 years, 65–74 years; ≥ 75 years, and all ages. These conditions include cardiovascular disease (ICD-9:390–429), with the ischemic heart disease subgroup (ICD-9:410–414) specifically coded. Stroke admissions data (ICD-9:430–438) were also collected, using the same age groups as cardiovascular.

The definition of an emergency case was provided by each group contributing data on hospital admissions. Barcelona excluded emergency room visits that did not lead to admission. In Italy and Paris, emergency admissions could not be specifically identified, and an effort was made to select emergency cases by exclusions based mainly on diagnosis.

Meteorologic and Influenza Data Information was collected from all participating cities on daily values of: 24-hour average minimum and maximum temperature (°C); 24-hour average; percent relative humidity; day of the week, holidays, and any available data on influenza epidemics.

Pollution Data Based on the WHO and Regional Office for Europe revised guidelines for time averaging (WHO 2000); data for the following pollutants was collected: CO–maximum 8-hour average; O_3–maximum 8-hour average (mostly calculated as the 8-hour running moving average or the 8-hour average from 9 am to 5 pm); O_3–maximum 1-hour daily value; NO_2–maximum 1-hour daily value (or 24-hour average when 1-hour not available, as in the four Polish cities); SO_2–24-hour average; black smoke–24-hour average; and either PM_{10}–24-hour average or total suspended particles (TSP)–24-hour average.

The daily measurements of air pollution were obtained from the monitoring stations of the cities that participated in APHEA2. The European Union regulates the measurement methods of air pollutants (Commission of European Communities 1999), and recently most Central-Eastern European countries (that were not European Union members) have tried to comply with this regulation. Nevertheless, the recent directive of the European Union for the measurement of PM_{10}, replacing an older directive for black smoke, had not been applied during the time intervals studied, resulting in variability in PM_{10} measurement methods used. The mean daily concentration of each pollutant was calculated from as many monitoring stations as possible. A station was included in the calculation if it satisfied certain completeness criteria (Katsouyanni et al. 1996). More specifically, pollutant measurements were obtained from networks of monitoring stations situated in fixed and urban locations predetermined to be representative of each city (Schwartz et al. 1996). They were generally designated as either background or urban background monitors. Road-influenced sites, such as curbside or roadside sites were excluded, as were those influenced by point sources. Specifically for O_3, measurements in suburban regions were also included, because of the characteristics of O_3 dispersion. A station was excluded from the analysis if it had missing values for more than 25% of the study period. Despite the completeness criteria, a few missing values remained and were replaced according to Equation 1.

A missing value on day i of year k from monitor j was replaced by a weighted average $\hat{x}_{ijk}$ of the values of the other monitoring stations as follows:

$$\hat{x}_{ijk} = \overline{x}_{i.k}\left(\overline{x}_{.jk}/\overline{x}_{..k}\right) \quad (1)$$

where $\overline{x}_{i.k}$ is the mean value on day i of year k among all monitors reporting; $\overline{x}_{.jk}$ is the mean value for monitor j in year k; and $\overline{x}_{..k}$ is the overall mean level in year k. For days when all monitoring stations had missing values, provided that the total number of such days accounted for < 5% of the study period, the values were calculated by taking the mean of the values from the previous and next days. In the case of consecutive such days, the values remained missing in the final series. A dummy variable was included in the pollution data file for each pollutant, indicating the days with missing values from all selected monitoring stations.

All 31 cities provided data on SO_2, 19 cities provided data on CO, 21 on the O_3 maximum 8-hour average, and 23 on the O_3 maximum 1-hour daily value. Twenty-six cities (all except the Polish cities and Dublin) provided data on daily NO_2 maximum 1-hour value and 17 cities provided data on NO_2 24-hour average. Fifteen cities provided data on black smoke, 12 cities on PM_{10}, and 12 cities on TSP. For the PM_{10} analysis, 10 cities (Athens, Basel, Budapest, Cracow, Erfurt, Geneva, Milano, Rome, Turin, Zurich) had PM_{10} concentrations estimated using several methods: a regression model relating colocated PM_{10} measurements to black smoke measurements (Athens, Cracow), or to TSP measurements (Budapest, Erfurt), or as a percentage of TSP (based on measurements for the other cities). The details of the estimation are given below.

Of the 12 cities providing PM_{10} data, 10 had PM_{10} measurements (Barcelona, Birmingham, Helsinki, London, Madrid, Netherlands, Prague, Stockholm, Tel Aviv, and Teplice),

and 2 had PM_{13} measurements (Paris and Lyon). The PM_{13} data were accepted as a reasonable approximation to PM_{10} (Yvon LeMoullec 1999, personal communication).

For Basel and Geneva in 1993, the PM_{10}/TSP ratio was calculated in the Swiss Study on Air Pollution and Lung Disease in Adults (SAPALDIA) data as 0.71% (Ackermann-Liebrich et al. 1997). Harvard impactors were used for the PM_{10} measurements. For Basel and Zurich, using digital high-volume samplers for PM_{10} measurements in 1998, the ratios were 80% and 89%, respectively. For the APHEA study period of 1990–1995, Dr. C. Monn, the scientist involved with measurements in SAPALDIA, suggested using a conversion factor of 90% (personal communication 1998; Gehrig 1999).

A PM_{10}/TSP ratio of 75% was used for Milan, based on the results of Gianelle and colleagues (1991). For Turin, the PM_{10}/TSP ratio was considered to be 60%, based on the results of Cadum and colleagues (1999). For Rome, a value of 72% was used, based on the results of D'Innocenzio and colleagues (1998). For Cracow, colocated black smoke and PM_{10} monitors in two stations for 2042 days provided data for two regression lines used to predict PM_{10} concentration using the following equation:

$$PM_{10} = \alpha + \beta_1(\text{season}) + \beta_2(\text{black smoke } \mu g/m^3) \quad (2)$$

where season = 0 (April through September) or 1 (October through March). The two predictions were:

PM_{10} predicted = 34.017 − 5.248 (season) + 0.525 (black smoke $\mu g/m^3$), and

PM_{10} predicted = 50.816 − 13.851 (season) + 0.343 (black smoke $\mu g/m^3$).

The PM_{10} concentrations for Cracow were then based on the average of these two predictions.

For Athens, PM_{10} concentrations were estimated from a regression based on monitoring at neighboring stations (1 station for PM_{10} and 2 nearby stations for black smoke).

PM_{10} = 19.472 − 0.8017 (season) + 0.3326 (black smoke $\mu g/m^3$),

where season = 0 (April through September) or 1 (October through March).

The data set created using these approaches was a daily time series of pollutant concentrations. The number of monitoring sites per city ranged from 1 to 12. The number of observations per site were required to adhere to the APHEA completeness criteria; consequently the database contained no systematically missing values and the number of randomly missing values was small. Each research group was responsible for selecting monitors to use, for inputting or estimating the daily averages, and for filling in missing data before sending the data for the time-series analyses. The method for developing PM_{10} measurements used in each city is not included in the APHEA files; that is, the days that are predicted are flagged, but the specific monitoring method is not identified.

To assess the consequences of any error from this approach to developing a complete set of daily PM_{10} time-series data for the 31 cities, the investigators compared estimates of the effect of PM_{10} on mortality from all 31 cities to estimates from the 12 cities that had actual PM_{10} measurements. When all 31 cities were used in the second-stage analysis, the estimated percentage increase in mortality per 10-$\mu g/m^3$ PM_{10} was 0.68% (95% CI; 0.6 to 0.8) using the fixed effects model and 0.62% (0.4 to 0.8) using the random effects model. When only the 12 cities with original measurements were used, the estimates were 0.70% (0.6 to 0.9) using the fixed effects model and 0.70% (0.5 to 0.9) using the random effects model.

Canadian Studies

The Canadian data sets used in APHENA had originally been developed for a number of analyses that involved multiple cities and examined mortality or hospitalization as outcomes. Twelve Canadian cities were selected based on available air pollution monitoring data (Table 1). The geographic region defining the study area was based on the location of the air pollution monitoring stations in the surrounding region. The definition of study area was based on either the census division or combination of census subdivisions. The Canadian outcome data were obtained through the Canadian Institute for Health Information (CIHI), established in 1994 by the federal, provincial, and territorial ministers of health in response to the need for coordinating national health information. To ensure the quality of the data, CIHI established a comprehensive and systematic data-quality program with periodic data-quality checking, particularly in 2001–2002 during the implementation of the 10th ICD (International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Canada) and Canadian Classification of Health Interventions.

Mortality Data Computerized database for mortality information has been available in Canada since 1959. These data were obtained for analysis. Information includes demographics of the deceased person (name, sex, age, place of residence), as well as the single underlying cause of death and decedent information. Data were obtained directly from the vital statistics department of each province or territory

Table 1. The Twelve Canadian Cities Selected, Based on Available Air Pollution Monitoring Data[a]

City	CD	CSD	Number of Monitors PM	Number of Monitors O_3
Halifax		1209018, 1209021, 1209022, 1209024	3	4
St. John		1301006, 1305008, 1305009, 1305010, 1305012, 1305051, 1305053, 1305056, 1305058	2	5
Quebec	2423	2424020, 2425025	3	6
Montreal	2466		5	19
Ottawa	2481, 3506		2	5
Toronto	3520		5	17
Hamilton		3525018	1	5
Windsor		3537039	2	2
Winnipeg		4611040	1	2
Edmonton		4811061	1	3
Calgary		4806016	2	3
Vancouver	5915		3	23

[a] Geographic region defining the study area was based on the location of the air pollution monitoring stations in the surrounding region. Each community study area was defined based on either the census division (CD) or combination of census subdivisions (CSD) listed.

and are available at provincial and national levels. The mortality data can be obtained from official agencies such as Canada Vital Statistics Information System and CIHI at the Discharge Abstract Database that registers death discharges from hospitals.

Deaths reported in the Canada Vital Statistics Information System and the Discharge Abstract Database were classified by cause of death using ICD-9. Mortality data were collected for the ICD-9 codes specified in the APHENA protocol (Appendix B).

Hospital Admissions Data The CIHI Discharge Abstract Database covers the population attended by all acute care hospitals in participating provinces and contains data on discharges, deaths, sign-outs, and transfers. All hospitals are bound by provincial or territorial legislation to maintain health records for every person seen by medical personnel. All reporting units are primarily controlled by the respective provinces and territories and are required to submit data on a yearly basis. Until 2001 the databases contained a mixture of data originally submitted to CIHI with three classification schemes (ICD-10-CA, ICD-9, and ICD-9-CM). CIHI made an effort to produce a nationally-comparable data set, creating conversion tables that are used to map ICD-10-CA diagnosis back to ICD-9. An extensive quality assurance program is in place. Compilation of the Canadian database for hospital admissions followed the APHENA protocol (Appendix B).

Pollution Data The Canadian pollution data were obtained from Environment Canada, which administers the National Air Pollution Surveillance (NAPS) program. The NAPS, in existence since 1970, is a cooperative partnership of federal, provincial, territorial, and some regional governments that measure air quality. The goal of the NAPS program is to provide accurate and long-term air quality data of a uniform standard throughout Canada. An extensive quality assurance program is in place. Data from the NAPS program are included in the Canada-wide Air Quality Database and are published in annual air quality data summary reports available at *www.etc-cte.ec.gc.ca/publications/napsreports_e.htm*.

The number of sites used to derive daily measurements for each of the cities in the Canadian component of APHENA are presented in Table 1. Although these numbers remained constant for the entire study period, daily measurements were derived from fewer monitors on days when one or more monitors were not functional.

For O_3, measurements were available on an hourly basis. Information for a given day was considered to be missing if more than 25% of the hourly measurements were missing. A single daily measurement was obtained for a city by averaging the results for all the monitors in that city, even if the daily measurement was missing for one or more monitors on that day. Single daily values for PM_{10} were obtained by averaging the 24-hour cumulative mass measurements among the monitoring stations measuring PM_{10} within each community.

FIRST-STAGE ANALYSIS: CITY-SPECIFIC TIME-SERIES DATA MODELING APPROACH

Background

A variety of analytic methods have been applied in time-series studies of mortality and morbidity (Bell et al. 2004b). In particular, approaches to control potential temporal confounding by such time-varying factors as season, weather, and influenza epidemics have not been uniform across studies. Achieving adequate control for potential time-variable confounding factors is critical for air pollution studies, but over- or under-control can lead to spurious

associations. Methods for control of such factors have evolved substantially over the last 15 years. Some of the earliest time-series studies used techniques to control for season and weather that would now be regarded as simplistic, such as indicator variables for season and hot days, and linear terms for weather factors (Schwartz and Marcus 1990). GLM and, in particular, Poisson regression techniques with parametric functions of time and weather variables (linear, quadratic, or sinusoidal curves of differing periodicities) replaced these early techniques about a decade ago (Fairley 1990; Schwartz et al. 1996; Kelsall et al. 1997). Since then, such standard regression techniques have in turn been almost fully supplanted by the more flexible GAM with nonparametric smoothing functions for modeling nonlinear relationships of time and weather variables (Dominici et al. 2000b; Samet et al. 2000a; Katsouyanni et al. 2001; Touloumi et al. 2004). Sinusoidal terms were used in the APHEA Phase 1 (APHEA1) project (Katsouyanni et al. 1996; Schwartz et al. 1996), but were replaced by GAM in APHEA2.

Thus, before APHENA was implemented, no one method had been accepted among investigators for carrying out the first-stage analyses in individual cities. Approaches had evolved as new methodologies and software became available and limitations of previous methods were identified. The APHENA investigators needed to select a preferred approach to smoothing and also to identify the optimum number of degrees of freedom. An understanding of the sensitivity of findings to these key aspects of model development was needed to assure transparency of the findings.

The APHENA investigators also probed the differences among the methods used in prior analyses. The NMMAPS (Dominici et al. 2000b; Samet et al. 2000a) and APHEA2 (Katsouyanni et al. 2001; Touloumi et al. 2004) projects involved analysis of mortality and morbidity data from a large number of cities from the United States and throughout Europe, respectively; the Canadian investigators had also analyzed multiple cities. In the NMMAPS and APHEA2 studies, a two-stage hierarchical approach was adopted. In the first stage, data from each city were analyzed separately using GAM with nonparametric smoothing functions for time and weather variables; in the second stage, evidence across cities was combined. The results from the second stage may have been affected by the methods of the first-stage analysis, particularly if they created biased estimates. Therefore, in APHENA, several sets of sensitivity analyses were performed with the objective of characterizing the analytic approaches used in the two studies, and assessing the sensitivity of overall findings to the choice and specification of models.

Simulation Studies

To assess the merits of the many different statistical models and procedures used in the literature, the APHENA investigators conducted extensive simulation studies to test the various methods. The simulations focused on two critical modeling aspects: (1) the choice of the smoothing method and basis functions used to estimate the smooth function of time in the city-specific models; and (2) the number of degrees of freedom to be used in the smooth function of time. The investigators also evaluated whether each city should be assigned the same model specification or whether each city-specific model should depend on city-specific characteristics. For the former, the same degrees of freedom (ranging from 1 to 20 *df*/year of data) were assigned to the smooth function of time for every city. The range was determined by choosing the minimum possible degrees of freedom per year up to a maximum degrees of freedom per year that essentially removed all variation in the data beyond time scales of one week. Also, the collective experience of the investigators indicated that using more than 20 *df*/year does not substantially affect the risk estimates. For the latter approach, the degrees of freedom for the smooth function of time were chosen separately for each city using a fit criterion, such as the Akaike Information Criterion (AIC), or by minimizing the PACF of the residuals.

The APHENA investigators chose the smoothing method used to estimate the smooth function of time in the city-specific models from among the methods commonly used in the literature, namely locally-weighted scatterplot smoothing (LOESS), SS, NS, and PS. For choosing the specific degrees of freedom to be used when estimating the smooth function of time, a number of automatic methods were employed. These methods were minimizing AIC, minimizing the PACF of the residuals, and minimizing the generalized cross-validation (GCV) criterion.

The decision to examine the smoothing techniques and the degrees of freedom was motivated in part by uncertainty in the literature as to the appropriate approach for analysis of these data. While there are many other aspects of model development, these two issues were considered central to the robustness of model estimates and to their interpretation. In particular, the degrees of freedom used determine the level of adjustment for unmeasured confounders. The point estimates of risk associated with air pollution can be greatly affected by using too few or too many degrees of freedom.

A no-threshold model was assumed for the simulation studies. Previous work (Daniels et al. 2000) gave little evidence of a threshold for short-term effects of PM_{10}. Therefore, the investigators chose to simulate data from a no-threshold model to simplify the analysis. A separate analysis of thresholds conducted within the APHENA group indicated that if such a threshold were to exist, it would be relatively difficult to detect given the inherently small relative risks being estimated.

The approach to the simulation studies first required generating Poisson time-series data from a specified model that used parametric basis functions to represent the smooth function of time. Parameters for simulating the outcome data were calibrated to match mortality levels in data sets from Europe and the United States. For Europe, data from London, Cracow, and Madrid were used, while for the United States, data from Minneapolis–St. Paul were used. Given a simulated data set, each combination of basis (NS, PS, LOESS, SS) and automatic fit criterion (AIC, PACF, GCV) was applied and an estimate of the degrees of freedom and risk coefficient obtained. This process was repeated for each of the simulated data sets, and the estimated values were compared to the true values. For each basis–criterion combination, the bias and variance of the estimator were examined. Full details for all the simulations conducted, including values used to simulate the data, were published by Touloumi and colleagues (2003) and by Peng and colleagues (2006a).

Comparison of Time-Series Models: Results from the Sensitivity Analyses

At the time the APHENA project was initiated, GAM Poisson models had been selected for use in the first stage of analysis. The models are of the form:

$$\ln\left(\mu_t^c\right) = \ln\left[\mathrm{E}\left(Y_t^c\right)\right] = \alpha_0 + \sum_{j=1}^{q} \mathrm{f}_j^c\left(X_{tj}^c\right) + \beta^c P_t^c + \sum_{d=1}^{6} \beta_d DOW_d \qquad (3)$$

with variances $v_t^c = \varphi^c \mu_t^c$. Y_t^c is the observed count of the relevant health outcome at city c on day t, β^c is the effect estimate for P_t^c, the pollutant concentration in city c on day t, (PM_{10} in this case), X_{tj}^c are the nonpollution predictor variables (i.e., time, mean daily temperature and mean daily relative humidity), f_j^c are smooth functions of these variables, and DOW_d are indicator variables for the day-of-the-week, μ_t^c is the expected count of the relevant health outcome in city c on day t, and φ^c is the overdispersion parameter. Two smoothing functions, LOESS or SS were evaluated. (Both methods are described in detail in several books; see for example, Hastie and Tibshirani [1990].)

However, reports published during the planning stage of APHENA showed that there are at least two major limitations to applying GAM in time-series studies of air pollution and health. First, use of the default convergence criteria in the S-Plus statistical software may result in biased effect estimates, although this bias can be remedied by using more stringent convergence criteria (Dominici et al. 2002b). Second, in the presence of concurvity, the nonparametric analogue of multicolinearity, GAMs give seriously underestimated standard errors of the model parameters (Ramsay et al. 2003). The underestimation happens because, in most statistical software, standard errors of the linear part of the GAMs are estimated based on a linear approximation of the smooth (nonlinear) part of the model (Ramsay et al. 2003). Concurvity in air pollution data has been found to be as high as 0.6 (Dominici et al. 2002b; Ramsay et al. 2003). An additional issue that became apparent was the sensitivity of estimates to using too much or too little smoothing (Health Effects Institute 2003).

In response to the identification of these problems, a return to GLM using NS for secular and seasonal trends was considered, whereas PS emerged as an attractive alternative that would bridge the GLM with the GAM approach. Regardless of the class of models to be selected, the more general issue regarding the criteria to be used for selecting the appropriate degree of smoothness for temporal confounders remained unsolved. Consequently, a set of additional sensitivity analyses was carried out to guide the selection of the models and the degrees of freedom to be used. Parallel simulation studies were carried out by the APHEA2 and NMMAPS groups.

Within the APHEA2 group, a simulation study was conducted by Touloumi and colleagues (2006). Poisson time-series data were generated from a fully parametric model. They tested the basic functions (LOESS, NS, SS, and PS), and evaluated several criteria for choosing the degree of smoothing, including minimization of the absolute value of the sum of the PACF, AIC, fixed degrees of freedom, and GCV. The results of the simulation study are presented in Appendix Table A.1. Overall bias in air pollution effect estimates ranged from −31% to 15%, depending on the method. Splines (both NS and PS) tended to underestimate the air pollution effect, whereas LOESS tended to overestimate it. Nonparametric methods underestimated the standard error of the air pollution regression coefficient, PS gave relatively small bias, and PACF in combination with PS performed relatively well in terms of bias.

In parallel, another simulation study was carried out by the NMMAPS group, and results were reviewed at an investigators' meeting. Data were generated from nonparametric models under several scenarios reflecting different levels of confounding. The performance of each modeling

method was evaluated using the mean squared error (MSE) of the estimated air pollution effect as the criterion. While results were relatively robust to different methods, the overall conclusion was that the bias decreases as the effective degrees of freedom increase, and that model selection methods that optimize prediction (like MSE) may not be suitable for obtaining an estimate with small bias (Peng et al. 2006a).

Further sensitivity analyses designed to evaluate the robustness of the air pollution effect estimates with respect to the degrees of freedom were carried out within APHENA. Given the different seasonal patterns that different health outcome series exhibit, those analyses were carried out for all outcomes in APHENA.

Based on the simulation results, the APHENA group decided in 2003 to use NS and PS for initial exploratory city-specific analyses with a wide range of degrees of freedom to demonstrate any sensitivity of results to the degrees of freedom selected (i.e., 6, 8, 10, 12, 14, or 16 *df*/year for mortality data; 6, 8, 10, 12, 14, 16, or 18 *df*/year for respiratory admissions; 4, 6, 8, 10, or 12 *df*/year for cardiovascular disease admissions). Because previous results have shown that air pollution effect estimates tend to stabilize after a certain number of effective degrees of freedom (Dominici et al. 2004), the best model in each city for both methods (NS, PS) was chosen to be the model with the degrees of freedom corresponding to the second estimate after the values of the estimates stabilized. In addition, the PACF criterion for the optimal choice of degrees of freedom was used. Both approaches lead to city-specific choices. However, further sensitivity analysis clarified that overall effect estimates (i.e., estimates pooled over several cities) tended to stabilize at high degrees of freedom, while such stabilization was not observed for all city-specific results. The results also showed that NS is more sensitive than PS to fitted degrees of freedom.

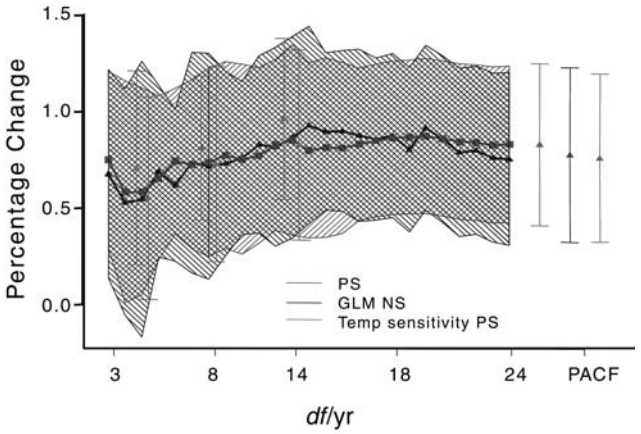


Figure 1. Sensitivity analyses for three European cities: Pooled percentage change and 95% CIs for all-cause mortality per 10-µg/m^3 PM_{10} increase.

Regarding temperature control, results from additional sensitivity analyses showed that air pollution mortality estimates were robust to increasing the degrees of freedom for temperature to more than three (see Figure 1). Additional sensitivity analyses for handling missing data were carried out. These analyses showed that particular attention is needed in specifying days with missing air pollution data when these are not systematic (i.e., random). Additionally, in the presence of systematic, missing data (i.e., measurements available every 6 days), the performance of the methods varied (e.g., PACF seemed not to perform as well). In light of the above results, and provided that estimates from a core model or a limited set of core models were needed for the second-stage analysis, decisions about the first-stage analysis were once again revised, and the first-stage analytic protocol was finalized.

Final Models for City-Specific Analysis

The final models used for city-specific analysis are described in Table 2. NS and PS were used as smooth functions for trend, seasonality and temperature control. The

Table 2. Description of Final Models Used for City-Specific Analyses Including Outcomes Examined, Pollutants, Exposure Lags, and Smoothers

- Outcomes
  - Mortality
    - Cardiovascular disease (< 75, ≥ 75)
    - Respiratory (≥ 75, all ages)
    - Total non-accidental (< 75, ≥ 75, all ages)
  - Hospital admissions
    - Cardiovascular disease (≥ 65)
    - Respiratory (≥ 65)
- Pollutants
  - O_3 (daily 1-hour maximum)
  - PM_{10} (24-hour average)
- Exposure lags
  - Lag 1 (all pollutants)
  - Average of lags 0 and 1 (O_3 and PM_{10} where complete)
  - Distributed lag (O_3 and PM_{10} where complete)
- Smoothers for time trends
  - Natural splines (NS)
  - Penalized splines (PS)
- Degrees of freedom per year for smooth function of time
  - 3, 8, 12, and *df* chosen by minimizing PACF
- Other variables
  - Temperature at lag 0 and lag 1, day of the week indicator, holiday indicator

selected degrees of freedom were 3, 8, and 12 per year. Additionally, for complete time series (i.e., series without systematic missing values), minimization of the PACF was used to select the optimal degrees of freedom using the PS method. The degrees of freedom chosen by the PACF were then used to fit an NS model. For temperature control, smooth terms of lag 0 and lag 1 were introduced in the model. The degrees of freedom were set to three for both terms. Relative humidity and dew point were not included in the models. Dummy variables for day of week and holidays were included in the final models, but variables for influenza epidemics were not. For models based on the minimization of the PACF criterion, autoregressive terms were introduced if necessary (i.e., significant autocorrelation remained in the final model's residuals). The pollutants considered in the models were PM_{10} (24-hr average) and O_3 (1-hr maximum).

Hospital Admissions Data The protocol stated that different degrees of freedom for different periods within a year may be necessary for admissions series, given that admissions data show seasonal patterns such as sharp drops during vacation seasons and steep increases during respiratory infection epidemics. Control for influenza was accomplished through a separate smoother for winter respiratory epidemics when appropriate.

O_3 Seasonal Analysis Apart from the annual data, the effects of O_3 were also analyzed by season. The models using only half-year data used dummy variables for the months (by year) instead of splines. The warm season was defined as April through September. Temperature lags were used for summer-only models (for NS, 2 *df* for lag 0 and linear term for lag 1; for PS, 3 *df* for lag 0 and linear term for lag 1).

For a detailed description of the APHENA methodology protocol, see Appendix B.

Examining Concentration–Response and Threshold Analysis for APHENA

Methodological work was carried out to meet another of APHENA's goals: assessment of the potential for characterizing the form of the concentration–response relationships between PM_{10} and mortality and between O_3 and mortality. First, a simulation study was conducted to explore the behavior of the methodology for detecting a nonlinear concentration–response relationship and to determine whether sufficient information was available from the actual data to estimate a threshold. Second, the NMMAPS database was analyzed to determine if there was any evidence in the data for such a nonlinear concentration–response relationship.

The particular nonlinear concentration–response model used was a threshold model, or *broken line* model, which assumes the effect of exposure to be zero for values of the exposure below some specified value h. For values of the exposure greater than h, the effect is assumed to be linear with exposure. Such a model takes the outcome Y_t, which can be a daily count of mortality or hospital admissions, and relates it to the pollutant concentration (a measure of exposure) on day t, P_t, as $Y_t = \beta_0 + \beta_1 (P_t - h)_+ +$ confounders, where $(P_t - h)_+$ is zero when $P_t < h$, β_0 is the baseline of daily counts, and β_1 is the concentration response for the pollutant.

For the simulation study, the APHENA investigators simulated mortality and PM_{10} or O_3 data, assuming four different relationships (expressed as log-relative risk) for the association between PM_{10} or O_3 and all-cause mortality. For each of those four log-relative risks: 0.01, 0.005, 0.001, and 0.0005, the investigators generated 250 data sets. In each data set, the true threshold value was assumed to be zero (i.e., no threshold) and the length of the daily time series was 10 years. The reason for choosing the true threshold value to be zero was to determine if the model would correctly provide enough substantial evidence to favor a threshold of zero in the case where no threshold existed, and to examine whether such a model would incorrectly identify strong evidence of a threshold when no threshold existed. Given that any threshold, if it exists, would likely exist at relatively low concentrations, the investigators considered it appropriate to run the simulations using the most conservative scenarios.

For each data set, the AIC was calculated for a GLM fit to the data using a range of thresholds. The thresholds chosen for both PM_{10} and O_3 were 0, 5, 10, 15, … , 75 $\mu g/m^3$ (ppb in Canadian analyses of O_3). After all models were fitted and AIC values computed, the model that minimized the AIC was selected. This process was repeated for each of the simulated data sets. This approach is similar to that used by Daniels and colleagues (2000).

The simulation studies indicated that detecting a threshold in a broken line type of model would be difficult if the true association between PM_{10} and mortality were relatively small (i.e., $\beta_1 = 0.0005$), as is the case in many time-series studies. For a single-city, there appears to be little potential for discriminating among possible thresholds if the thresholds are at relatively low (i.e., $h < 15$ $\mu g/m^3$) concentrations of PM_{10}.

In multicity studies, evidence can be combined from multiple locations, a feature that may assist in an investigation of concentration–response relationships. To complement the simulation analysis, the investigators analyzed the NMMAPS database using the methodology of Daniels and

colleagues (2000) to determine if there was evidence of a threshold in the 15 NMMAPS cities for which daily PM_{10} data were available. The APHENA investigators analyzed the relationship between PM_{10} lag 1 and all-cause mortality for each city, using a broken line model similar to that used in the simulation study. In this model they also controlled for the relevant confounders specified in the first-stage protocol. This model was fit for a number of different threshold values, and the AIC was computed each time. Then, for a given threshold, the investigators averaged the AIC values across cities to obtain an average AIC, and then selected a threshold to be the minimizer of this average AIC. The full procedure is summarized below.

1. Choose a grid of threshold values 0, 5, 10, … , 75 $\mu g/m^3$ of the pollutant.
2. Choose a threshold value h in the grid.
   a. Choose a city c.
   b. Fit the threshold model using
      i. All-cause mortality, all ages
      ii. 8 *df* for the smooth function of time
      iii. NS to represent all of the smooth functions
   c. Given the fitted model, compute the AIC of the model:

      $AIC_c(h)$ = Deviance + 2(# parameters),

      where the number of parameters does not include h as a parameter.
   d. Go back to Step 2a and repeat for all cities.
3. After the AIC values for all cities are computed, construct the sequence $AIC_1(h)$, … , $AIC_M(h)$, where M is the total number of cities. Compute the average AIC value for a given threshold as $AIC_{avg}(h) = (1/M)(AIC_1(h) + AIC_2(h) + \ldots + AIC_M(h))$.
4. Go to back to Step 2 and repeat for all values of h.
5. Given $AIC_{avg}(h)$ for all values of h, we can choose the value which minimizes $AIC_{avg}(h)$, in other words $h^* = \arg\min AIC_{avg}(h)$.

The APHENA investigators fit the threshold models for PM_{10} (all cities) and O_3 (cities with full-year data only).

SECOND-STAGE ANALYSIS

Analytic Approaches for Mortality

For the second stage analysis of the APHENA project, the investigators developed and compared two different approaches for combining information and exploring heterogeneity across the United States, Europe, and Canada: hierarchical and metaregression models. These methods were needed to accomplish the overall objective of exploring potential contributors to any heterogeneity in the effects of PM_{10} and O_3 in the APHENA regions. Previous analyses of the three databases, and of the broader air pollution literature in general, provided the basis for considerations of potential determinants of heterogeneity. Potential sources of heterogeneity in the effects of the two air pollutants were grouped broadly as:

- methodological, reflective of differences in the nature of the data on air pollutant concentrations and health outcomes or
- biological, reflective of differences in the nature of the air pollution mixtures or in the underlying susceptibility of the populations to air pollution.

Furthermore, given the range of pollutant concentrations represented by the three geographic areas, the investigators recognized that the earlier decision to largely fit linear models could also create seeming heterogeneity in the effects of air pollution if the true concentration–response relationships were nonlinear.

Hierarchical Models The APHENA investigators developed hierarchical models (Morris and Normand 1992) to pool the city-specific estimates within the United States, Europe, and Canada and to explore determinants of heterogeneity. Hierarchical models provide a unified and flexible framework for estimating pollutant effects for particular cities, regions (*region* refers to the United States, Canada, and Europe), or combinations thereof while accounting for covariate effects and components of variation. They facilitate more precise estimation of relative rates within each city than can be accomplished by analysis of data for each city individually. In addition, the hierarchical approach allows the estimation of city-specific, region-specific, and overall air pollution mortality relative rates, while accounting for variability in the air pollution mortality relative rates across locations within a region and across regions. For example, if there is substantial heterogeneity in the relative mortality rates across cities within a region (or across regions), relative rate estimates for the region (or across regions) will be less precise. The estimation of the city-specific, region-specific, and overall relative rate, and of between-city and between-region variability can be carried out with Monte Carlo Markov Chain methods. One useful feature of these methods is that they provide, in addition to the point estimate and confidence interval (CI), an approximation of the entire distribution of the unknown parameters.

To estimate city-specific, region-specific, and overall pollution relative rates of mortality for the multiple locations

in Canada, the selected European countries, and the United States, we applied a three-stage hierarchical model describing: (1) within-city variability; (2) within-region variability, and (3) between-region variability. In the first stage, we applied the agreed-upon Poisson log-linear regression models to estimate the city-specific relative rates of mortality. We assumed:

$$Y_t^c \sim \text{Poisson}(\mu_t^c)$$

$$\log \mu_t^c = \beta^c{}_r P_t^c + \text{confounders} \quad (4)$$

where Y_t^c and μ_t^c are the observed and the expected number of deaths on day t in city c, P_t^c is the pollution concentration on day t in city c, and $\beta^c{}_r$ is the parameter of interest that measures the log-relative rates of mortality for a unit of increase in air pollution in city c within region r. Additional terms were included in the model to adjust for trend, seasonality, weather, and other potential confounding factors.

In the second stage of the model (the within-region analysis), we described between-city variation in the true log-relative rates within a region and evaluated heterogeneity across cities. Demographic and socioeconomic factors were considered as potential determinants of heterogeneity. We assumed:

$$\beta_r^c = \beta_r^* + \sum_{j=1}^{c} \beta_{jr}\left(X_j^c - \bar{X}_j\right) + error^c \quad (5)$$

We centered all predictors, j, with respect to their mean, so that the intercept, β^*_r, can be interpreted as the log-relative rate of mortality for air pollution in region r when all the predictors are centered at their mean values. We considered the following potential effect modifiers, X: mean levels of pollutants, temperature, dew-point, mortality rates, sociodemographic variables, urbanization, and variables related to measurement error. For this second stage to be informative, data from a sufficient number of locations were needed.

The third stage describes between-region variation in the true city-specific coefficients and estimates the overall air pollution effects for the U.S., European, and Canadian cities, and for all three regions combined. We assume:

$$\beta_r^* = \beta^* + error_r \quad (6)$$

where β^* denotes the overall (U.S., Canadian, and European locations) log-relative rate of air pollution. Because of the limited number of regions, estimates of the between-region variability and of potential effect modification are sensitive to prior distributions used for the unknown parameter.

The Bayesian formulation of the model is completed by specifying dispersed but proper baseline conjugate prior distributions. To approximate the posterior distribution of all of the unknown parameters, we implemented a Monte Carlo Markov Chain algorithm with a block Gibbs Sampler. The full conditional distributions are available in closed form. Their derivations are routine and not described here.

Metaregression Models Meta-analytic techniques for the aggregation and synthesis of prior research have a long history in observational and experimental epidemiology. Meta-analysis is essential for obtaining reproducible summaries of study results and valuable for discovering patterns among study results. Meta-analysis of observational data remains a controversial topic. Meta-analyses of independently published results may be limited by such problems as selection bias, comparability of health outcomes and exposure variables, statistical analysis, and lack of information on confounding factors. However, a planned second-stage meta-analysis of comparable, analyzed first-stage data — as proposed for this project — addresses most of these issues.

To provide a quantitative summary of all city-specific results and to investigate potential effect modifiers, the APHENA investigators applied univariate (for one-pollutant models) or multivariate (for multiple-pollutant models) metaregression models. In such models, fixed-effects pooled regression coefficients are estimated by weighted regression of city-specific estimates on potential effect modifiers (at city level) with weights inversely proportional to their city-specific variances. If significant heterogeneity among city results remains beyond the variation associated with fixed effects, random-effects metaregression models can be applied.

Models for multivariate metaregression follow the form:

$$\beta^c = X_i^c \alpha_i + \delta^c + \epsilon^c \quad (7)$$

where β^c is the vector of the estimates of interest that measures the log-relative rate of mortality for a unit of increase in air pollution in city c for a particular pollutant; $\mathbf{X}_i^c$ is a matrix containing the observed city-level covariates i for city c; its pattern of entries indicates which pollutant's effect appears in each row of β^c and to which pollutant effect each covariate relates; α_i is the vector of regression coefficients being estimated; it may include a separate intercept for each pollutant and a separate slope for each pollutant against each corresponding covariate; δ^c is a vector of p random effects associated with city c representing, for each pollutant, the city's deviation from the average of all cities having the same values of covariates, and ϵ^c (assumed

independent from δ^c) is the vector of random sampling errors within each city.

The $p \times p$ matrix $\text{cov}(\delta^c) = \mathbf{D}$ (to be estimated) represents the between-cities covariance that is unexplained by the fixed effects (i.e., the regression).

It is assumed that:

$$\begin{aligned} \delta^c &\sim \text{MVN}\ (\mathbf{0},\mathbf{D}) \\ \epsilon^c &\sim \text{MVN}\ \left(\mathbf{0},\mathbf{S}^c\right) \\ \beta^c &\sim \text{MVN}\ \left(\mathbf{X}^c\alpha,\mathbf{D}+\mathbf{S}^c\right) \end{aligned} \quad (8)$$

where MVN is multivariate normal, $\mathbf{S}^c$ is the estimated variance-covariance matrix in city c for the p pollutants. When $\mathbf{D} \approx 0$ we have the corresponding fixed-effects estimates, while when $\mathbf{D} \neq 0$ we have the random-effects estimates.

To estimate the model parameters, the method described by Berkey and coworkers (1998) was applied. Contrary to the usual metaregression, in which results from each pollutant are analyzed separately, the multivariate metaregression provides more accurate estimates by incorporating the correlation among pollutants within each city. Specific S-plus functions were written to fit the univariate and multivariate metaregression models.

Univariate metaregression models represent a specific case of the multivariate metaregression models where p = 1 (one-pollutant city-specific models). In such cases, the between-cities variance is estimated from the data using the maximum likelihood method described by Berkey and coworkers (1998) and added to the city-specific variances.

Comparison of the Hierarchical and Metaregression Models While we expect that the two approaches should yield similar results regarding the pooled estimates of pollutant effects within each region (i.e., Europe, the U.S., and Canada), they may differ substantially in explaining heterogeneity across regions.

In the hierarchical model approach, region constitutes the highest level of hierarchy (level 3); cities nested within each region constitute a second level (level 2); and successive days are the lowest level of hierarchy (level 1). At each level of hierarchy, factors that differ across that level (e.g., across cities) can be used as potential covariates. However, given the small number of regions in this analysis, potential effect modifiers at that level could not be examined.

The metaregression models are conceptually similar to two-level hierarchical models with cities constituting the highest level (level 2) of hierarchy and successive days constituting the lowest level (level 1) of hierarchy. In such an approach, regions can be considered as another covariate. In this structure, we can examine whether the effects of several modifiers are similar across regions, as well as investigate whether region explains part of the possible heterogeneity among city-specific results that is unexplained by other effect modifiers. However, metaregression models easily can be extended to incorporate nations or regions as an additional level of pooling data. In such an approach, metaregression models are applied to pool results in each region. Then overall estimates are obtained by pooling the region-specific estimates. Effect modification can then be explored at each level of aggregation using the methodology described previously.

The joint mortality analysis, which is designed to obtain pooled estimates and explore heterogeneity, relied on a database comprised of data from 30 cities in Europe (APHEA2), 90 cities in the United States (NMMAPS), and 11 cities in Canada.

Analytic Approaches for Hospital Admissions

For the hospital admissions series, the individual city analysis and the methodology for pooled analysis and exploration of heterogeneity was the same as for the NMMAPS and for the APHEA2 mortality time series. A smaller number of cities were analyzed for admissions than for mortality, due to difficulties involved in compiling admissions data as well as to limited data availability. The analysis thus had less power for explaining heterogeneity.

NMMAPS data on hospital admissions for people ≥ 65 years from 14 cities were analyzed using distributed lag (DL) models developed by Schwartz (2000b). These models are based on the knowledge that the health effects of pollution on one day may spread over several subsequent days. They assume that:

$$\text{Log}\,[\text{E}(Y)] = \alpha_0 + \text{f}_1(X_1) + \ldots + \text{f}_p(X_p) + \beta_0 P_0 + \ldots + \beta_p P_p \quad (9)$$

where X_i are covariates for pollutants 1 through p, α_0 is the intercept, β_0 estimates the corresponding pollutants, f is a smooth function of X, and P_0 is the pollutant concentration on the concurrent day, P_1 on the previous day, etc. We can take advantage of our multicity analysis to fit this model without constraints on the β's. Control for confounding by season and weather was consistent with the approach used for APHEA2. Effect modifiers examined in the second-stage metaregression included the regression coefficients between pollutants and social factors, similar to those used in the NMMAPS mortality analysis. The hospital admissions analysis shared the same database on potential effect modifiers as described in the section on mortality.

The results of the individual city analysis were combined and heterogeneity explored using the methods described in the mortality meta-analysis section.

SIMULATION STUDY TO COMPARE METHODS FOR POOLING RESULTS ACROSS CITIES

One of the tasks of the APHENA project was to carry out a comparison between the hierarchical approaches used by the NMMAPS and APHENA groups for pooling relative-rate estimates across cities. The NMMAPS group had used a normal–normal hierarchical model and Monte Carlo Markov Chain methods for a computation called two-level normal independent sampling estimation (TLNISE) (Everson and Morris 2000). The APHENA group had used the same normal–normal hierarchical model but a different computational strategy based on the Berkey maximum likelihood estimation (MLE-Berkey; Berkey et al. 1998). To compare the two computational approaches for fitting the same hierarchical model, the investigators carried out a simulation study.

Data Generating Mechanism

Starting with a model of the form

$$\begin{aligned} \beta^c &\sim N\left(\alpha,\sigma^2\right) \\ \hat{\beta}^c \mid \beta^c &\sim N\left(\beta^c, V^c\right) \end{aligned} \tag{10}$$

where α is the national average log-relative risk and σ^2 is the heterogeneity variance of the city-specific log-relative risk β^c. β^c and V^c are known and equal to the NMMAPS estimates using 1987–2000 data and lag 1. The investigators generated 1000 data sets (j = 1, ... ,1000) for each *true value* parameter combination of α and σ shown in Table 3. For each data set, the investigators re-estimated the model parameters, once with TLNISE and once with MLE-Berkey. The estimates of α and σ obtained under the two methods are denoted as $\hat{\alpha}_{TL}$ and $\hat{\sigma}_{TL}$, $\hat{\alpha}_{BM}$ and $\hat{\sigma}_{BM}$, respectively.

Mean Squared Error Calculation

The investigators then calculated the MSE of each estimator. For example, for $\hat{\alpha}_{TL}$,

$$MSE = \left[\frac{1}{1000}\sum_{j=1}^{1000}\left(\hat{\alpha}_{TL,j}-\alpha\right)\right]^2 + \frac{1}{1000-1}\sum_{j=1}^{1000}\left(\hat{\alpha}_{TL,j}-\bar{\hat{\alpha}}_{TL}\right)^2. \tag{11}$$

Table 3 also summarizes the MSE of the unknown parameters of interest (α,σ) obtained by using either the TLNISE or the MLE-Berkey method. As expected, the efficiency of the two computational approaches was very similar. However, TLNISE was slightly more efficient at estimating σ than was MLE-Berkey (smaller MSE) for small values of σ and for moderate to large values of α. For small values of α, TLNISE was slightly worse for estimating σ. TLNISE was always better for estimating α for any combination of α and σ.

Table 3. Evaluation of NMMAPS and APHENA Computational Methods Used in Hierarchical Models for Meta-Analyses: Comparison of Mean Squared Errors of Estimated Parameters

True Values			MSE	
α	σ	Method	$\hat{\alpha}$	$\hat{\sigma}$
0.1	0.05	TLNISE	0.001803	0.064804
0.1	0.05	Berkey	0.002515	0.044224
0.1	0.1	TLNISE	0.001814	0.066649
0.1	0.1	Berkey	0.002481	0.047479
0.1	0.5	TLNISE	0.006386	0.184938
0.1	0.5	Berkey	0.006789	0.212599
0.1	1	TLNISE	0.015428	0.837419
0.1	1	Berkey	0.015731	0.805371
0.5	0.05	TLNISE	0.001664	0.02114
0.5	0.05	Berkey	0.002519	0.043033
0.5	0.1	TLNISE	0.001784	0.02025
0.5	0.1	Berkey	0.002399	0.038552
0.5	0.5	TLNISE	0.005946	0.004201
0.5	0.5	Berkey	0.00614	0.007865
0.5	1	TLNISE	0.014273	0.272957
0.5	1	Berkey	0.014558	0.25298
1.0	0.05	TLNISE	0.001624	0.416514
1.0	0.05	Berkey	0.002319	0.490227
1.0	0.1	TLNISE	0.001908	0.412235
1.0	0.1	Berkey	0.002596	0.479993
1.0	0.5	TLNISE	0.006238	0.227546
1.0	0.5	Berkey	0.006477	0.201019
1.0	1	TLNISE	0.01573	0.011163
1.0	1	Berkey	0.01618	0.009516

RESULTS

DESCRIPTION OF THE DATABASE

The databases analyzed are described in Appendix Tables D.1–D.6 (available on the Web). Table D.1 provides descriptive data on the 12 Canadian cities with mortality data for 1987 to 1996 (10 years). The populations ranged from about 100,000 to over 2 million. The cities had systematically missing PM_{10} data (measurements are for one out of every six days), but all had daily O_3 data. The median PM_{10} concentrations ranged from 11.4 μg/m^3 to 27.5 μg/m^3, and the

median O_3 concentration ranged from 6.6 μg/m^3, to 9.8 μg/m^3. The temperature ranged from 2.7° to 10.5°C (annual average), and the daily total number of deaths (all causes) ranged from 3 to 49.

Appendix Table D.2 gives the descriptive data for the European cities with mortality data. Of the 31 cities from the original database, 24 had data for PM or O_3 or both, and are therefore included in this table. These cities contributed data for three to seven years between 1990 and 1997. The data for the European cities were compiled from different countries, and the data collection system was specific to the national institutions and research groups collaborating in the APHEA2 project. The city populations ranged from slightly above 200,000 to about 7 million; however, the Netherlands (population 15 million) was considered as one urban area because of its density and urban character. The mean daily number of deaths from all causes ranged from 6 to 169 in the cities and was 347 in the Netherlands. For the cities with PM_{10} measurements, missing values were random. For 10 cities, PM_{10} concentrations were estimated using TSP or black smoke measurements. Two cities (Ljubljana and Valencia) did not have PM data in any form. The O_3 data cover the whole year for all cities except Cracow, for which they were not available. The median PM_{10} concentrations ranged from 13 μg/m^3 (Stockholm) to 65 μg/m^3 (Turin and Prague); those for O_3 ranged from 22 μg/m^3 (London) to 82 μg/m^3 (Athens).

Appendix Table D.3 describes the U.S. database, which includes the 90 cities with the largest populations in the United States and extends for 10 years from 1987 to 1996. Population size varied substantially among cities (from about 250,000 to above 9 million) and, consequently, so did the daily number of deaths, which ranged from 5 to 198. For some cities, the average daily number of deaths for less common causes was close to zero. Of the 90 cities, 15 had daily PM measurements. All had daily O_3 measurements, but 36 of those had O_3 measurements only in the summer. The median PM_{10} concentrations ranged from 14 μg/m^3 (Anchorage) to 43 μg/m^3 (Fresno), and the median O_3 concentrations ranged from 26 μg/m^3 (Honolulu) to 75.3 μg/m^3 (Bakersfield).

Appendix Table D.4 provides descriptive information concerning the hospital admissions data from Canada. The cities included in this database are the same as those included in the Canadian mortality data set, but the time period covered is shorter: three years per city, with PM_{10} measured once every six days. Only four cities had populations near or greater than 1 million. For people 65 years or older, the mean daily number admitted for cardiovascular disease ranged from 5 to 50, while those admitted for respiratory disease ranged from 2 to 19.

Appendix Table D.5 describes the European hospital admissions database, which includes eight cities or areas, each with a population greater than 1 million. The mean daily number of people 65 years or older admitted for cardiovascular disease ranged from 11 to 81, and the admissions for respiratory disease ranged from 5 to 58. The air pollution data series for most cities were randomly missing fewer than 10% of their observations. The exceptions were the O_3 series for Birmingham (17% missing) and the O_3 series for Rome (15% missing). The PM_{10} series is available for Amsterdam only since 1992 and for Stockholm since 1994.

Appendix Table D.6 gives descriptive characteristics of the hospital admissions data for the United States from 1985 through 1994. The data comprise 14 cities, with 5 cities having populations over 1 million. The mean daily number of people 65 years or older admitted for cardiovascular disease ranged from 2 to 102; admissions for respiratory disease ranged from 1 to 53. For PM_{10}, the missing values occurred randomly but quite frequently in some cities (20%–30%) (Samet 2000b). For O_3, seven cities had measurements for the summer only; O_3 data were not available for Minneapolis.

SUMMARY ANALYSIS OF DATABASES

In Figure 2A, box plots for mortality outcomes for people 75 years or older and for people younger than 75 years in Canada, Europe, and the United States show that the median city-specific counts for daily deaths are higher in the European data, reflecting the larger city populations. The same is true for hospital admissions of people with respiratory or cardiovascular disease for people 65 years or older and people younger than 65 years (Figure 2B).

Box plots of pollution concentrations for the days included in the mortality data set are shown in Figure 3A. Both PM_{10} and O_3 concentrations are higher in the European cities, followed by those in the United States. In Canada, PM_{10} concentrations are comparable to those in the United States, but O_3 levels are much lower. Figure 3B shows the pollution data for the admissions databases. For Europe, these databases comprise a subset of the cities in the mortality database, while for the United States the admissions data include several cities not in the mortality database. However, the majority of cities in the admissions database are a subset of those in the mortality database. The PM_{10} levels in Europe and the United States are similar; Canadian levels are slightly lower. O_3 levels are highest for U.S. cities, followed by European cities, and finally Canadian cities with much lower levels of O_3.

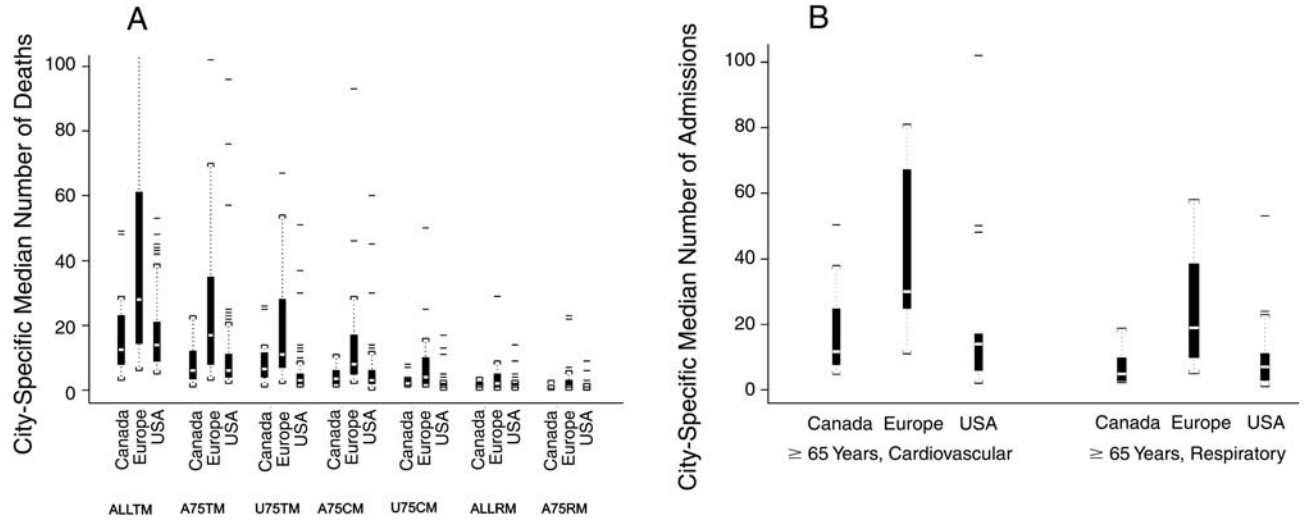


Figure 2. Canada, Europe and the United States A: city-specific median mortality counts (**ALLTM** indicates all ages, all-cause mortality; **A75TM** indicates ≥ 75 years, all-cause mortality; **U75TM** indicates < 75 years, all-cause mortality; **A75CM** indicates ≥ 75 years, cardiovascular mortality; **U75CM** indicates < 75 years, cardiovascular mortality; **ALLRM** indicates all ages, respiratory mortality; **A75RM** indicates ≥ 75 years, respiratory mortality.) **B**: city-specific median counts for number of people ≥ 65 years admitted to the hospital with cardiovascular or respiratory disease.

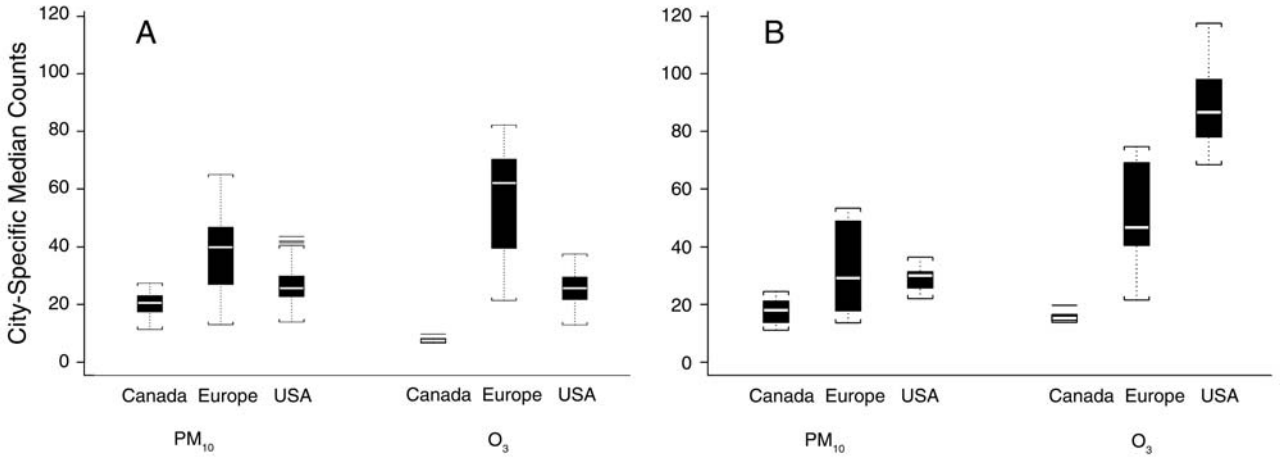


Figure 3. Canada, Europe, and the United States: 24-hour average PM_{10} and O_3 maximum 1-hour daily pollution concentrations for the A: mortality database; **B**: admissions database.

Figure 4 box plots show the correlation between PM_{10} and O_3 for each of the three centers by season. Summer median correlations are positive for all centers and range from 0.27 to 0.40. In the winter season the median correlations of PM_{10} and O_3 are around zero or negative.

EFFECTS OF PM_{10} AND O_3 ON MORTALITY

This section provides an overview of the outcome of the analyses conducted to evaluate the sensitivity of the effect estimates for PM_{10} and O_3 to various model specifications summarized earlier in Table 2. Investigators from each of the three research centers fit the agreed upon first-stage model to individual cities in their regions, then fit second-stage models across cities within each region. Metaregressions were then conducted to pool results across regions; results were combined only for Europe and the United States, given incompatibilities with the Canadian data sets.

We first discuss the individual region results for mortality and for hospital admissions. We then present a comparison of the three regions and an evaluation of the combined results.

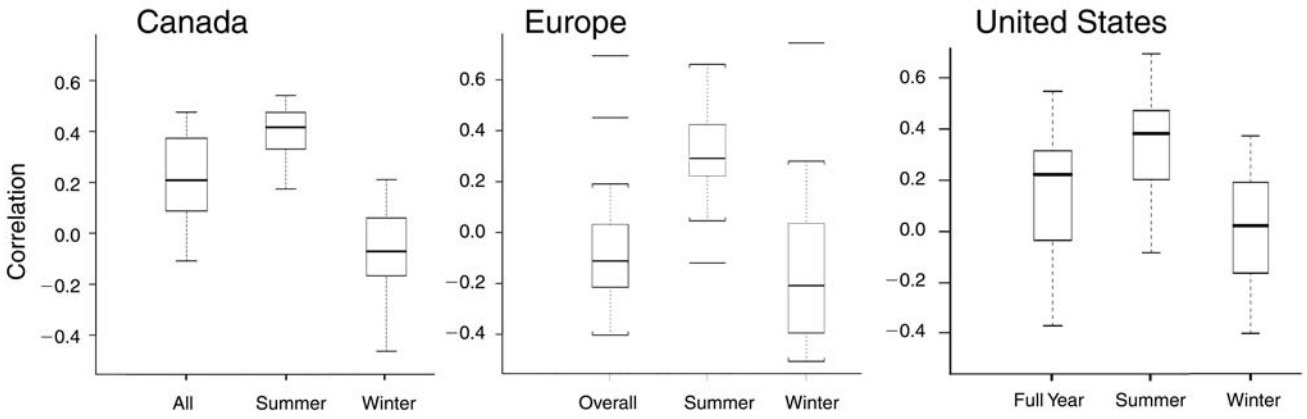


Figure 4. Correlations between 24-hour average PM_{10} and O_3 maximum 1-hour daily pollution concentrations by season for Canada, Europe, and the United States.

Table 4. Canada: Percentage Change in All-Cause Mortality per 10-μg/m^3 Increase in PM_{10}, Lag 1

Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} Results		
All ages		
3 *df*/year	1.2 (0.68, 1.7)	1.1 (0.58, 1.6)
8 *df*/year	0.86 (0.32, 1.4)	0.84 (0.3, 1.4)
12 *df*/year	0.8 (0.24, 1.4)	0.82 (0.24, 1.4)
PACF	1.1 (0.64, 1.6)	1.1 (0.58, 1.6)
≥ 75 Years		
3 *df*/year	1.6 (0.88, 2.3)	1.4 (0.69, 2.1)
8 *df*/year	1.1 (0.35, 1.9)	1 (0.25, 1.8)
12 *df*/year	1 (0.26, 1.8)	1.1 (0.3, 1.9)
PACF	1.5 (0.82, 2.2)	1.4 (0.69, 2.1)
< 75 Years		
3 *df*/year	0.76 (0.09, 1.4)	0.74 (0.055, 1.4)
8 *df*/year	0.58 (−0.16, 1.3)	0.63 (−0.12, 1.4)
12 *df*/year	0.51 (−0.27, 1.3)	0.5 (−0.31, 1.3)
PACF	0.74 (0.06, 1.4)	0.74 (0.055, 1.4)
Controlling for O_3		
All ages		
3 *df*/year	1 (0.51, 1.5)	0.92 (0.4, 1.4)
8 *df*/year	0.74 (0.19, 1.3)	0.76 (0.2, 1.3)
12 *df*/year	0.69 (0.12, 1.3)	0.75 (0.15, 1.4)
PACF	1 (0.5, 1.5)	0.92 (0.4, 1.4)
≥ 75 Years		
3 *df*/year	1.5 (0.76, 2.2)	1.3 (0.59, 2)
8 *df*/year	1 (0.25, 1.8)	0.98 (0.18, 1.8)
12 *df*/year	0.98 (0.17, 1.8)	1.1 (0.23, 1.9)
PACF	1.4 (0.73, 2.2)	1.3 (0.59, 2)
< 75 Years		
3 *df*/year	0.54 (−0.15, 1.2)	0.5 (−0.2, 1.2)
8 *df*/year	0.41 (−0.35, 1.2)	0.51 (−0.26, 1.3)
12 *df*/year	0.36 (−0.44, 1.2)	0.39 (−0.44, 1.2)
PACF	0.55 (−0.14, 1.2)	0.5 (−0.2, 1.2)

Canada

PM As noted in the section describing the database, all Canadian cities had PM_{10} measurements for one out of six days. Therefore, in accordance with the protocol, only PM_{10} exposure with lag 1 could be used in the models; the corresponding results are reported here.

Table 4 summarizes the effects of a 10-μg/m^3 PM_{10}, lag 1 increase on the total daily number of deaths (all-cause mortality) for the three age groups (all ages, ≥ 75 years, and < 75 years) using the various models. For all ages, each of the models estimated effects that were statistically significant at the 95% level; the estimates from models using PS and NS were similar in size. Thus, mortality increased by 0.86% (95% CI; 0.3 to 1.4) when associated with a 10-μg/m^3 PM_{10} increase with PS models and 8 *df* and by 0.84% (0.3 to 1.4) with NS models and 8 *df*. The estimates were slightly lower with adjustment for O_3. As will be discussed in the section on combined results, the effects were larger (about double) in Canada compared with estimates for Europe and the United States.

For people 75 years or older and people younger than 75 years, the investigators observed the same pattern in these Canadian results as in the European results; the effect estimate for all-cause mortality was consistently larger for the older age group than for the younger age group. An increase of 10 μg/m^3 in PM_{10}, lag 1, was associated with an increase of 1.1% (0.35 to 1.9) in the daily number of deaths among the older age group (model with 8 *df*, PS) and with a 0.58% (−0.16 to 1.3) increase in mortality among the younger age group. In the older age group, the PM_{10} effects remained unchanged after adjusting for O_3; in the younger age group, the estimated effects were slightly lower.

Table 5. Canada: Percentage Change in Cardiovascular Mortality per 10-μg/m^3 Increase in PM_{10}, Lag 1

Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} Results		
≥ 75 Years		
3 *df*/year	1.8 (0.81, 2.8)	1.6 (0.56, 2.6)
8 *df*/year	1.4 (0.29, 2.5)	1.3 (0.19, 2.4)
12 *df*/year	1.4 (0.23, 2.6)	1.5 (0.34, 2.8)
PACF	1.8 (0.75, 2.8)	1.6 (0.56, 2.6)
< 75 Years		
3 *df*/year	0.41 (−0.76, 1.6)	0.43 (−0.77, 1.6)
8 *df*/year	0.045 (−1.2, 1.3)	0.07 (−1.3, 1.4)
12 *df*/year	−0.18 (−1.5, 1.2)	−0.026 (−1.5, 1.4)
PACF	0.43 (−0.74, 1.6)	0.43 (−0.77, 1.6)
Controlling for O_3		
≥ 75 Years		
3 *df*/year	1.7 (0.64, 2.7)	1.5 (0.41, 2.5)
8 *df*/year	1.2 (0.098, 2.4)	1.1 (−0.014, 2.3)
12 *df*/year	1.2 (0.048, 2.5)	1.4 (0.16, 2.6)
PACF	1.6 (0.61, 2.7)	1.5 (0.41, 2.5)
< 75 Years		
3 *df*/year	0.33 (−0.88, 1.6)	0.3 (−0.93, 1.5)
8 *df*/year	0.013 (−1.3, 1.4)	0.039 (−1.3, 1.4)
12 *df*/year	−0.2 (−1.6, 1.2)	−0.036 (−1.5, 1.5)
PACF	0.33 (−0.89, 1.6)	0.3 (−0.93, 1.5)

Table 6. Canada: Percentage Change in Respiratory Mortality per 10-μg/m^3 Increase in PM_{10}, Lag 1

Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} Results		
All ages		
3 *df*/year	−0.43 (−2.1, 1.2)	−0.62 (−2.3, 1.1)
8 *df*/year	−1.4 (−3.1, 0.46)	−1.4 (−3.2, 0.5)
12 *df*/year	−1.4 (−3.2, 0.55)	−1.3 (−3.3, 0.69)
PACF	−0.36 (−2, 1.3)	−0.62 (−2.3, 1.1)
≥ 75 Years		
3 *df*/year	0.27 (−1.8, 2.3)	0.14 (−1.9, 2.3)
8 *df*/year	−0.7 (−2.9, 1.6)	−0.74 (−3, 1.6)
12 *df*/year	−0.69 (−3, 1.7)	−0.41 (−2.9, 2.1)
PACF	0.4 (−1.6, 2.5)	0.14 (−1.9, 2.3)
Controlling for O_3		
All ages		
3 *df*/year	−1.4 (−4, 1.3)	−0.76 (−2.5, 1)
8 *df*/year	−2.4 (−5.2, 0.57)	−2.4 (−5.3, 0.71)
12 *df*/year	−1.4 (−3.3, 0.6)	−1.4 (−3.4, 0.73)
PACF	−0.46 (−2.2, 1.3)	−0.76 (−2.5, 1)
≥ 75 Years		
3 *df*/year	0.2 (−1.9, 2.4)	0.048 (−2.1, 2.2)
8 *df*/year	−2.2 (−6.2, 2.1)	−2.2 (−6.4, 2.1)
12 *df*/year	−2.1 (−6.3, 2.2)	−0.32 (−2.8, 2.3)
PACF	0.19 (−2, 2.4)	0.048 (−2.1, 2.2)

Table 5 shows the PM_{10} effect estimates for cardiovascular mortality for people 75 years or older and for people younger than 75 years, for all models. Statistically significant estimates, comparable in size to those observed for all-cause mortality, were found for the older age group. These estimates were not changed after adjusting for O_3. In contrast, no effect on cardiovascular mortality was found for the older age group.

Table 6 shows the estimated PM_{10} effects on respiratory mortality for all ages and for people 75 years or older. No significant effect was found for all ages or for the older age group using any model.

O_3 Canadian cities had daily O_3 measurements, so models were run for all lags (1, 0–1 [the average of lags 0 and 1], and DL for the cumulative effects of lags 0, 1, & 2).

Table 7 provides the results for O_3 effects on all-cause mortality for three age groups (all ages, ≥ 75, and < 75) for all models. The estimated effects were positive and significant for all age groups, except for DL models for the younger age group. Effects using DL were more similar to those using lag 0–1 for the two older age groups. The estimated effects using lag 0–1 were higher than those of lag 1. Increasing the degrees of freedom led to a decrease in the effect estimates in all age groups. Controlling for PM_{10}, lag 1, had a somewhat inconsistent influence on the effect estimates, generally lowering them in the older age group, but in some models increasing them in the younger age group. The annual effects estimates shown here were similar for both age groups to the summer effects shown in Table 8.

Table 9 shows the O_3 effects on cardiovascular mortality for people 75 years or older and the corresponding results for people younger than 75 years. The effects were positive and statistically significant for the older age group and positive but not statistically significant for the younger age group. Adjusting for PM_{10} reduced the estimates.

Table 10 displays the O_3 effects on respiratory mortality for all ages and for people 75 years or older. No statistically significant effects were found.

Figures 5 and 6 show the results from the more extensive sensitivity analyses done to explore the influence of using a broader range of degrees of freedom (from 2 to 20 *df*) on all-cause mortality and respiratory mortality for people 75 years or older from exposure to PM_{10} and O_3. The

Table 7. Canada: Percentage Change in All-Cause Mortality per 10-μg/m^3 Increase in O_3

Seasonality Control	Average of Lags 0–1		Lag 1		Distributed Lags	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 Results						
All ages						
3 *df*/year	1 (0.73, 1.3)	1.1 (0.74, 1.4)	0.73 (0.48, 0.99)	0.74 (0.47, 1)	0.84 (0.35, 1.3)	1 (0.52, 1.5)
8 *df*/year	0.85 (0.51, 1.2)	0.81 (0.47, 1.2)	0.56 (0.28, 0.83)	0.52 (0.24, 0.8)	0.75 (0.25, 1.3)	0.73 (0.23, 1.2)
12 *df*/year	0.78 (0.44, 1.1)	0.75 (0.4, 1.1)	0.5 (0.22, 0.78)	0.48 (0.2, 0.77)	0.68 (0.18, 1.2)	0.66 (0.15, 1.2)
PACF	0.97 (0.67, 1.3)	1 (0.69, 1.3)	0.68 (0.43, 0.94)	0.7 (0.43, 0.96)	0.79 (0.3, 1.3)	1 (0.53, 1.5)
≥ 75 Years						
3 *df*/year	1.3 (0.85, 1.7)	1.3 (0.87, 1.8)	0.87 (0.51, 1.2)	0.88 (0.5, 1.3)	1.1 (0.37, 1.8)	1.3 (0.61, 2)
8 *df*/year	0.98 (0.5, 1.5)	0.9 (0.41, 1.4)	0.61 (0.22, 1)	0.54 (0.14, 0.94)	0.9 (0.19, 1.6)	0.83 (0.11, 1.6)
12 *df*/year	0.88 (0.39, 1.4)	0.78 (0.28, 1.3)	0.53 (0.13, 0.93)	0.46 (0.059, 0.87)	0.79 (0.076, 1.5)	0.68 (−0.043, 1.4)
PACF	1.2 (0.73, 1.6)	1.3 (0.81, 1.7)	0.79 (0.42, 1.2)	0.84 (0.46, 1.2)	1 (0.3, 1.7)	1.3 (0.58, 2)
< 75 Years						
3 *df*/year	0.82 (0.42, 1.2)	0.84 (0.41, 1.3)	0.62 (0.27, 0.96)	0.62 (0.27, 0.98)	0.64 (−0.026, 1.3)	0.76 (0.084, 1.4)
8 *df*/year	0.74 (0.28, 1.2)	0.73 (0.26, 1.2)	0.52 (0.14, 0.89)	0.51 (0.13, 0.89)	0.62 (−0.062, 1.3)	0.65 (−0.043, 1.3)
12 *df*/year	0.7 (0.23, 1.2)	0.73 (0.25, 1.2)	0.48 (0.1, 0.87)	0.5 (0.12, 0.89)	0.62 (−0.067, 1.3)	0.64 (−0.056, 1.3)
PACF	0.82 (0.41, 1.2)	0.84 (0.41, 1.3)	0.61 (0.26, 0.95)	0.62 (0.27, 0.98)	0.63 (−0.034, 1.3)	0.76 (0.084, 1.4)
Controlling for PM_{10}						
All ages						
3 *df*/year			0.67 (0.058, 1.3)	0.76 (0.14, 1.4)		
8 *df*/year			0.48 (−0.18, 1.2)	0.4 (−0.28, 1.1)		
12 *df*/year			0.45 (−0.23, 1.1)	0.27 (−0.44, 0.99)		
PACF			0.67 (0.052, 1.3)	0.76 (0.14, 1.4)		
≥ 75 Years						
3 *df*/year			0.12 (−0.73, 0.99)	0.22 (−0.65, 1.1)		
8 *df*/year			−0.067 (−0.99, 0.87)	−0.13 (−1.1, 0.84)		
12 *df*/year			−0.06 (−1, 0.91)	−0.21 (−1.2, 0.8)		
PACF			0.13 (−0.73, 0.99)	0.22 (−0.65, 1.1)		
< 75 Years						
3 *df*/year			1.2 (0.38, 2.1)	1.3 (0.47, 2.2)		
8 *df*/year			1 (0.11, 2)	0.96 (−0.0018, 1.9)		
12 *df*/year			0.96 (−0.0037, 1.9)	0.75 (−0.25, 1.8)		
PACF			1.2 (0.37, 2.1)	1.3 (0.47, 2.2)		

Table 8. Canada: Percentage Change in Mortality per 10-μg/m^3 Increase in O_3—Summer-Only Analysis[a]

	Average of Lags 0–1 % (95% CI)	Lag 1 % (95% CI)	Distributed Lags % (95% CI)
All-cause mortality			
All ages	0.95 (0.65, 1.20)	0.71 (0.46, 0.95)	0.42 (0.16, 0.67)
Ages ≥ 75	1.2 (0.51, 2.0)	0.88 (0.35, 1.4)	0.82 (0.25, 1.4)
Ages < 75	0.60 (0.20, 1.00)	0.45 (0.12, 0.78)	0.42 (0.16, 0.67)
Cardiovascular mortality			
Ages ≥ 75	0.76 (−0.22, 1.8)	0.51 (−0.36, 1.4)	0.19 (−0.36, 0.74)
Ages < 75	−0.26 (−0.97, 0.44)	−0.23 (−0.81, 0.35)	−0.13 (−0.55, 0.29)
Respiratory mortality			
All ages	2.8 (1.8, 3.8)	2.2 (1.3, 3.0)	3.0 (1.6, 4.5)
Ages ≥ 75	2.4 (1.1, 3.7)	2 (0.93, 3)	2.3 (0.28, 4.4)

[a] O_3 not controlled for PM_{10}.

Table 9. Canada: Percentage Change in Cardiovascular Mortality per 10-μg/m^3 Increase in O_3

Seasonality Control	Average of Lags 0–1		Lag 1		Distributed Lags	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 Results						
≥ 75 Years						
3 *df*/year	1.4 (0.77, 2)	1.4 (0.81, 2.1)	1 (0.49, 1.5)	1 (0.5, 1.6)	1.3 (0.28, 2.3)	1.6 (0.55, 2.6)
8 *df*/year	1.1 (0.46, 1.8)	1 (0.34, 1.8)	0.79 (0.22, 1.4)	0.7 (0.12, 1.3)	1.2 (0.18, 2.3)	1.1 (0.095, 2.2)
12 *df*/year	1.1 (0.34, 1.8)	0.96 (0.24, 1.7)	0.71 (0.13, 1.3)	0.65 (0.066, 1.2)	1.1 (0.053, 2.1)	1 (0.0029, 2.1)
PACF	1.4 (0.77, 2)	1.4 (0.81, 2.1)	1 (0.48, 1.5)	1 (0.5, 1.6)	1.3 (0.28, 2.3)	1.6 (0.55, 2.6)
< 75 Years						
3 *df*/year	0.71 (−0.007, 1.4)	0.68 (−0.069, 1.4)	0.44 (−0.16, 1)	0.41 (−0.22, 1)	0.91 (−0.26, 2.1)	1.1 (−0.092, 2.3)
8 *df*/year	0.43 (−0.38, 1.2)	0.37 (−0.45, 1.2)	0.18 (−0.48, 0.85)	0.14 (−0.53, 0.82)	0.89 (−0.31, 2.1)	0.87 (−0.35, 2.1)
12 *df*/year	0.3 (−0.54, 1.1)	0.35 (−0.49, 1.2)	0.081 (−0.59, 0.76)	0.12 (−0.57, 0.81)	0.79 (−0.43, 2)	0.88 (−0.34, 2.1)
PACF	0.7 (−0.015, 1.4)	0.68 (−0.069, 1.4)	0.44 (−0.17, 1)	0.41 (−0.22, 1)	0.92 (−0.26, 2.1)	1.1 (−0.092, 2.3)
Controlling for PM_{10}						
≥ 75 Years						
3 *df*/year			0.15 (−1.1, 1.4)	0.31 (−0.96, 1.6)		
8 *df*/year			0.15 (−1.2, 1.5)	0.24 (−1.2, 1.7)		
12 *df*/year			0.27 (−1.2, 1.7)	0.31 (−1.2, 1.8)		
PACF			0.15 (−1.1, 1.4)	0.31 (−0.96, 1.6)		
< 75 Years						
3 *df*/year			0.2 (−1.3, 1.7)	0.57 (−0.96, 2.1)		
8 *df*/year			−0.14 (−1.8, 1.5)	−0.34 (−2, 1.4)		
12 *df*/year			−0.27 (−2, 1.4)	−0.6 (−2.4, 1.2)		
PACF			0.072 (−1.4, 1.6)	0.57 (−0.96, 2.1)		

Table 10. Canada: Percentage Change in Respiratory Mortality per 10-μg/m^3 Increase in O_3

Seasonality Control	Average of Lags 0–1 % (95% CI)		Lag 1 % (95% CI)		Distributed Lags % (95% CI)	
	Penalized Splines	Natural Splines	Penalized Splines	Natural Splines	Penalized Splines	Natural Splines
O_3 Results						
All ages						
3 *df*/year	0.86 (−0.15, 1.9)	−0.026 (−1.1, 1)	0.52 (−0.34, 1.4)	−0.12 (−1, 0.78)	0.76 (−0.91, 2.5)	−0.012 (−1.7, 1.7)
8 *df*/year	0.1 (−1, 1.2)	0.15 (−1, 1.3)	0.067 (−0.86, 1)	0.11 (−0.84, 1.1)	0.04 (−1.6, 1.8)	0.13 (−1.6, 1.9)
12 *df*/year	0.054 (−1.1, 1.2)	0.016 (−1.2, 1.2)	0.051 (−0.9, 1)	0.037 (−0.92, 1)	0.0051 (−1.7, 1.7)	−0.053 (−1.8, 1.7)
PACF	0.44 (−0.61, 1.5)	0.33 (−0.76, 1.4)	0.25 (−0.63, 1.1)	0.21 (−0.7, 1.1)	0.3 (−1.4, 2)	0.15 (−1.5, 1.9)
≥ 75 Years						
3 *df*/year	0.66 (−0.59, 1.9)	−0.17 (−1.5, 1.2)	0.47 (−0.59, 1.5)	−0.11 (−1.2, 1)	0.29 (−1.8, 2.4)	−0.39 (−2.5, 1.8)
8 *df*/year	−0.29 (−1.7, 1.1)	−0.33 (−1.8, 1.1)	−0.068 (−1.2, 1.1)	−0.085 (−1.3, 1.1)	−0.62 (−2.7, 1.5)	−0.6 (−2.7, 1.6)
12 *df*/year	−0.44 (−1.9, 1)	−0.6 (−2.1, 0.89)	−0.14 (−1.3, 1.1)	−0.24 (−1.4, 0.98)	−0.78 (−2.9, 1.4)	−0.95 (−3.1, 1.2)
PACF	0.32 (−0.96, 1.6)	−0.23 (−1.6, 1.1)	0.23 (−0.84, 1.3)	−0.097 (−1.2, 1)	−0.14 (−2.2, 2)	−0.25 (−2.4, 1.9)
Controlling for PM_{10}						
All ages						
3 *df*/year			1.2 (−0.9, 3.4)	1.3 (−0.86, 3.5)		
8 *df*/year			2.5 (−1.2, 6.3)	2.6 (−1.7, 7.1)		
12 *df*/year			2.7 (−1.8, 7.5)	2.6 (−2.4, 7.9)		
PACF			1.2 (−0.92, 3.5)	1.3 (−0.86, 3.5)		
≥ 75 Years						
3 *df*/year			0.25 (−2.4, 2.9)	0.65 (−2, 3.4)		
8 *df*/year			0.0096 (−2.8, 2.9)	0.16 (−2.8, 3.2)		
12 *df*/year			−0.079 (−3, 2.9)	−0.28 (−3.4, 2.9)		
PACF			0.14 (−2.5, 2.9)	0.65 (−2, 3.4)		

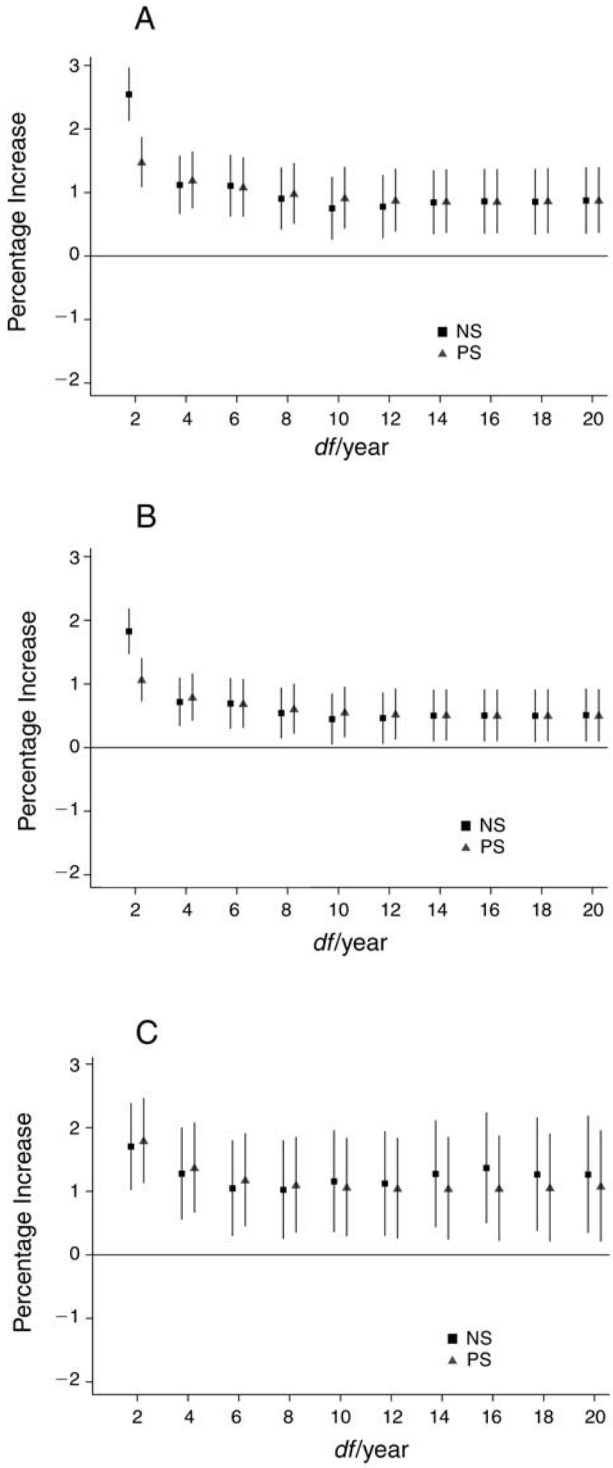


Figure 5. Canada (≥ 75 years): Extended sensitivity analysis of impact of lag and *df*/year on percentage change in all-cause mortality associated with a 10-µg/m^3 increase in A: O_3, lag 0–1; **B**: O_3, lag 1; **C**: PM_{10}, lag 1.

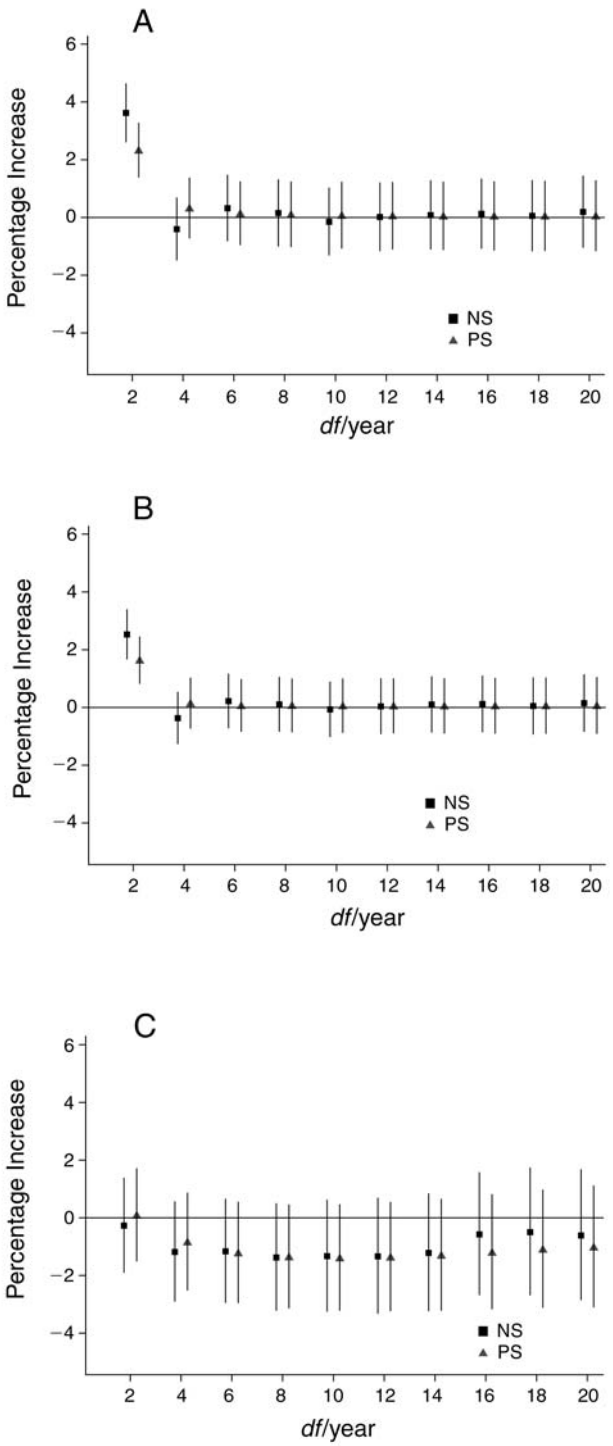


Figure 6. Canada (all ages): Extended sensitivity analysis of impact of lag and *df*/year on percentage change in respiratory mortality associated with a 10-µg/m^3 increase in A: O_3, lag 0–1; **B**: O_3, lag 1; **C**: PM_{10}, lag 1.

Table 11. Europe: Percentage Change in All-Cause Mortality per 10-μg/m^3 Increase in PM_{10}

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} Results						
All ages						
3 *df*/year	0.62 (0.39, 0.85)	0.31 (0.13, 0.49)	0.65 (0.46, 0.83)	0.36 (0.21, 0.51)	0.50 (0.27, 0.73)	0.21 (−0.01, 0.44)
8 *df*/year	0.29 (0.14, 0.45)	0.20 (0.05, 0.35)	0.33 (0.22, 0.44)	0.26 (0.15, 0.36)	0.20 (−0.01, 0.42)	0.10 (−0.12, 0.32)
12 *df*/year	0.27 (0.12, 0.43)	0.23 (0.08, 0.38)	0.30 (0.20, 0.41)	0.28 (0.18, 0.39)	0.19 (−0.02, 0.40)	0.16 (−0.05, 0.37)
PACF	0.41 (0.24, 0.59)	0.27 (0.12, 0.43)	0.45 (0.29, 0.61)	0.32 (0.20, 0.44)	0.30 (0.09, 0.52)	0.15 (−0.07, 0.36)
≥ 75 Years						
3 *df*/year	0.84 (0.54, 1.14)	0.41 (0.17, 0.66)	0.88 (0.63, 1.12)	0.48 (0.28, 0.68)	0.68 (0.39, 0.98)	0.29 (−0.005, 0.59)
8 *df*/year	0.39 (0.19, 0.59)	0.28 (0.08, 0.47)	0.44 (0.29, 0.58)	0.34 (0.20, 0.48)	0.32 (0.04, 0.60)	0.20 (−0.08, 0.49)
12 *df*/year	0.36 (0.16, 0.56)	0.30 (0.10, 0.51)	0.40 (0.26, 0.53)	0.37 (0.23, 0.51)	0.31 (0.03, 0.58)	0.28 (−0.002, 0.56)
PACF	0.56 (0.34, 0.79)	0.42 (0.18, 0.65)	0.58 (0.40, 0.75)	0.40 (0.27, 0.53)	0.48 (0.20, 0.76)	0.37 (0.09, 0.65)
< 75 Years						
3 *df*/year	0.48 (0.33, 0.64)	0.27 (0.12, 0.43)	0.47 (0.32, 0.62)	0.28 (0.13, 0.42)	0.33 (0.01, 0.65)	0.18 (−0.15, 0.50)
8 *df*/year	0.25 (0.09, 0.42)	0.14 (−0.03, 0.31)	0.25 (0.10, 0.40)	0.17 (0.01, 0.32)	0.11 (−0.20, 0.43)	0.03 (−0.29, 0.35)
12 *df*/year	0.25 (0.08, 0.42)	0.18 (0.00, 0.36)	0.24 (0.09, 0.39)	0.20 (0.04, 0.35)	0.11 (−0.21, 0.43)	0.07 (−0.25, 0.39)
PACF	0.39 (0.20, 0.57)	0.21 (0.05, 0.38)	0.38 (0.22, 0.54)	0.21 (0.06, 0.36)	0.25 (−0.07, 0.57)	0.11 (−0.21, 0.43)
Controlling for O_3						
All ages						
3 *df*/year	0.59 (0.34, 0.84)	0.29 (0.11, 0.48)	0.62 (0.43, 0.82)	0.34 (0.20, 0.49)		
8 *df*/year	0.29 (0.13, 0.45)	0.18 (0.05, 0.32)	0.32 (0.21, 0.42)	0.25 (0.14, 0.36)		
12 *df*/year	0.26 (0.10, 0.42)	0.21 (0.05, 0.37)	0.30 (0.19, 0.40)	0.27 (0.16, 0.38)		
PACF	0.42 (0.23, 0.62)	0.28 (0.12, 0.43)	0.45 (0.29, 0.62)	0.30 (0.18, 0.43)		
≥ 75 Years						
3 *df*/year	0.81 (0.49, 1.12)	0.39 (0.14, 0.64)	0.83 (0.59, 1.08)	0.44 (0.25, 0.64)		
8 *df*/year	0.39 (0.20, 0.58)	0.26 (0.09, 0.44)	0.41 (0.27, 0.54)	0.32 (0.18, 0.47)		
12 *df*/year	0.35 (0.16, 0.55)	0.30 (0.09, 0.50)	0.38 (0.24, 0.52)	0.35 (0.21, 0.49)		
PACF	0.56 (0.34, 0.79)	0.37 (0.19, 0.55)	0.54 (0.38, 0.70)	0.36 (0.22, 0.50)		
< 75 Years						
3 *df*/year	0.44 (0.26, 0.62)	0.26 (0.09, 0.42)	0.43 (0.28, 0.58)	0.28 (0.13, 0.43)		
8 *df*/year	0.21 (0.03, 0.38)	0.12 (−0.06, 0.30)	0.23 (0.07, 0.39)	0.18 (0.01, 0.34)		
12 *df*/year	0.20 (0.02, 0.38)	0.16 (−0.03, 0.34)	0.22 (0.06, 0.38)	0.20 (0.03, 0.36)		
PACF	0.38 (0.17, 0.59)	0.16 (−0.02, 0.35)	0.34 (0.18, 0.49)	0.21 (0.06, 0.37)		

estimated effects for all-cause mortality were highest at 2 *df*, but decreased and stabilized for PM_{10} models at 6 *df* and for the O_3 models at closer to 8–10 *df* for both lag 0–1 and lag 1 models; the results were similar for PS and NS. Increasing the number of degrees of freedom generally had little effect on estimates of respiratory mortality for either PM_{10} or O_3 (except at 2 *df*). All models with > 4 *df* had similar results.

The analysis of summer-only O_3 effects on all-cause, cardiovascular, and respiratory mortality are shown in Table 8 for the same age groups as in the full-year analyses presented above. While estimates were similar for all-cause cardiovascular mortality in all-year and summer-only models, the estimates for respiratory mortality in the summer were much higher.

Europe

PM Table 11 summarizes the effects of PM_{10}, lags 1 and 0–1 and DL (covering lags 0–2), on all-cause mortality for all three age groups, and for the different models applied. The effects were larger when applying 3 *df* for seasonality control, especially with PS, while the other models gave comparable estimates. All effects estimated for the all-ages group with lag 0–1 and lag 1 exposures were positive and statistically significant; those estimated with DL models were somewhat lower than for lag 0–1 and many were not statistically significant. The overall effect for all ages at 8 *df* using PS is 0.29% (95% CI; 0.14 to 0.45) increase per 10-μg/m^3 increase in PM_{10} for lag 0–1 and 0.33% (0.22 to

Table 12. Europe: Percentage Change in Cardiovascular Mortality per 10-μg/m^3 Increase in PM_{10}

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} Results						
≥ 75 Years						
3 *df*/year	1.06 (0.67, 1.45)	0.63 (0.30, 0.96)	1.02 (0.70, 1.35)	0.62 (0.34, 0.90)	0.79 (0.40, 1.19)	0.43 (0.03, 0.83)
8 *df*/year	0.48 (0.20, 0.76)	0.31 (0.04, 0.59)	0.47 (0.23, 0.70)	0.32 (0.10, 0.54)	0.32 (−0.07, 0.71)	0.16 (−0.23, 0.56)
12 *df*/year	0.41 (0.13, 0.70)	0.34 (0.06, 0.62)	0.40 (0.16, 0.64)	0.34 (0.11, 0.56)	0.26 (−0.13, 0.65)	0.23 (−0.16, 0.63)
PACF	0.76 (0.42, 1.11)	0.38 (0.06, 0.70)	0.74 (0.45, 1.03)	0.42 (0.18, 0.67)	0.55 (0.16, 0.95)	0.30 (−0.10, 0.69)
< 75 Years						
3 *df*/year	0.55 (0.28, 0.81)	0.34 (0.07, 0.60)	0.53 (0.28, 0.78)	0.34 (0.10, 0.59)	0.40 (−0.13, 0.94)	0.20 (−0.33, 0.74)
8 *df*/year	0.21 (−0.07, 0.49)	0.13 (−0.16, 0.42)	0.23 (−0.02, 0.48)	0.17 (−0.09, 0.43)	0.08 (−0.45, 0.61)	0.00 (−0.54, 0.54)
12 *df*/year	0.20 (−0.09, 0.49)	0.16 (−0.14, 0.46)	0.22 (−0.04, 0.48)	0.19 (−0.08, 0.45)	0.08 (−0.46, 0.62)	0.01 (−0.53, 0.55)
PACF	0.51 (0.25, 0.78)	0.21 (−0.06, 0.48)	0.49 (0.25, 0.73)	0.29 (0.04, 0.53)	0.38 (−0.16, 0.91)	0.10 (−0.44, 0.64)
Controlling for O_3						
≥ 75 Years						
3 *df*/year	1.02 (0.62, 1.41)	0.60 (0.27, 0.94)	0.97 (0.64, 1.30)	0.58 (0.30, 0.87)		
8 *df*/year	0.48 (0.20, 0.75)	0.31 (0.05, 0.58)	0.44 (0.20, 0.67)	0.29 (0.08, 0.51)		
12 *df*/year	0.40 (0.12, 0.68)	0.33 (0.06, 0.59)	0.36 (0.12, 0.59)	0.28 (0.08, 0.49)		
PACF	0.74 (0.41, 1.08)	0.40 (0.12, 0.69)	0.70 (0.43, 0.97)	0.39 (0.16, 0.63)		
< 75 Years						
3 *df*/year	0.49 (0.19, 0.80)	0.31 (0.02, 0.60)	0.52 (0.24, 0.79)	0.34 (0.08, 0.59)		
8 *df*/year	0.17 (−0.12, 0.46)	0.10 (−0.20, 0.41)	0.23 (−0.04, 0.49)	0.17 (−0.10, 0.45)		
12 *df*/year	0.16 (−0.15, 0.46)	0.13 (−0.18, 0.45)	0.21 (−0.06, 0.49)	0.18 (−0.10, 0.47)		
PACF	0.49 (0.19, 0.78)	0.24 (−0.06, 0.54)	0.49 (0.24, 0.74)	0.29 (0.01, 0.57)		

0.44) for lag 1. The corresponding figures for NS are 0.20% (0.05 to 0.35) and 0.26% (0.15 to 0.36). When using the minimization of the PACF criterion for choosing the degrees of freedom, the estimates are 0.41% (0.24 to 0.59) increase in the number of deaths for PS, lag 0–1, and 0.45% (0.29 to 0.61) for lag 1. For NS they are 0.27% (0.12 to 0.43) and 0.32% (0.20 to 0.44), respectively.

Table 11 also allows comparisons between the effect estimates for the three age groups. The effect estimates of mortality in the ≥ 75 age group, although similar to those for the all-age group, were consistently higher using all models. They were also consistently higher than the effect estimates for the < 75 age group. For example, the effect estimate for the PS model using 8 *df*, was a 0.39% increase in deaths for the older age group (0.19 to 0.59) and a 0.25% increase for the younger age group (0.09 to 0.42). Using NS, the increases were 0.28% (0.08 to 0.47) and 0.14 % (−0.03 to 0.31) for the older and younger age groups. Thus, the estimates for the older age group were 50%–100% higher than for the younger age group.

The lower section of Table 11 shows that the estimates of PM_{10} effects on all-cause mortality were not altered after controlling for O_3 concentrations in any age group.

Table 12 presents the effect estimates of PM_{10} on cardiovascular mortality for two age groups (≥ 75 and < 75) for the various models. For the older age group, daily cardiovascular deaths were estimated to increase 0.48% (0.20 to 0.76) for lag 0–1 (PS, 8 *df*) and 0.47% (0.23 to 0.70) for lag 1. Although the DL model (PS, 8 *df*) had a positive estimate of 0.32% (−0.07 to 0.71), it was not statistically significant. The estimates varied somewhat by model, but most were statistically significant. Exceptions were a few models for the DL (cumulative effect). The results were not confounded by O_3 concentrations.

When compared with the effects on the older group, the effects of PM_{10} on cardiovascular mortality for the younger age group were much smaller (by almost half in many models). For several models, particularly those with 8 and 12 *df* and those using DL, the estimates were not statistically significant.

Table 13 show the effects of PM_{10} on respiratory mortality for all ages and for people 75 years or older. The daily number of deaths from respiratory causes was much smaller than that for all-cause or for cardiovascular mortality, and thus the corresponding CIs tended to be wider. Therefore, although positive effects were estimated using

Table 13. Europe: Percentage Change in Respiratory Mortality per 10-μg/m^3 Increase in PM_{10}

Seasonality Control	Average of Lags 0–1		Lag 1		Distributed Lags	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} Results						
All ages						
3 *df*/year	0.79 (0.35, 1.23)	−0.15 (−0.53, 0.24)	0.91 (0.54, 1.29)	0.07 (−0.27, 0.42)	0.75 (−0.03, 1.54)	−0.13 (−0.90, 0.66)
8 *df*/year	0.30 (−0.08, 0.68)	0.23 (−0.16, 0.62)	0.47 (0.13, 0.81)	0.37 (0.03, 0.73)	0.36 (−0.36, 1.09)	0.19 (−0.55, 0.95)
12 *df*/year	0.44 (0.05, 0.83)	0.45 (0.07, 0.85)	0.57 (0.23, 0.91)	0.55 (0.20, 0.89)	0.54 (−0.18, 1.26)	0.51 (−0.22, 1.24)
PACF	0.49 (0.11, 0.86)	0.26 (−0.12, 0.64)	0.62 (0.28, 0.96)	0.39 (0.06, 0.73)	0.60 (−0.13, 1.32)	0.34 (−0.38, 1.07)
≥ 75 Years						
3 *df*/year	0.97 (0.39, 1.55)	−0.05 (−0.54, 0.45)	1.05 (0.54, 1.56)	0.15 (−0.32, 0.62)	1.01 (0.11, 1.92)	0.08 (−0.82, 0.99)
8 *df*/year	0.34 (−0.10, 0.78)	0.26 (−0.19, 0.72)	0.47 (0.08, 0.86)	0.36 (−0.05, 0.77)	0.60 (−0.24, 1.45)	0.44 (−0.42, 1.32)
12 *df*/year	0.48 (0.03, 0.93)	0.51 (0.05, 0.97)	0.55 (0.15, 0.95)	0.54 (0.14, 0.95)	0.78 (−0.06, 1.62)	0.79 (−0.06, 1.65)
PACF	0.73 (0.23, 1.23)	0.22 (−0.23, 0.66)	0.79 (0.32, 1.26)	0.35 (−0.04, 0.75)	0.88 (0.03, 1.74)	0.60 (−0.26, 1.48)
Controlling for O_3						
All ages						
3 *df*/year	0.66 (0.27, 1.06)	−0.18 (−0.57, 0.22)	0.84 (0.49, 1.20)	0.06 (−0.30, 0.42)		
8 *df*/year	0.26 (−0.14, 0.65)	0.11 (−0.30, 0.52)	0.46 (0.11, 0.81)	0.33 (−0.04, 0.69)		
12 *df*/year	0.40 (−0.01, 0.80)	0.39 (−0.01, 0.80)	0.56 (0.21, 0.91)	0.55 (0.19, 0.91)		
PACF	0.41 (0.02, 0.80)	0.18 (−0.21, 0.57)	0.55 (0.20, 0.90)	0.31 (−0.04, 0.67)		
≥ 75 Years						
3 *df*/year	0.89 (0.31, 1.46)	−0.14 (−0.60, 0.31)	0.97 (0.47, 1.47)	0.05 (−0.36, 0.47)		
8 *df*/year	0.31 (−0.14, 0.77)	0.17 (−0.30, 0.65)	0.46 (0.06, 0.86)	0.32 (−0.11, 0.74)		
12 *df*/year	0.44 (−0.03, 0.91)	0.46 (−0.01, 0.94)	0.54 (0.13, 0.95)	0.53 (0.11, 0.95)		
PACF	0.58 (0.14, 1.03)	0.27 (−0.18, 0.73)	0.68 (0.26, 1.09)	0.36 (−0.05, 0.76)		

almost all models (except some with 3 *df*), a few were not statistically significant. The effects of PM_{10} on respiratory mortality for the older age group were generally larger compared with those at all ages, but for some models the difference was small. For the model using PS and 8 *df*, for example, an increase in PM_{10} lag 1 exposure of 10 μg/m^3 was associated with a 0.47% (0.13 to 0.81) increase in the daily number of respiratory deaths for all ages and with a 0.47% (0.08 to 0.86) increase for the older age group. The difference appeared more consistent and pronounced for the DL models. Lag 1 estimates tended to be larger than those of lag 0–1. An unexpected finding was the large difference in the effect estimates at 3 *df* between the NS and PS methods. The estimates were not substantially confounded by O_3 concentrations.

Figure 7A shows the results of a more extensive sensitivity analysis (using models from 2 to 22 *df* and the two smoothing methods, NS and PS) that was carried out for all-cause mortality of people 75 years or older exposed to a 10-ug/m^3 increase in PM_{10}, lag 1. The results using PS and NS differed when fewer degrees of freedom were used. PS estimates were larger than NS estimates, but the two models tended to converge with 8 *df* or more. All-cause mortality effect estimates using NS rose slightly. Using the minimization of the PACF criterion for model fitting, the investigators obtained estimates comparable with those models using 6 or 8 *df*.

A different pattern of sensitivity to model specification was observed in the analysis of respiratory mortality for the all-age group (see Figure 7B); the NS estimates were less stable compared with the PS estimates. Use of the two smoothers yielded comparable results with eight or more degrees of freedom; the estimates were lowest at 3 *df* (NS) and at 8 *df* (PS) and then tended to rise.

O_3 Table 14 shows the effect estimates for the change in all-cause mortality associated with an increase of 10 μg/m^3 in O_3 concentrations for lag 0–1, lag 1, and DL (lags 0–2); results for three age groups are compared (all ages, ≥ 75 and < 75). Interestingly, in contrast to the PM_{10} sensitivity analyses, the models using 3 *df* gave lower effects estimates in all three age groups than did models with more degrees of freedom, especially those using PS. All the effects estimates using more than 3 *df* were statistically significant. For all ages, the PS model using 8 *df* provided an estimated increase in the total daily number of deaths of 0.18% (0.07 to 0.30) with lag 0–1 and 0.17% (0.09 to 0.25) with lag 1.

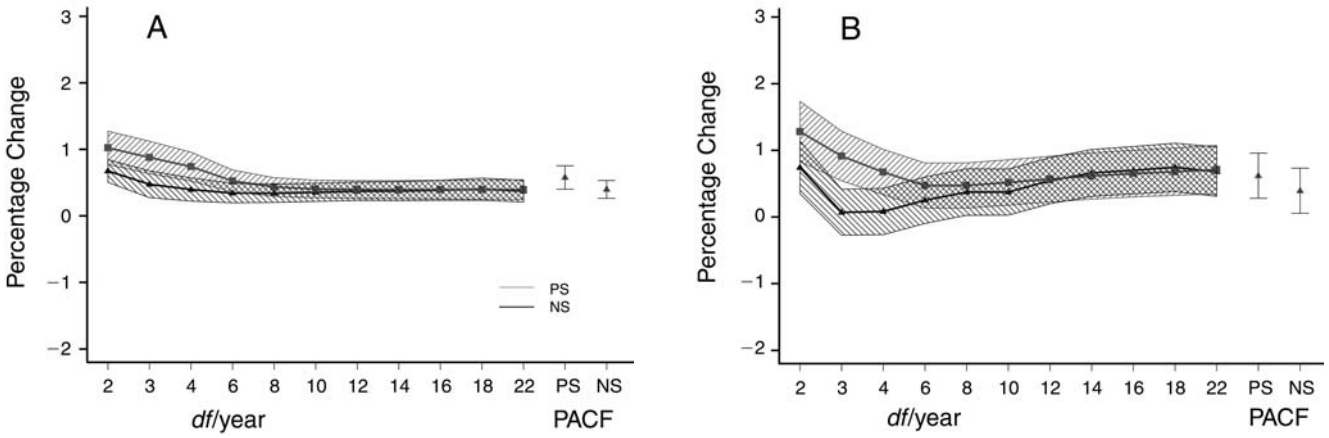


Figure 7. Europe sensitivity analyses: Percentage change associated with PM_{10} (lag 1) for A: all-cause mortality, ≥ 75 years; **B**: respiratory mortality, all ages.

Table 14. Europe: Percentage Change in All-Cause Mortality per 10-μg/m^3 Increase in O_3

Seasonality Control	Average of Lags 0–1		Lag 1		Distributed Lags	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 Results						
All ages						
3 *df*/year	−0.02 (−0.18, 0.14)	0.13 (0.00, 0.26)	0.02 (−0.10, 0.14)	0.14 (0.06, 0.22)	0.06 (−0.12, 0.24)	0.18 (0.02, 0.35)
8 *df*/year	0.18 (0.07, 0.30)	0.17 (0.05, 0.28)	0.17 (0.09, 0.25)	0.13 (0.05, 0.21)	0.25 (0.10, 0.40)	0.21 (0.06, 0.37)
12 *df*/year	0.17 (0.04, 0.30)	0.12 (−0.02, 0.26)	0.16 (0.08, 0.24)	0.10 (0.01, 0.20)	0.24 (0.09, 0.39)	0.17 (0.02, 0.33)
PACF	0.10 (−0.02, 0.22)	0.12 (−0.01, 0.24)	0.09 (0.01, 0.17)	0.09 (0.01, 0.17)	0.14 (−0.01, 0.29)	0.16 (0.01, 0.31)
≥ 75 Years						
3 *df*/year	−0.09 (−0.30, 0.11)	0.09 (−0.10, 0.27)	−0.02 (−0.17, 0.13)	0.10 (−0.03, 0.23)	−0.04 (−0.27, 0.20)	0.08 (−0.16, 0.31)
8 *df*/year	0.12 (−0.02, 0.26)	0.11 (−0.05, 0.26)	0.14 (0.04, 0.24)	0.10 (−0.02, 0.21)	0.17 (−0.03, 0.37)	0.11 (−0.10, 0.31)
12 *df*/year	0.11 (−0.04, 0.26)	0.07 (−0.12, 0.25)	0.13 (0.02, 0.24)	0.08 (−0.06, 0.21)	0.15 (−0.04, 0.34)	0.05 (−0.15, 0.27)
PACF	0.00 (−0.17, 0.17)	0.11 (−0.07, 0.30)	0.04 (−0.09, 0.16)	0.07 (−0.07, 0.21)	0.11 (−0.07, 0.29)	0.09 (−0.11, 0.29)
< 75 Years						
3 *df*/year	0.13 (−0.02, 0.29)	0.25 (0.10, 0.40)	0.13 (0.03, 0.24)	0.22 (0.11, 0.33)	0.22 (0.00, 0.44)	0.33 (0.11, 0.56)
8 *df*/year	0.25 (0.10, 0.40)	0.29 (0.14, 0.44)	0.18 (0.07, 0.29)	0.20 (0.08, 0.32)	0.37 (0.14, 0.59)	0.35 (0.12, 0.57)
12 *df*/year	0.23 (0.06, 0.40)	0.27 (0.12, 0.42)	0.17 (0.06, 0.28)	0.18 (0.06, 0.30)	0.36 (0.14, 0.59)	0.33 (0.10, 0.55)
PACF	0.15 (−0.03, 0.32)	0.22 (0.07, 0.37)	0.15 (0.04, 0.26)	0.18 (0.07, 0.29)	0.24 (0.02, 0.47)	0.29 (0.07, 0.52)
Controlling for PM_{10}						
All ages						
3 *df*/year	0.03 (−0.14, 0.19)	0.15 (0.00, 0.30)	0.07 (−0.05, 0.20)	0.17 (0.07, 0.27)		
8 *df*/year	0.21 (0.10, 0.31)	0.18 (0.07, 0.29)	0.19 (0.10, 0.28)	0.16 (0.06, 0.25)		
12 *df*/year	0.17 (0.07, 0.28)	0.10 (−0.01, 0.21)	0.16 (0.08, 0.25)	0.09 (0.00, 0.18)		
PACF	0.09 (−0.01, 0.19)	0.11 (−0.02, 0.23)	0.09 (0.00, 0.18)	0.10 (0.02, 0.19)		
≥ 75 Years						
3 *df*/year	−0.01 (−0.20, 0.18)	0.13 (−0.03, 0.30)	0.06 (−0.09, 0.21)	0.15 (0.02, 0.29)		
8 *df*/year	0.14 (0.00, 0.27)	0.10 (−0.04, 0.24)	0.16 (0.05, 0.28)	0.12 (0.00, 0.23)		
12 *df*/year	0.10 (−0.04, 0.23)	0.02 (−0.15, 0.19)	0.13 (0.02, 0.25)	0.06 (−0.08, 0.19)		
PACF	0.00 (−0.14, 0.14)	0.08 (−0.09, 0.26)	0.05 (−0.07, 0.16)	0.10 (−0.03, 0.23)		
< 75 Years						
3 *df*/year	0.15 (−0.05, 0.36)	0.23 (0.02, 0.44)	0.16 (0.02, 0.31)	0.24 (0.11, 0.37)		
8 *df*/year	0.31 (0.14, 0.47)	0.30 (0.12, 0.47)	0.24 (0.12, 0.37)	0.23 (0.10, 0.36)		
12 *df*/year	0.31 (0.15, 0.47)	0.25 (0.09, 0.42)	0.23 (0.10, 0.36)	0.18 (0.04, 0.31)		
PACF	0.15 (−0.05, 0.34)	0.24 (0.06, 0.42)	0.19 (0.06, 0.31)	0.22 (0.09, 0.34)		

Table 14 also indicates that, among people 75 years or older, most estimates of the effects of O_3 exposure on all-cause mortality were not statistically significant. If they were statistically significant, they were somewhat smaller than those for all ages. The effect estimates for people younger than 75 years were generally higher than for either of the other age groups. The increase estimated from the model using 8 *df* (PS) for lag 0–1 was 0.25% (0.10 to 0.40) and 0.18% (0.07 to 0.29) for lag 1 for the youngest age group, compared with corresponding estimates of 0.12% (−0.02 to 0.26) and 0.14% (0.04 to 0.24) for the oldest age group. The effect estimates increased slightly when adjusted for PM_{10}.

Table 15 compares the effect estimates for O_3 exposure on cardiovascular mortality for two age groups (≥ 75 and < 75). The effect estimates were higher compared with those for all-cause mortality in the same age groups. Thus, the increase estimated from the same model as above (8 *df*, PS) for lag 0–1 was 0.22% (0.00 to 0.45) and 0.17% (−0.02 to 0.36) for lag 1. Cardiovascular mortality effect estimates, especially in models using lag 0–1, were higher for the younger age group. Controlling for PM_{10} slightly increased the effect estimates from some models.

Table 16 displays the estimated O_3 effects on respiratory mortality for all ages and for people 75 years or older. The effect estimates were of a similar magnitude and were not statistically significant for either age group. For lag 0–1, some estimates did not indicate an increase at all, while for lag 1 the estimates were generally positive but not statistically significant.

Table 17 shows the results of the analyses for O_3 effects on all outcomes and for all age groups during the summer season. The effect estimates for all-cause mortality from the one-pollutant model (unadjusted for PM_{10}) were higher than those estimated from the models using all-year data. After adjusting for PM_{10}, the effects of O_3 were reduced and not statistically significant for lag 0–1 for any age group. They remained significant but smaller (especially for the older age group) for lag 1. A similar pattern was observed with cardiovascular mortality effect estimates. For respiratory mortality effects, one-pollutant models yielded much larger effect estimates, particularly for lag 1, than did the models using the annual data, but these effects appeared to be the result of confounding by PM_{10}. After adjusting for PM_{10}, the estimate for O_3 was greatly diminished.

Table 15. Europe: Percentage Change in Cardiovascular Mortality per 10-μg/m^3 Increase in O_3

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 Results						
≥ 75 Years						
3 *df*/year	−0.07 (−0.33, 0.18)	0.14 (−0.13, 0.40)	−0.03 (−0.23, 0.17)	0.12 (−0.09, 0.33)	0.04 (−0.24, 0.32)	0.20 (−0.10, 0.50)
8 *df*/year	0.22 (0.00, 0.45)	0.21 (−0.03, 0.44)	0.17 (−0.02, 0.36)	0.14 (−0.06, 0.34)	0.31 (0.03, 0.59)	0.26 (−0.03, 0.54)
12 *df*/year	0.24 (0.01, 0.47)	0.18 (−0.07, 0.44)	0.17 (−0.04, 0.38)	0.11 (−0.10, 0.33)	0.32 (0.05, 0.60)	0.23 (−0.05, 0.51)
PACF	0.08 (−0.18, 0.34)	0.14 (−0.11, 0.39)	0.06 (−0.14, 0.25)	0.08 (−0.13, 0.28)	0.13 (−0.15, 0.41)	0.24 (−0.04, 0.52)
< 75 Years						
3 *df*/year	0.24 (−0.03, 0.50)	0.32 (0.05, 0.60)	0.15 (−0.05, 0.35)	0.22 (0.02, 0.43)	0.23 (−0.15, 0.60)	0.31 (−0.07, 0.70)
8 *df*/year	0.35 (0.12, 0.58)	0.35 (0.11, 0.59)	0.18 (−0.02, 0.37)	0.17 (−0.03, 0.37)	0.26 (−0.12, 0.64)	0.25 (−0.14, 0.64)
12 *df*/year	0.33 (0.09, 0.57)	0.30 (0.05, 0.54)	0.15 (−0.05, 0.35)	0.12 (−0.09, 0.32)	0.22 (−0.16, 0.61)	0.16 (−0.23, 0.56)
PACF	0.22 (−0.03, 0.47)	0.36 (0.11, 0.61)	0.12 (−0.10, 0.33)	0.22 (0.03, 0.41)	0.25 (−0.13, 0.63)	0.28 (−0.10, 0.66)
Controlling for PM_{10}						
≥ 75 Years						
3 *df*/year	−0.01 (−0.29, 0.27)	0.16 (−0.13, 0.45)	0.04 (−0.19, 0.27)	0.16 (−0.08, 0.40)		
8 *df*/year	0.21 (−0.01, 0.43)	0.19 (−0.05, 0.44)	0.18 (−0.02, 0.39)	0.15 (−0.07, 0.37)		
12 *df*/year	0.25 (0.00, 0.49)	0.12 (−0.14, 0.37)	0.17 (−0.05, 0.39)	0.09 (−0.14, 0.32)		
PACF	0.03 (−0.22, 0.28)	0.14 (−0.12, 0.41)	0.06 (−0.15, 0.28)	0.13 (−0.09, 0.35)		
< 75 Years						
3 *df*/year	0.24 (−0.09, 0.56)	0.33 (−0.01, 0.67)	0.19 (−0.07, 0.44)	0.22 (−0.06, 0.50)		
8 *df*/year	0.36 (0.10, 0.62)	0.39 (0.12, 0.66)	0.23 (0.00, 0.46)	0.22 (−0.04, 0.47)		
12 *df*/year	0.33 (0.06, 0.60)	0.28 (0.00, 0.56)	0.20 (−0.03, 0.42)	0.14 (−0.09, 0.37)		
PACF	0.32 (0.03, 0.61)	0.39 (0.08, 0.70)	0.21 (−0.04, 0.45)	0.22 (−0.04, 0.48)		

Table 16. Europe: Percentage Change in Respiratory Mortality per 10-μg/m^3 Increase in O_3

Seasonality Control	Average of Lags 0–1		Lag 1		Distributed Lags	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 Results						
All ages						
3 *df*/year	−0.12 (−0.44, 0.20)	0.01 (−0.30, 0.33)	−0.04 (−0.41, 0.33)	0.10 (−0.22, 0.43)	0.17 (−0.38, 0.71)	0.13 (−0.42, 0.69)
8 *df*/year	−0.02 (−0.32, 0.28)	0.04 (−0.28, 0.35)	0.19 (−0.06, 0.45)	0.18 (−0.13, 0.48)	0.23 (−0.28, 0.73)	0.23 (−0.28, 0.75)
12 *df*/year	−0.11 (−0.41, 0.20)	−0.14 (−0.46, 0.19)	0.16 (−0.10, 0.42)	0.09 (−0.19, 0.37)	0.12 (−0.38, 0.62)	0.02 (−0.49, 0.53)
PACF	−0.14 (−0.45, 0.16)	−0.11 (−0.42, 0.20)	0.02 (−0.30, 0.35)	0.04 (−0.30, 0.37)	0.01 (−0.50, 0.51)	0.06 (−0.45, 0.58)
≥ 75 Years						
3 *df*/year	−0.15 (−0.49, 0.19)	−0.04 (−0.40, 0.32)	−0.02 (−0.41, 0.37)	0.11 (−0.24, 0.46)	0.08 (−0.54, 0.71)	0.11 (−0.53, 0.75)
8 *df*/year	−0.11 (−0.45, 0.25)	−0.05 (−0.41, 0.32)	0.15 (−0.14, 0.45)	0.15 (−0.22, 0.51)	0.12 (−0.46, 0.70)	0.14 (−0.45, 0.74)
12 *df*/year	−0.21 (−0.57, 0.14)	−0.26 (−0.63, 0.11)	0.10 (−0.20, 0.40)	0.03 (−0.30, 0.37)	−0.01 (−0.59, 0.56)	−0.14 (−0.73, 0.46)
PACF	−0.22 (−0.56, 0.13)	−0.09 (−0.45, 0.26)	0.02 (−0.38, 0.42)	0.15 (−0.25, 0.55)	−0.09 (−0.67, 0.49)	−0.10 (−0.69, 0.49)
Controlling for PM_{10}						
All ages						
3 *df*/year	0.11 (−0.24, 0.46)	0.15 (−0.20, 0.51)	0.10 (−0.28, 0.48)	0.23 (−0.07, 0.54)		
8 *df*/year	0.02 (−0.33, 0.36)	0.01 (−0.35, 0.37)	0.21 (−0.08, 0.50)	0.18 (−0.13, 0.48)		
12 *df*/year	−0.12 (−0.47, 0.23)	−0.20 (−0.56, 0.16)	0.13 (−0.17, 0.42)	0.01 (−0.29, 0.31)		
PACF	−0.16 (−0.50, 0.18)	−0.15 (−0.50, 0.20)	0.04 (−0.25, 0.33)	0.05 (−0.24, 0.35)		
≥ 75 Years						
3 *df*/year	0.06 (−0.32, 0.45)	0.02 (−0.39, 0.42)	0.13 (−0.24, 0.51)	0.15 (−0.20, 0.50)		
8 *df*/year	−0.12 (−0.52, 0.27)	−0.17 (−0.58, 0.25)	0.14 (−0.20, 0.48)	0.07 (−0.28, 0.42)		
12 *df*/year	−0.28 (−0.69, 0.12)	−0.39 (−0.80, 0.03)	0.05 (−0.29, 0.39)	−0.08 (−0.49, 0.33)		
PACF	−0.30 (−0.69, 0.10)	−0.27 (−0.68, 0.14)	−0.06 (−0.44, 0.31)	0.00 (−0.41, 0.42)		

Table 17. Europe: Percentage Change in Mortality per 10-μg/m^3 Increase in O_3—Summer-Only Analysis

	Average of Lags 0–1		Lag 1		Distributed Lags	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 Results						
All-cause mortality						
All ages	0.25 (0.10, 0.40)	0.21 (0.05, 0.37)	0.29 (0.19, 0.39)	0.26 (0.14, 0.37)	0.34 (0.16, 0.53)	0.30 (0.11, 0.49)
Ages ≥ 75	0.23 (0.07, 0.39)	0.18 (0.00, 0.35)	0.30 (0.17, 0.44)	0.26 (0.11, 0.41)	0.32 (0.08, 0.57)	0.26 (0.01, 0.51)
Ages < 75	0.29 (0.12, 0.46)	0.27 (0.10, 0.44)	0.27 (0.13, 0.41)	0.26 (0.12, 0.40)	0.36 (0.09, 0.63)	0.33 (0.06, 0.60)
Cardiovascular mortality						
Ages ≥ 75	0.37 (0.09, 0.66)	0.31 (0.01, 0.61)	0.39 (0.13, 0.65)	0.32 (0.05, 0.60)	0.51 (0.17, 0.86)	0.46 (0.12, 0.81)
Ages < 75	0.38 (0.06, 0.70)	0.36 (0.04, 0.68)	0.22 (−0.08, 0.52)	0.21 (−0.09, 0.52)	0.29 (−0.18, 0.76)	0.28 (−0.19, 0.75)
Respiratory mortality						
All ages	0.31 (−0.10, 0.72)	0.24 (−0.18, 0.65)	0.58 (0.25, 0.91)	0.54 (0.21, 0.88)	0.57 (−0.08, 1.22)	0.48 (−0.17, 1.13)
Ages ≥ 75	0.17 (−0.30, 0.65)	0.08 (−0.39, 0.56)	0.49 (0.11, 0.88)	0.44 (0.05, 0.82)	0.43 (−0.32, 1.18)	0.31 (−0.44, 1.06)
Controlling for PM_{10}						
All-cause mortality						
All ages	0.10 (−0.07, 0.27)	0.05 (−0.14, 0.23)	0.19 (0.07, 0.32)	0.16 (0.02, 0.29)		
Ages ≥ 75	0.06 (−0.13, 0.25)	0.00 (−0.21, 0.22)	0.19 (0.00, 0.34)	0.15 (0.00, 0.29)		
Ages < 75	0.19 (−0.04, 0.42)	0.15 (−0.09, 0.39)	0.24 (0.07, 0.40)	0.21 (0.04, 0.38)		
Cardiovascular mortality						
Ages ≥ 75	0.12 (−0.16, 0.41)	0.06 (−0.26, 0.37)	0.24 (−0.04, 0.52)	0.20 (−0.09, 0.50)		
Ages < 75	0.28 (−0.10, 0.67)	0.29 (−0.09, 0.67)	0.21 (−0.12, 0.54)	0.21 (−0.13, 0.55)		
Respiratory mortality						
All ages	−0.11 (−0.58, 0.37)	−0.19 (−0.66, 0.29)	0.20 (−0.19, 0.58)	0.15 (−0.23, 0.54)		
Ages ≥ 75	−0.33 (−0.88, 0.22)	−0.45 (−1.00, 0.11)	0.05 (−0.39, 0.50)	−0.02 (−0.46, 0.43)		

Table 18. United States: Percentage Change in All-Cause Mortality per 10-μg/m^3 Increase in PM_{10}

Seasonality Control	Average of Lags 0–1		Lag 1		Distributed Lags	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} Results						
All ages						
3 *df*/year	0.48 (0.19, 0.76)	0.23 (−0.03, 0.48)	0.63 (0.49, 0.77)	0.38 (0.26, 0.5)	0.69 (0.31, 1.06)	0.33 (0.05, 0.61)
8 *df*/year	0.14 (−0.12, 0.4)	0.11 (−0.16, 0.38)	0.29 (0.18, 0.4)	0.26 (0.15, 0.37)	0.26 (−0.08, 0.61)	0.15 (−0.16, 0.47)
12 *df*/year	0.12 (−0.14, 0.39)	0.08 (−0.2, 0.35)	0.27 (0.15, 0.38)	0.23 (0.11, 0.34)	0.25 (−0.1, 0.6)	0.15 (−0.17, 0.47)
PACF	0.6 (0.29, 0.9)	0.25 (0.03, 0.46)	0.62 (0.39, 0.86)	0.36 (0.15, 0.56)	0.54 (0.16, 0.92)	0.33 (0.08, 0.58)
≥ 75 Years						
3 *df*/year	0.68 (0.25, 1.11)	0.30 (−0.07, 0.68)	0.95 (0.76, 1.14)	0.59 (0.42, 0.76)	0.97 (0.44, 1.5)	0.46 (0.05, 0.87)
8 *df*/year	0.19 (−0.19, 0.56)	0.13 (−0.25, 0.51)	0.47 (0.31, 0.63)	0.41 (0.25, 0.57)	0.33 (−0.16, 0.82)	0.14 (−0.31, 0.59)
12 *df*/year	0.16 (−0.21, 0.52)	0.10 (−0.28, 0.49)	0.43 (0.26, 0.59)	0.37 (0.21, 0.53)	0.29 (−0.2, 0.78)	0.16 (−0.29, 0.6)
PACF	0.87 (0.43, 1.32)	0.36 (0.02, 0.69)	0.89 (0.57, 1.21)	0.50 (0.19, 0.82)	0.75 (0.16, 1.33)	0.50 (0.12, 0.89)
< 75 Years						
3 *df*/year	0.28 (0.01, 0.55)	0.16 (−0.12, 0.44)	0.32 (0.16, 0.48)	0.17 (0.03, 0.32)	0.43 (0, 0.85)	0.22 (−0.11, 0.55)
8 *df*/year	0.09 (−0.2, 0.38)	0.08 (−0.22, 0.39)	0.12 (−0.02, 0.27)	0.11 (−0.04, 0.26)	0.20 (−0.24, 0.63)	0.15 (−0.2, 0.5)
12 *df*/year	0.08 (−0.22, 0.39)	0.05 (−0.27, 0.37)	0.11 (−0.04, 0.26)	0.09 (−0.07, 0.24)	0.19 (−0.25, 0.63)	0.12 (−0.24, 0.48)
PACF	0.36 (0.08, 0.63)	0.24 (−0.02, 0.5)	0.38 (0.15, 0.61)	0.25 (0.04, 0.45)	0.48 (0.05, 0.91)	0.29 (−0.01, 0.6)
Controlling for O_3						
All ages						
3 *df*/year	0.44 (−0.11, 0.98)	0.24 (−0.16, 0.64)	0.60 (0.4, 0.79)	0.35 (0.18, 0.52)		
8 *df*/year	0.11 (−0.42, 0.64)	0.13 (−0.36, 0.61)	0.24 (0.08, 0.41)	0.24 (0.07, 0.41)		
12 *df*/year	0.09 (−0.47, 0.64)	0.07 (−0.5, 0.64)	0.22 (0.05, 0.39)	0.19 (0.01, 0.37)		
PACF	0.63 (0.22, 1.05)	0.37 (−0.05, 0.79)	0.68 (0.38, 0.98)	0.50 (0.17, 0.83)		
≥ 75 Years						
3 *df*/year	0.74 (0.06, 1.42)	0.29 (−0.26, 0.84)	0.85 (0.64, 1.06)	0.51 (0.29, 0.72)		
8 *df*/year	0.24 (−0.42, 0.9)	0.15 (−0.52, 0.83)	0.37 (0.16, 0.59)	0.35 (0.13, 0.57)		
12 *df*/year	0.19 (−0.47, 0.86)	0.09 (−0.6, 0.79)	0.33 (0.11, 0.55)	0.28 (0.04, 0.51)		
PACF	0.88 (0.24, 1.53)	0.57 (−0.05, 1.2)	0.93 (0.45, 1.4)	0.68 (0.15, 1.21)		
< 75 Years						
3 *df*/year	0.21 (−0.37, 0.78)	0.21 (−0.39, 0.82)	0.33 (0.07, 0.6)	0.19 (−0.05, 0.43)		
8 *df*/year	0 (−0.64, 0.64)	0.10 (−0.54, 0.75)	0.1 (−0.13, 0.34)	0.12 (−0.12, 0.37)		
12 *df*/year	0 (−0.73, 0.73)	0.03 (−0.76, 0.82)	0.1 (−0.14, 0.33)	0.08 (−0.16, 0.32)		
PACF	0.40 (0.01, 0.79)	0.22 (−0.19, 0.64)	0.46 (0.17, 0.76)	0.34 (0, 0.67)		

United States

PM Table 18 summarizes the effects of a 10-μg/m^3 increase in PM_{10}, using lags 0–1 and 1, on all-cause mortality for three age groups (all ages, ≥ 75, and < 75). Note that the effects of PM_{10} lag 0–1 could only be calculated for cities with daily PM measurements (15 cities for the U.S.), while estimates for lag 1 were calculated for all cities (n = 89). As a result, the CIs tended to be wider for the lag 0–1 effects. The effects of lag 1 were statistically significant for all models, while for lag 0–1 they were smaller for some models and less significant. For all ages, a 10-μg/m^3 PM_{10} increase with lag 0–1 and 8 *df* was associated with a 0.14% (−0.12 to 0.4) increase in the daily number of deaths using PS and 0.11% (−0.16 to 0.38) using NS. A similar increase in lag 1 PM_{10} was associated with a 0.29% (0.18 to 0.4) increase in the number of deaths using PS and a 0.26% (0.15 to 0.37) increase using NS. Distributed lag models (lags 0–2) showed an increase in the number of deaths of 0.26% (−0.08 to 0.61) using PS and 0.15% (−0.16 to 0.47) using NS. Adjusting for O_3 slightly decreased the estimated effects.

In Table 18, the corresponding effect estimates for people 75 years or older and people younger than 75 years indicate the same patterns as for all ages. However, the effect estimates were generally higher for the older age group, especially for lag 1 PM_{10}.

Table 19. United States: Percentage Change in Cardiovascular Mortality per 10-μg/m^3 Increase in PM_{10}

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} Results						
≥ 75 Years						
3 *df*/year	0.86 (0.33, 1.39)	0.48 (0.02, 0.94)	1.03 (0.79, 1.26)	0.67 (0.45, 0.88)	1.11 (0.44, 1.79)	0.64 (0.16, 1.11)
8 *df*/year	0.34 (−0.13, 0.81)	0.30 (−0.18, 0.78)	0.51 (0.29, 0.73)	0.46 (0.23, 0.68)	0.41 (−0.24, 1.06)	0.26 (−0.27, 0.8)
12 *df*/year	0.31 (−0.15, 0.77)	0.27 (−0.2, 0.75)	0.45 (0.23, 0.67)	0.40 (0.17, 0.63)	0.36 (−0.29, 1.01)	0.31 (−0.24, 0.85)
PACF	1.04 (0.49, 1.59)	0.67 (0.13, 1.22)	1.01 (0.6, 1.42)	0.74 (0.33, 1.16)	1.11 (0.38, 1.84)	0.74 (0.28, 1.21)
< 75 Years						
3 *df*/year	0.25 (−0.22, 0.73)	0.12 (−0.4, 0.64)	0.37 (0.15, 0.6)	0.21 (−0.03, 0.45)	0.53 (−0.23, 1.28)	0.35 (−0.23, 0.94)
8 *df*/year	0.03 (−0.52, 0.59)	0.06 (−0.52, 0.64)	0.19 (−0.05, 0.44)	0.21 (−0.05, 0.47)	0.28 (−0.51, 1.08)	0.38 (−0.27, 1.02)
12 *df*/year	0.01 (−0.55, 0.57)	−0.05 (−0.62, 0.51)	0.18 (−0.07, 0.43)	0.16 (−0.11, 0.42)	0.30 (−0.5, 1.1)	0.21 (−0.44, 0.86)
PACF	0.35 (−0.11, 0.82)	0.18 (−0.29, 0.66)	0.42 (0.06, 0.79)	0.36 (−0.02, 0.73)	0.63 (−0.13, 1.39)	0.41 (−0.14, 0.95)
Controlling for O_3						
≥ 75 Years						
3 *df*/year	0.63 (−0.45, 1.71)	0.17 (−0.8, 1.14)	0.93 (0.62, 1.23)	0.59 (0.29, 0.9)		
8 *df*/year	0.23 (−0.84, 1.3)	0.03 (−1.09, 1.15)	0.42 (0.12, 0.73)	0.41 (0.09, 0.73)		
12 *df*/year	0.20 (−0.86, 1.26)	0.10 (−0.95, 1.15)	0.35 (0.04, 0.66)	0.33 (0, 0.66)		
PACF	1.06 (0.24, 1.87)	0.77 (0, 1.54)	1.09 (0.5, 1.69)	0.86 (0.23, 1.49)		
< 75 Years						
3 *df*/year	0.28 (−0.88, 1.44)	0.15 (−0.63, 0.93)	0.57 (0.24, 0.91)	0.40 (0.05, 0.75)		
8 *df*/year	0.01 (−1.29, 1.3)	−0.05 (−1.24, 1.13)	0.34 (−0.02, 0.69)	0.37 (−0.01, 0.75)		
12 *df*/year	−0.01 (−1.43, 1.42)	−0.04 (−1.55, 1.48)	0.33 (−0.04, 0.69)	0.31 (−0.1, 0.73)		
PACF	0.40 (−0.31, 1.11)	0.16 (−0.51, 0.82)	0.52 (−0.03, 1.07)	0.41 (−0.11, 0.92)		

Table 19 shows the effect estimates, per 10-μg/m^3 increase in PM_{10}, for lags 0–1 and 1, on cardiovascular mortality for people 75 years or older and people younger than 75 years. For the older age group, the increase in the daily number of cardiovascular deaths using a model with 8 *df* was about 0.30% for lag 0–1 (both for PS and NS, though not statistically significant for either) but it was even larger for lag 1: 0.51% (0.29 to 0.73) for PS and 0.46% (0.23 to 0.68) for NS. The estimates of the DL models were similar to those of lag 0–1.

The effects of O_3 on cardiovascular mortality were generally smaller for people younger than 75 years than those for the older age group. The effect estimates were close to zero for lag 0–1 and about half the values of the older age group for lag 1 using either smoothing method. The DL effects were again smaller than those in the older age group using PS but larger in the younger age group using NS; however all DL estimates were statistically insignificant for the younger age group. The effects on cardiovascular deaths in people younger than 75 years were only significant for lag 1 in models using 3 *df*/year or the minimal number of degrees of freedom determined by the PACF method.

Adjusting for O_3 generally reduced most effect estimates slightly for the older age group, while increasing them slightly for the younger age groups. However, for lag 0–1 with the NS method, adding O_3 to the model with 8 *df* reduced the estimate to about zero in the older age group.

Table 20 shows the estimates of the effects of a 10-μg/m^3 increase in PM_{10} on respiratory mortality for all ages and for people 75 years or older for all the models used in the sensitivity analysis. In the all-ages group, the effects were not statistically significant for models using 8 and 12 *df*. They were significant only in models using PS and either 3 *df* or the PACF method for degrees of freedom. The patterns in the effect estimates among people 75 years or older were generally similar to those in the all-ages group.

As for the Canadian and European data, we conducted a more extensive sensitivity analysis using degrees of freedom ranging from 2 up to 22; the results are shown in Figure 8A for all ages and Figure 8B for people 75 years or older. The effect estimates were highest with 2 *df* but declined and became stable with ≥ 8 *df*.

Table 20. United States: Percentage Change in Respiratory Mortality per 10-μg/m^3 Increase in PM_{10}

Seasonality Control	Average of Lags 0–1		Lag 1		Distributed Lags	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} Results						
All ages						
3 *df*/year	1.17 (0.41, 1.93)	0.40 (−0.33, 1.14)	1.03 (0.63, 1.42)	0.39 (0.02, 0.75)	1.50 (0.36, 2.65)	0.56 (−0.3, 1.42)
8 *df*/year	0.25 (−0.51, 1.01)	0.22 (−0.55, 0.99)	0.19 (−0.23, 0.61)	0.15 (−0.28, 0.59)	0.36 (−0.8, 1.51)	0.07 (−0.91, 1.05)
12 *df*/year	0.15 (−0.65, 0.94)	0.02 (−0.77, 0.8)	0.11 (−0.33, 0.56)	0.09 (−0.36, 0.54)	0.21 (−0.98, 1.4)	−0.05 (−1.06, 0.96)
PACF	1.55 (0.8, 2.3)	0.41 (−0.29, 1.11)	1.34 (0.73, 1.96)	0.49 (−0.12, 1.09)	1.20 (0.01, 2.39)	0.71 (−0.18, 1.6)
≥ 75 Years						
3 *df*/year	1.41 (0.44, 2.37)	0.51 (−0.42, 1.45)	1.20 (0.72, 1.68)	0.44 (−0.03, 0.92)	1.55 (0.12, 2.99)	0.28 (−0.92, 1.47)
8 *df*/year	0.59 (−0.35, 1.52)	0.54 (−0.39, 1.46)	0.39 (−0.13, 0.9)	0.35 (−0.19, 0.89)	0.48 (−0.94, 1.9)	0.12 (−1.11, 1.35)
12 *df*/year	0.48 (−0.51, 1.47)	0.31 (−0.71, 1.34)	0.34 (−0.2, 0.88)	0.23 (−0.31, 0.77)	0.37 (−1.08, 1.82)	−0.11 (−1.36, 1.15)
PACF	1.90 (0.86, 2.93)	0.80 (−0.18, 1.78)	0.46 (−1.56, 2.49)	0.40 (−0.38, 1.19)	−12.47 (−44.77, 19.83)	0.67 (−0.6, 1.94)
Controlling for O_3						
All ages						
3 *df*/year	0.79 (−0.74, 2.32)	0.43 (−0.49, 1.35)	0.86 (0.24, 1.47)	0.24 (−0.33, 0.8)		
8 *df*/year	0.13 (−1.42, 1.68)	0.24 (−0.71, 1.18)	−0.01 (−0.67, 0.65)	0 (−0.73, 0.73)		
12 *df*/year	0.02 (−1.6, 1.63)	−0.13 (−1.72, 1.46)	−0.07 (−0.77, 0.63)	−0.07 (−0.8, 0.66)		
PACF	1.52 (0.67, 2.36)	0.73 (−0.16, 1.63)	1.26 (0.56, 1.97)	0.62 (−0.14, 1.38)		
≥ 75 Years						
3 *df*/year	1.32 (−0.35, 2.99)	0.56 (−0.66, 1.79)	1 (0.22, 1.78)	0.29 (−0.48, 1.06)		
8 *df*/year	0.63 (−1.09, 2.36)	0.54 (−0.68, 1.75)	0.15 (−0.76, 1.06)	0.24 (−0.72, 1.21)		
12 *df*/year	0.57 (−1.24, 2.39)	0.31 (−1.67, 2.29)	0.08 (−0.86, 1.03)	0.11 (−0.84, 1.06)		
PACF	1.92 (0.61, 3.23)	1.05 (−0.17, 2.28)	1.24 (0.17, 2.31)	0.58 (−0.51, 1.67)		

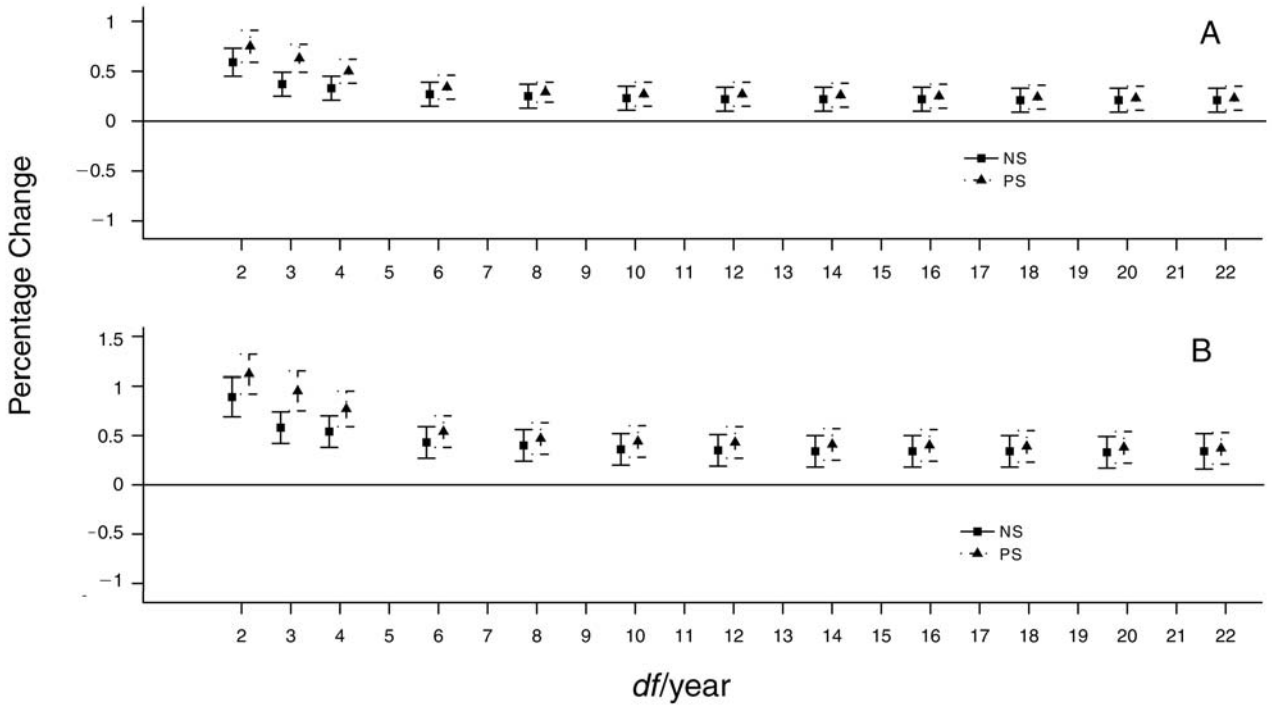


Figure 8. United States sensitivity analysis: Percentage change in all-cause mortality associated with a 10-μg/m^3 PM_{10} increase with lag 1 for A: all ages; **B**: ≥ 75 years.

O_3 Table 21 shows the estimated effects of O_3 exposure on all-cause mortality for all three age groups, for all model variations. All U.S. cities had daily data, so the number of cities used in the analyses was the same for all lags. For all ages, the effects for lag 0–1, lag 1, and DL models were significant when 8 and 12 *df*/year were used, but not when 3 *df* or the PACF criterion were used. In models using 8 *df*, the estimated increase in all-cause mortality per 10-μg/m^3 O_3 at lag 0–1 was 0.31% (0.09 to 0.52) for PS and 0.34% (0.13 to 0.55) for NS; for lag 1 it was 0.18% for both PS (0 to 0.35) and NS (0.01 to 0.35). For the DL model (lags 0–2), the effect estimates were 0.43% (0.11 to 0.75) for PS and 0.38% (0.14 to 0.61) for NS. The estimates were lower and not significant when adjusted for PM.

Table 21. United States: Percentage Change in All-Cause Mortality per 10-μg/m^3 Increase in O_3

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 Results						
All ages						
3 *df*/year	−0.55 (−0.88, −0.22)	0.08 (−0.17, 0.33)	−0.51 (−0.8, −0.22)	−0.01 (−0.21, 0.2)	−0.73 (−1.15, −0.31)	0.01 (−0.26, 0.29)
8 *df*/year	0.31 (0.09, 0.52)	0.34 (0.13, 0.55)	0.18 (0, 0.35)	0.18 (0.01, 0.35)	0.43 (0.11, 0.75)	0.38 (0.14, 0.61)
12 *df*/year	0.33 (0.12, 0.54)	0.32 (0.12, 0.52)	0.19 (0.01, 0.36)	0.16 (−0.01, 0.32)	0.48 (0.16, 0.79)	0.36 (0.13, 0.6)
PACF	−0.62 (−0.98, −0.27)	0.41 (0.17, 0.64)	−0.57 (−0.89, −0.26)	0.22 (0.02, 0.41)	−0.34 (−0.74, 0.06)	0.43 (0.16, 0.7)
≥ 75 Years						
3 *df*/year	−0.77 (−1.23, −0.31)	0.07 (−0.29, 0.43)	−0.69 (−1.09, −0.3)	−0.01 (−0.31, 0.28)	−1.09 (−1.67, −0.51)	−0.08 (−0.5, 0.33)
8 *df*/year	0.33 (0.02, 0.64)	0.36 (0.06, 0.66)	0.21 (−0.04, 0.46)	0.22 (−0.02, 0.47)	0.4 (−0.05, 0.85)	0.32 (−0.04, 0.67)
12 *df*/year	0.37 (0.06, 0.67)	0.33 (0.04, 0.63)	0.24 (−0.01, 0.48)	0.2 (−0.04, 0.44)	0.46 (0.01, 0.91)	0.31 (−0.03, 0.65)
PACF	−0.87 (−1.35, −0.39)	0.5 (0.13, 0.87)	−0.79 (−1.22, −0.37)	0.31 (0.02, 0.61)	−0.61 (−1.17, −0.05)	0.32 (−0.11, 0.75)
< 75 Years						
3 *df*/year	−0.34 (−0.66, −0.02)	0.11 (−0.17, 0.39)	−0.34 (−0.62, −0.07)	0.01 (−0.23, 0.24)	−0.29 (−0.74, 0.15)	0.13 (−0.17, 0.44)
8 *df*/year	0.3 (0.04, 0.56)	0.33 (0.07, 0.59)	0.15 (−0.07, 0.37)	0.15 (−0.07, 0.37)	0.5 (0.09, 0.92)	0.45 (0.15, 0.75)
12 *df*/year	0.33 (0.06, 0.59)	0.31 (0.04, 0.57)	0.16 (−0.06, 0.38)	0.12 (−0.1, 0.34)	0.54 0.13, 0.96)	0.43 (0.12, 0.73)
PACF	−0.41 (−0.75, −0.08)	0.41 (0.13, 0.68)	−0.41 (−0.7, −0.11)	0.27 (0.03, 0.5)	−0.18 (−0.62, 0.26)	0.46 (0.16, 0.75)
Controlling for PM_{10}						
All ages						
3 *df*/year	0.19 (−0.82, 1.2)	0.19 (−0.82, 1.2)	−0.68 (−1.1, −0.26)	−0.68 (−1.1, −0.26)		
8 *df*/year	0.52 (−0.57, 1.61)	0.52 (−0.57, 1.61)	0.13 (−0.18, 0.44)	0.13 (−0.18, 0.44)		
12 *df*/year	0.51 (−0.62, 1.63)	0.51 (−0.62, 1.63)	0.12 (−0.2, 0.45)	0.12 (−0.2, 0.45)		
PACF	0.14 (−0.76, 1.03)	0.14 (−0.76, 1.03)	−0.75 (−1.2, −0.3)	−0.75 (−1.2, −0.3)		
≥ 75 Years						
3 *df*/year	−0.58 (−2.79, 1.63)	−0.58 (−2.79, 1.63)	−0.85 (−1.51, −0.19)	−0.85 (−1.51, −0.19)		
8 *df*/year	−0.33 (−3.01, 2.36)	−0.33 (−3.01, 2.36)	0.12 (−0.39, 0.63)	0.12 (−0.39, 0.63)		
12 *df*/year	−0.25 (−3.09, 2.58)	−0.25 (−3.09, 2.58)	0.12 (−0.4, 0.65)	0.12 (−0.4, 0.65)		
PACF	−0.44 (−2.3, 1.42)	−0.44 (−2.3, 1.42)	−0.9 (−1.57, −0.24)	−0.9 (−1.57, −0.24)		
< 75 Years						
3 *df*/year	0.68 (−0.58, 1.93)	0.68 (−0.58, 1.93)	−0.49 (−0.97, −0.01)	−0.49 (−0.97, −0.01)		
8 *df*/year	1.06 (−0.53, 2.65)	1.06 (−0.53, 2.65)	0.12 (−0.33, 0.56)	0.12 (−0.33, 0.56)		
12 *df*/year	0.95 (−0.68, 2.57)	0.95 (−0.68, 2.57)	0.12 (−0.34, 0.57)	0.12 (−0.34, 0.57)		
PACF	0.58 (−0.47, 1.62)	0.58 (−0.47, 1.62)	−0.57 (−1.06, −0.08)	−0.57 (−1.06, −0.08)		

The estimated effects of O_3 exposure on all-cause mortality for the older age group (≧ 75) showed a similar pattern across the various models as for all ages (Table 21). However, the estimates moved toward the null after adjusting for PM_{10}. For people younger than 75 years, the estimated effects using the lag 0–1, lag 1, and the DL models were significant when 8 or 12 *df*/year were used. When adjusted for PM_{10}, the effects estimated using lag 0–1 and 1 became nonsignificant (DL not estimated).

Table 22 summarizes the effects of O_3 on cardiovascular mortality for two age groups (≧ 75 and < 75). The effects of O_3 were higher at lag 0–1 with 8 or 12 *df*/year when com-

Table 22. United States: Percentage Change in Cardiovascular Mortality per 10-μg/m^3 Increase in O_3

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 Results						
≧ 75 Years						
3 *df*/year	−0.77 (−1.3, −0.24)	0.05 (−0.37, 0.48)	−0.74 (−1.21, −0.27)	−0.07(−0.41, 0.28)	−1.07 (−1.78, −0.36)	−0.08 (−0.56, 0.4)
8 *df*/year	0.34 (−0.03, 0.71)	0.31 (−0.07, 0.69)	0.18 (−0.13, 0.49)	0.14 (−0.17, 0.46)	0.48 (−0.12, 1.09)	0.29 (−0.17, 0.75)
12 *df*/year	0.38 (0.01, 0.75)	0.35 (−0.04, 0.73)	0.2 (−0.11, 0.51)	0.17 (−0.15, 0.49)	0.55 (−0.05, 1.15)	0.36 (−0.09, 0.82)
PACF	−0.89 (−1.46, −0.32)	0.63 (0.18, 1.09)	−0.85 (−1.37, −0.33)	0.28 (−0.05, 0.62)	−0.68 (−1.36, −0.01)	0.3 (−0.17, 0.77)
< 75 Years						
3 *df*/year	−0.37 (−0.78, 0.04)	0.09 (−0.31, 0.49)	−0.51 (−0.86, −0.16)	−0.17 (−0.51, 0.17)	−0.26 (−0.97, 0.44)	0.13 (−0.32, 0.59)
8 *df*/year	0.37 (−0.03, 0.78)	0.33 (−0.11, 0.77)	0.05 (−0.30, 0.40)	−0.02 (−0.39, 0.36)	0.56 (−0.15, 1.26)	0.48 (−0.02, 0.98)
12 *df*/year	0.39 (−0.04, 0.81)	0.30 (−0.15, 0.74)	0.05 (−0.31, 0.41)	−0.05 (−0.43, 0.33)	0.58 (−0.12, 1.29)	0.44 (−0.07, 0.95)
PACF	−0.37 (−0.83, 0.08)	0.72 (0.31, 1.13)	−0.51 (−0.9, −0.11)	0.32 (−0.03, 0.66)	−0.23 (−0.94, 0.47)	0.82 (0.38, 1.27)
Controlling for PM_{10}						
≧ 75 Years						
3 *df*/year	0.19 (−2.46, 2.84)	0.19 (−2.46, 2.84)	−0.9 (−1.73, −0.07)	−0.9 (−1.73, −0.07)		
8 *df*/year	0.49 (−2.61, 3.6)	0.49 (−2.61, 3.6)	0.06 (−0.6, 0.72)	0.06 (−0.6, 0.72)		
12 *df*/year	0.56 (−2.82, 3.95)	0.56 (−2.82, 3.95)	0.08 (−0.6, 0.75)	0.08 (−0.6, 0.75)		
PACF	0.19 (−1.99, 2.36)	0.19 (−1.99, 2.36)	−0.97 (−1.83, −0.11)	−0.97 (−1.83, −0.11)		
< 75 Years						
3 *df*/year	1.2 (−0.31, 2.7)	1.2 (−0.31, 2.7)	−0.2 (−0.8, 0.39)	−0.2 (−0.8, 0.39)		
8 *df*/year	1.2 (−0.67, 3.08)	1.2 (−0.67, 3.08)	0.17 (−0.47, 0.82)	0.17 (−0.47, 0.82)		
12 *df*/year	1.21 (−0.76, 3.18)	1.21 (−0.76, 3.18)	0.16 (−0.51, 0.84)	0.16 (−0.51, 0.84)		
PACF	1.56 (0.19, 2.93)	1.56 (0.19, 2.93)	−0.17 (−0.8, 0.46)	−0.17 (−0.8, 0.46)		

pared with lag 1 for both groups; however, the only estimate that was significant was for the older age group using the model with lag 0–1, PS, and 12 *df*. The effects with DL (lags 0–2) and both PS and NS were higher than with lag 0–1, 8 or 12 *df*, but all estimates were nonsignificant. Effect estimates using 3 *df* or degrees of freedom determined using the PACF method were generally negative. The effects at lag 0–1 remained positive, but in most cases not significant, after adjusting for PM_{10}.

The effects of O_3 on respiratory mortality for all ages and for people 75 years or older are displayed in Table 23. When 8 or 12 *df*/year were used, the estimates were generally

Table 23. United States: Percentage Change in Respiratory Mortality per 10-μg/m^3 Increase in O_3

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 Results						
All ages						
3 *df*/year	−1.89 (−2.79, −0.99)	−0.49 (−1.24, 0.25)	−1.52 (−2.28, −0.76)	−0.34 (−0.95, 0.27)	−2.27 (−3.46, −1.08)	−0.66 (−1.5, 0.19)
8 *df*/year	0.19 (−0.44, 0.83)	0.27 (−0.39, 0.92)	0.28 (−0.25, 0.81)	0.31 (−0.24, 0.85)	0.4 (−0.63, 1.43)	0.32 (−0.43, 1.08)
12 *df*/year	0.23 (−0.4, 0.86)	0.24 (−0.42, 0.89)	0.31 (−0.22, 0.84)	0.27 (−0.28, 0.82)	0.48 (−0.55, 1.51)	0.32 (−0.46, 1.09)
PACF	−1.49 (−2.52, −0.46)	0.24 (−0.53, 1.01)	−1.29 (−2.17, −0.41)	0.3 (−0.3, 0.89)	−0.71 (−1.82, 0.41)	0.43 (−0.36, 1.21)
≥ 75 Years						
3 *df*/year	−2.36 (−3.36, −1.36)	−0.71 (−1.55, 0.13)	−1.94 (−2.81, −1.07)	−0.53 (−1.26, 0.2)	−2.73 (−4.2, −1.27)	−0.9 (−1.92, 0.12)
8 *df*/year	−0.03 (−0.8, 0.74)	0.13 (−0.69, 0.94)	0.08 (−0.59, 0.75)	0.19 (−0.5, 0.89)	0.13 (−1.19, 1.44)	0.14 (−0.85, 1.13)
12 *df*/year	0.04 (−0.75, 0.83)	0.10 (−0.73, 0.93)	0.14 (−0.55, 0.83)	0.16 (−0.55, 0.86)	0.25 (−1.07, 1.57)	0.13 (−0.9, 1.17)
PACF	−1.98 (−3.13, −0.83)	0.52 (−0.22, 1.25)	−1.74 (−2.75, −0.74)	0.43 (−0.22, 1.08)	−1.16 (−2.57, 0.24)	0.64 (−0.19, 1.47)
Controlling for PM_{10}						
All ages						
3 *df*/year	0.65 (−1.89, 3.19)	0.65 (−1.89, 3.19)	−1.21 (−2.34, −0.07)	−1.21 (−2.34, −0.07)		
8 *df*/year	0.62 (−1.98, 3.21)	0.62 (−1.98, 3.21)	0.44 (−0.55, 1.43)	0.44 (−0.55, 1.43)		
12 *df*/year	0.28 (−2.42, 2.97)	0.28 (−2.42, 2.97)	0.51 (−0.54, 1.56)	0.51 (−0.54, 1.56)		
PACF	1.64 (−0.96, 4.24)	1.64 (−0.96, 4.24)	−0.89 (−2.11, 0.33)	−0.89 (−2.11, 0.33)		
≥ 75 Years						
3 *df*/year	0.62 (−2.61, 3.84)	0.62 (−2.61, 3.84)	−1.3 (−2.46, −0.14)	−1.3 (−2.46, −0.14)		
8 *df*/year	0.63 (−2.58, 3.84)	0.63 (−2.58, 3.84)	0.36 (−0.79, 1.5)	0.36 (−0.79, 1.5)		
12 *df*/year	0.49 (−2.68, 3.67)	0.49 (−2.68, 3.67)	0.41 (−0.76, 1.58)	0.41 (−0.76, 1.58)		
PACF	1.52 (−2.01, 5.06)	1.52 (−2.01, 5.06)	−0.71 (−2.22, 0.81)	−0.71 (−2.22, 0.81)		

positive but not statistically significant for both age groups; when 3 *df* or the minimum degrees of freedom based on the PACF method were used, the estimates were generally negative, and using PS, statistically significant. Adjusting for PM_{10} brought the estimated effects close to the null for most models except lag 1, 3 *df,* and PACF. In most cases, adjustment for PM_{10} resulted in slight increases in the effect estimates, but results were still not statistically significant.

Table 24 shows the summer-only effects of O_3 on the various mortality outcomes. An increase in O_3 of 10 μg/m^3 is associated with an increase in the daily number of deaths (all ages) of 0.65% (using PS) and 0.48% (using NS) for the DL model; 0.57% (PS) and 0.53% (NS) for lags 0–1, and 0.49% (PS) and 0.54% (NS) for lag 1. All of these effects were significant. They remained positive, but became nonsignificant when adjusted for PM_{10}. The pattern was similar for people 75 years or older.

Effects of O_3 on cardiovascular disease mortality in the older age group were positive and statistically significant for one-pollutant models (with the exception of the DL model using NS), but became statistically insignificant with adjustment for PM_{10}. For people in the younger age group, the estimates were positive and statistically significant for all models.

Table 24 also shows the summer-only effects of O_3 on respiratory mortality for all ages and for the older age group. The effects were positive and marginally significant for one-pollutant models, but after adjusting for PM_{10} remained positive only for lag 1.

To check the methodology for pooling the O_3 estimates, the investigators compared the Berkey metaregression method with the TLNISE hierarchical model approach (Table 25). As shown in the table, the pooled mortality estimates from each model were very close when compared for the same degrees of freedom.

Table 25. Results for O_3 Using Berkey MLE, Average of Lags 0–1, Penalized Splines

Seasonality Control	Berkey	TLNISE
3 *df*/year	−0.54 (−0.88, 0.20)	−0.55 (−0.88, −0.22)
8 *df*/year	0.30 (0.11, 0.50)	0.31 (0.09, 0.52)
12 *df*/year	0.34 (0.15, 0.53)	0.33 (0.12, 0.54)
PACF	−0.62 (−1.01, −0.22)	−0.62 (−0.98, −0.27)

Table 24. United States: Percentage Change in Mortality per 10-μg/m^3 Increase in O_3—Summer-Only Analysis

	Average of Lags 0–1		Lag 1		Distributed Lags	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 Results						
All-cause mortality						
All ages	0.57 (0.35, 0.79)	0.53 (0.31, 0.75)	0.49 (0.29, 0.69)	0.54 (0.28, 0.80)	0.65 (0.33, 0.97)	0.48 (0.24, 0.72)
Ages ≥ 75	0.63 (0.33, 0.93)	0.58 (0.28, 0.88)	0.58 (0.32, 0.84)	0.54 (0.28, 0.80)	0.62 (0.16, 1.08)	0.39 (0.03, 0.75)
Ages < 75	0.50 (0.22, 0.78)	0.48 (0.20, 0.76)	0.40 (0.16, 0.64)	0.38 (0.14, 0.62)	0.70 (0.28, 1.12)	0.58 (0.28, 0.88)
Cardiovascular mortality						
Ages ≥ 75	0.56 (0.16, 0.96)	0.50 (0.08, 0.92)	0.49 (0.15, 0.83)	0.45 (0.11, 0.79)	0.65 (0.03, 1.27)	0.40 (−0.06, 0.86)
Ages < 75	0.64 (0.22, 1.06)	0.58 (0.14, 1.02)	0.45 (0.09, 0.81)	0.40 (0.04, 0.76)	0.96 (0.24, 1.68)	0.84 (0.34, 1.34)
Respiratory mortality						
All ages	0.66 (−0.02, 1.34)	0.63 (−0.07, 1.33)	0.77 (0.17, 1.37)	0.75 (0.15, 1.35)	0.73 (−0.39, 1.85)	0.55 (−0.27, 1.37)
Ages ≥ 75	0.70 (−0.08, 1.48)	0.67 (−0.21, 1.55)	0.71 (−0.07, 1.49)	0.70 (−0.08, 1.48)	0.79 (−0.65, 2.23)	0.51 (−0.55, 1.57)
Controlling for PM_{10}						
All-cause mortality						
All ages	0.15 (−0.51, 0.81)	0.14 (−0.52, 0.80)	0.27 (−0.07, 0.61)	0.24 (−0.10, 0.58)		
Ages ≥ 75	0.07 (−1.01, 1.15)	0.11 (−0.97, 1.19)	0.26 (−0.24, 0.76)	0.21 (−0.29, 0.71)		
Ages < 75	0.16 (−0.76, 1.08)	0.14 (−0.80, 1.08)	0.27 (−0.17, 0.71)	0.27 (−0.17, 0.71)		
Cardiovascular mortality						
Ages ≥ 75	−0.31 (−1.91, 1.29)	−0.21 (−1.77, 1.35)	−0.09 (−0.75, 0.57)	−0.15 (−0.81, 0.51)		
Ages < 75	1.08 (−0.74, 2.90)	1.10 (−0.72, 2.92)	0.22 (−0.52, 0.96)	0.16 (−0.58, 0.90)		
Respiratory mortality						
All ages	0.05 (−2.81, 2.91)	−0.13 (−2.99, 2.73)	0.99 (−0.33, 2.31)	0.87 (−0.45, 2.19)		
Ages ≥ 75	0.20 (−3.38, 3.78)	−0.18 (−3.72, 3.36)	0.72 (−0.88, 2.32)	0.56 (−1.00, 2.12)		

EFFECTS OF PM_{10} AND O_3 ON HOSPITAL ADMISSIONS FOR PEOPLE ≥ 65 YEARS WITH CARDIOVASCULAR OR RESPIRATORY DISEASE

Canada

PM Tables 26 and 27 show the effects of PM_{10} on respiratory and cardiovascular disease admissions. No effect was observed for the models using 8 or 12 *df*, while the PACF model yielded positive estimates significant only for respiratory admissions. Lag 1 was the only lag assessed because lag 0–1 models require daily data, and Canada has PM_{10} data for 1 of every 6 days.

O_3 Tables 28 and 29 show the effects of O_3 on respiratory and cardiovascular disease admissions for the various models investigated. The estimated effects on respiratory admissions were positive for lags 0–1, 1, and DL models, and were statistically significant except for the models with 8 *df*/year (lag 1) and 12 *df*/year (lag 0–1 and lag 1). They were somewhat larger for lag 1. When adjusted for PM_{10}, most effects for lag 1 decreased, and all became insignificant. For cardiovascular disease admissions, the effects using any model with 8 or 12 *df* were close to the null, while positive and significant effects were estimated with the lag 1 models and either 3 *df* or the PACF method. Controlling for PM_{10} in

Table 26. Canada: Percentage Change in Respiratory Admissions ≥ 65 Years per 10-μg/m^3 Increase in PM_{10}, Lag 1

Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} results		
3 *df*/year	1.20 (−0.04, 2.40)	1.40 (0.18, 2.60)
8 *df*/year	−0.46 (−1.70, 0.80)	−0.05 (−1.30, 1.20)
12 *df*/year	−0.23 (−1.50, 1.00)	−0.11 (−1.30, 1.10)
PACF	1.20 (−0.04, 2.40)	1.20 (0.01, 2.40)
Controlling for O_3		
3 *df*/year	0.9 (−0.36, 2.2)	1.1 (−0.17, 2.3)
8 *df*/year	−0.67 (−2.0, 0.64)	−0.19 (−1.5, 1.1)
12 *df*/year	−0.33 (−1.6, 1.0)	−0.2 (−1.5, 1.1)
PACF	0.01 (−1.2, 1.4)	0.6 (−0.65, 1.9)

Table 27. Canada: Percentage Change in Cardiovascular Disease Admissions ≥ 65 Years per 10-μg/m^3 Increase in PM_{10}, Lag 1

Seasonality Control	Penalized Splines % (95%CI)	Natural Splines % (95%CI)
PM_{10} results		
3 *df*/year	0.36 (−0.34, 1.1)	0.39 (−0.31, 1.1)
8 *df*/year	0.13 (−0.61, 0.89)	0.21 (−0.56, 0.99)
12 *df*/year	0.075 (−0.7, 0.86)	0.044 (−0.75, 0.85)
PACF	0.3 (−0.4, 1)	0.39 (−0.31, 1.1)
Controlling for O_3		
3 *df*/year	0.082 (−0.65, 0.82)	0.12 (−0.62, 0.87)
8 *df*/year	−0.066 (−0.85, 0.73)	−0.011 (−0.82, 0.8)
12 *df*/year	−0.1 (−0.92, 0.72)	−0.086 (−0.92, 0.75)
PACF	0.06 (−0.68, 0.8)	0.12 (−0.62, 0.87)

Table 28. Canada: Percentage Change in Respiratory Admissions ≥ 65 Years per 10-μg/m^3 Increase in O_3

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 results						
3 *df*/year	2.7 (1.8, 3.6)	1.9 (0.95, 2.9)	2 (1.2, 2.7)	1.3 (0.55, 2.1)	3.3 (1.3, 5.3)	3.6 (1.6, 5.7)
8 *df*/year	1.1 (0.1, 2)	1 (0.031, 2)	0.74 (−0.048, 1.5)	0.69 (−0.12, 1.5)	2.2 (0.28, 4.1)	2.4 (0.51, 4.4)
12 *df*/year	0.78 (−0.2, 1.8)	0.59 (−0.39, 1.6)	0.53 (−0.26, 1.3)	0.4 (−0.41, 1.2)	2 (0.17, 4)	1.9 (0.012, 3.8)
PACF	2.1 (1.2, 3.1)	1.5 (0.57, 2.5)	1.5 (0.74, 2.3)	1.1 (0.32, 1.9)	2.2 (0.35, 4.2)	2.2 (0.27, 4.1)
Controlling for PM_{10}						
3 *df*/year			1.3 (−0.071, 2.8)	1.1 (−0.41, 2.5)		
8 *df*/year			0.39 (−1.1, 1.9)	0.64 (−0.87, 2.2)		
12 *df*/year			0.086 (−1.4, 1.6)	−0.18 (−1.7, 1.4)		
PACF			1 (−0.4, 2.5)	1.3 (−0.16, 2.8)		

Table 29. Canada: Percentage Change in Cardiovascular Disease Admissions ≥ 65 Years per 10-μg/m^3 Increase in O_3

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 results						
3 *df*/year	0.55 (0.052, 1.1)	0.51 (−0.01, 1)	0.55 (0.13, 0.98)	0.53 (0.094, 0.97)	0.34 (−0.73, 1.4)	0.51 (−0.57, 1.6)
8 *df*/year	0.17 (−0.39, 0.74)	−0.028 (−0.6, 0.54)	0.26 (−0.2, 0.73)	0.11 (−0.35, 0.58)	0.28 (−0.79, 1.4)	0.11 (−0.97, 1.2)
12 *df*/year	0.077 (−0.5, 0.66)	0.067 (−0.52, 0.66)	0.19 (−0.28, 0.66)	0.17 (−0.3, 0.65)	0.27 (−0.81, 1.4)	0.35 (−0.75, 1.5)
PACF	0.51 (−0.0033, 1)	0.49 (−0.031, 1)	0.53 (0.1, 0.96)	0.53 (0.089, 0.97)	0.35 (−0.71, 1.4)	0.51 (−0.57, 1.6)
Controlling for PM_{10}						
3 *df*/year			0.96 (0.16, 1.8)	0.99 (0.16, 1.8)		
8 *df*/year			0.63 (−0.22, 1.5)	0.68 (−0.2, 1.6)		
12 *df*/year			0.5 (−0.38, 1.4)	0.62 (−0.83, 2.1)		
PACF			1 (0.2, 1.8)	0.99 (0.16, 1.8)		

Table 30. Canadian O_3 Summer-Only Analysis for Ages ≥ 65[a]

	Average of Lags 0–1 % (95% CI)	Lag 1 % (95% CI)	Distributed Lags % (95% CI)
Respiratory admissions	3.6 (2.2, 5.1)	2.5 (1.8, 3.3)	4.1 (1.4, 6.8)
Cardiovascular disease admissions	−0.23 (−0.71, 0.26)	−0.07 (−0.47, 0.34)	−0.24 (−0.59, 0.11)

[a] O_3 not controlled for PM_{10}.

Table 31. Europe: Percentage Change in Respiratory Admissions ≥ 65 Years per 10-μg/m^3 Increase in PM_{10}

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} results						
3 *df*/year	1.14 (0.77, 1.51)	0.74 (0.20, 1.27)	0.98 (0.65, 1.31)	0.50 (0.03, 0.97)	1.87 (1.27, 2.46)	1.36 (0.27, 2.47)
8 *df*/year	0.60 (0.25, 0.95)	0.16 (−0.21, 0.54)	0.30 (−0.14, 0.74)	0.01 (−0.41, 0.44)	0.89 (0.00, 1.79)	0.59 (−0.39, 1.59)
12 *df*/year	0.58 (0.24, 0.93)	0.50 (0.10, 0.90)	0.20 (−0.30, 0.70)	0.11 (−0.45, 0.67)	0.64 (−0.67, 1.97)	1.05 (−0.15, 2.27)
PACF	0.89 (0.43, 1.34)	0.49 (0.13, 0.85)	0.49 (0.06, 0.91)	0.20 (−0.25, 0.64)	1.21 (0.36, 2.07)	1.03 (0.07, 1.99)
Controlling for O_3						
3 *df*/year	1.17 (0.80, 1.54)	0.79 (0.39, 1.18)	1.01 (0.67, 1.34)	0.59 (0.21, 0.98)		
8 *df*/year	0.53 (0.18, 0.89)	0.17 (−0.20, 0.55)	0.32 (−0.12, 0.75)	0.05 (−0.34, 0.45)		
12 *df*/year	0.51 (0.16, 0.86)	0.51 (0.15, 0.87)	0.22 (−0.27, 0.72)	0.15 (−0.37, 0.68)		
PACF	0.81 (0.33, 1.29)	0.43 (0.01, 0.85)	0.48 (0.04, 0.91)	0.16 (−0.32, 0.64)		

the lag 1 models increased the effect estimates, though those models with 8 or 12 *df* remained insignificant. Using DL models, the percentage change in cardiovascular disease admissions was positive, but insignificant.

Table 30 shows the summer-only analysis for O_3 effects on cardiovascular and respiratory admissions using lag 0–1, lag 1, and DL models. Effect estimates for cardiovascular admissions were all negative; those for respiratory admissions were positive and significant for all three models.

Europe

PM In Table 31, the estimated effects of PM_{10} on respiratory admissions for the eight European cities with data are shown. The effects estimated were higher for lag 0–1 compared with lag 1 and, as expected, the effect estimates generally decreased as the degrees of freedom for controlling seasonality in the model increased. The effects were positive, but reached statistical significance consistently only for lag 0–1. Adjusting for O_3 did not change the magnitude or significance of the estimated effects. The DL effects tended to be larger and closer to significance, though remained insignificant in models using 8 or 12 *df*/year.

Table 32 shows the effects of PM_{10} on cardiovascular disease admissions. Again the effects of lags 0–1 and DL (lags 0–2) were larger and consistently statistically significant. Adjusting for O_3 left the lag 1 and lag 0–1 estimates unchanged.

O_3 Table 33 displays the effects of O_3 on respiratory admissions. The one-pollutant model effects were positive in

Table 32. Europe: Percentage Change in Cardiovascular Disease Admissions ≥ 65 Years per 10-μg/m^3 Increase in PM_{10}

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} results						
3 *df*/year	1.10 (0.80, 1.40)	0.90 (0.60, 1.30)	0.80 (0.60, 1.10)	0.70 (0.40, 0.90)	1.20 (0.80, 1.60)	1.10 (0.50, 1.70)
8 *df*/year	0.60 (0.20, 1.00)	0.40 (0.00, 0.80)	0.30 (0.10, 0.60)	0.20 (−0.10, 0.50)	0.80 (0.40, 1.30)	0.90 (0.50, 1.30)
12 *df*/year	0.40 (0.00, 0.80)	0.30 (−0.20, 0.70)	0.20 (−0.10, 0.50)	0.10 (−0.20, 0.40)	0.70 (0.30, 1.20)	0.80 (0.30, 1.20)
PACF	0.50 (0.00, 1.00)	0.40 (0.00, 0.70)	0.60 (0.40, 0.90)	0.40 (0.10, 0.60)	1.20 (0.70, 1.60)	1.10 (0.50, 1.60)
Controlling for O_3						
3 *df*/year	1.10 (0.80, 1.40)	0.90 (0.60, 1.30)	0.80 (0.60, 1.00)	0.70 (0.50, 0.90)		
8 *df*/year	0.50 (0.10, 0.90)	0.50 (0.00, 0.90)	0.30 (0.00, 0.60)	0.30 (0.00, 0.60)		
12 *df*/year	0.30 (−0.10, 0.70)	0.30 (−0.10, 0.70)	0.20 (−0.10, 0.50)	0.20 (−0.10, 0.40)		
PACF	0.50 (0.00, 1.00)	0.60 (0.00, 1.10)	0.70 (0.50, 1.00)	0.60 (0.30, 1.00)		

Table 33. Europe: Percentage Change in Respiratory Admissions ≥ 65 Years per 10-μg/m^3 Increase in O_3

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 results						
3 *df*/year	0.47 (−0.09, 1.03)	0.75 (0.09, 1.41)	0.35 (−0.46, 1.15)	0.76 (0.35, 1.17)	−0.22 (−1.11, 0.67)	0.61 (−0.13, 1.36)
8 *df*/year	0.36 (0.08, 0.63)	0.20 (−0.12, 0.52)	0.19 (−0.28, 0.67)	0.37 (0.13, 0.61)	0.01 (−0.67, 0.70)	0.10 (−0.58, 0.79)
12 *df*/year	0.28 (0.04, 0.52)	0.10 (−0.29, 0.49)	0.10 (−0.34, 0.53)	0.28 (0.05, 0.51)	0.00 (−0.70, 0.71)	−0.01 (−0.91, 0.89)
PACF	0.48 (0.20, 0.76)	0.47 (−0.07, 1.02)	0.35 (−0.16, 0.86)	0.54 (0.26, 0.82)	0.52 (−0.22, 1.26)	0.58 (−0.19, 1.36)
Controlling for PM_{10}						
3 *df*/year	0.28 (−0.37, 0.94)	0.59 (0.01, 1.17)	0.44 (0.06, 0.81)	0.68 (0.39, 0.97)		
8 *df*/year	0.11 (−0.29, 0.50)	0.11 (−0.23, 0.45)	0.32 (0.05, 0.60)	0.30 (0.01, 0.58)		
12 *df*/year	−0.05 (−0.37, 0.28)	−0.10 (−0.57, 0.36)	0.25 (−0.01, 0.52)	0.11 (−0.16, 0.38)		
PACF	0.10 (−0.35, 0.55)	0.14 (−0.24, 0.52)	0.35 (0.08, 0.62)	0.35 (0.07, 0.63)		

most cases, but not consistent among different models. Hence, while effects were higher for lags 0–1 using PS, the lag 1 effects were higher when the NS were applied. Adjusting for PM_{10} affects the lag 0–1 estimates only. The effects estimates using DL models were not significant.

O_3 exposure appeared to have no effect on admissions for cardiovascular disease (Table 34).

Table 35 shows the summer-only analysis for O_3 effects for the single pollutant, lag 0–1 and lag 1 models. No effect on cardiovascular disease admissions was observed. Respiratory disease admissions were elevated, but were not statistically significant.

United States

PM Table 36 shows the PM_{10} effects on respiratory admissions. For many of the one-pollutant models, the effect estimates, although positive, were not statistically significant, particularly those using 8 or 12 *df*/year. However, the effect estimates generally increased and became statistically significant with adjustment for O_3. Lag 1 models with 8 or 12 *df* remained nonsignificant.

Table 37 provides the corresponding results for PM_{10} effects on cardiovascular admissions. Significant associations were found for lag 0–1 and for lag 1, and the effect size was higher for lag 0–1 than for lag 1. The effect size decreased slightly as the degrees of freedom for controlling seasonality in the model increased. The DL effect appeared to be larger. After adjusting for O_3, the estimated effects increased slightly.

In a study that examined more cities, we found significant effects of both PM_{10} and O_3 on admissions for pneumonia and COPD (Medina-Ramon et al. 2006).

O_3 The effects of O_3 on respiratory admissions are shown in Table 38. Effect estimates for all one-pollutant models were positive, but mostly statistically insignificant. The most unstable results were yielded by the model with 3 *df*

Table 34. Europe: Percentage Change in Cardiovascular Disease Admissions ≥ 65 Years per 10-μg/m^3 Increase in O_3

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 results						
3 *df*/year	−0.06 (−0.50, 0.39)	0.04 (−0.41, 0.49)	0.11 (−0.24, 0.47)	0.19 (−0.16, 0.54)	−0.09 (−0.70, 0.53)	0.07 (−0.65, 0.80)
8 *df*/year	−0.10 (−0.46, 0.27)	−0.12 (−0.47, 0.24)	0.09 (−0.19, 0.36)	0.06 (−0.19, 0.32)	−0.05 (−0.42, 0.33)	0.00 (−0.40, 0.40)
12 *df*/year	−0.10 (−0.43, 0.23)	−0.08 (−0.41, 0.25)	0.08 (−0.16, 0.33)	0.09 (−0.14, 0.32)	0.05 (−0.27, 0.37)	0.11 (−0.22, 0.45)
PACF	0.01 (−0.29, 0.31)	0.02 (−0.30, 0.33)	0.21 (0.00, 0.42)	0.15 (−0.06, 0.36)	0.13 (−0.23, 0.48)	−0.10 (−0.91, 0.72)
Controlling for PM_{10}						
3 *df*/year	0.49 (0.48, 0.50)	0.42 (0.34, 0.50)	0.35 (0.23, 0.46)	0.42 (0.28, 0.56)		
8 *df*/year	0.64 (0.36, 0.91)	0.65 (0.54, 0.77)	0.53 (0.42, 0.63)	0.58 (0.56, 0.59)		
12 *df*/year	1.12 (1.02, 1.21)	0.66 (0.28, 1.04)	0.56 (0.48, 0.64)	0.47 (0.30, 0.63)		
PACF	0.54 (0.48, 0.59)	0.45 (0.27, 0.64)	0.44 (0.36, 0.52)	0.35 (0.34, 0.37)		

Table 35. European O_3 Summer-Only Analysis for Ages ≥ 65[a]

	Average of Lags 0–1		Lag 1	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
Respiratory admissions	0.05 (−0.41, 0.50)	0.03 (−0.43, 0.49)	0.33 (−0.05, 0.72)	0.31 (−0.08, 0.69)
Cardiovascular disease admissions	−0.20 (−0.80, 0.40)	−0.20 (−0.80, 0.40)	0.00 (−0.50, 0.40)	0.00 (−0.40, 0.40)

[a] O_3 not controlled for PM_{10}.

Table 36. United States: Percentage Change in Respiratory Admissions ≥ 65 Years per 10-μg/m^3 Increase in PM_{10}

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} results						
3 *df*/year	1.22 (0.58, 1.87)	0.66 (0.09, 1.22)	1.01 (0.51, 1.51)	0.38 (0.02, 0.73)	1.06 (0.25, 1.88)	0.57 (−0.10, 1.25)
8 *df*/year	0.27 (−0.09, 0.63)	0.13 (−0.21, 0.47)	0.07 (−0.24, 0.38)	−0.06 (−0.36, 0.24)	0.12 (−0.36, 0.60)	−0.02 (−0.45, 0.41)
12 *df*/year	0.19 (−0.15, 0.52)	0.02 (−0.34, 0.37)	−0.02 (−0.31, 0.27)	−0.13 (−0.43, 0.17)	0.03 (−0.40, 0.46)	−0.34 (−0.85, 0.17)
PACF	1.05 (0.47, 1.63)	0.36 (0.04, 0.69)	0.79 (0.35, 1.23)	0.12 (−0.15, 0.38)	1.28 (0.61, 1.95)	0.37 (−0.26, 1.01)
Controlling for O_3						
3 *df*/year	2.06 (1.58, 2.53)	1.16 (0.67, 1.64)	1.50 (1.11, 1.89)	0.69 (0.29, 1.09)		
8 *df*/year	0.73 (0.30, 1.17)	0.47 (0.02, 0.92)	0.32 (−0.03, 0.67)	0.13 (−0.23, 0.50)		
12 *df*/year	0.52 (0.10, 0.94)	0.42 (−0.02, 0.86)	0.14 (−0.20, 0.48)	0.06 (−0.29, 0.41)		
PACF	1.87 (1.31, 2.44)	0.63 (0.20, 1.07)	1.20 (0.69, 1.72)	0.20 (−0.15, 0.56)		

Table 37. United States: Percentage Change in Cardiovascular Disease Admissions ≥ 65 Years per 10-μg/m^3 Increase in PM_{10}

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
PM_{10} results						
3 *df*/year	1.04 (0.80, 1.29)	0.96 (0.76, 1.17)	0.67 (0.48, 0.86)	0.51 (0.35, 0.68)	1.07 (0.78, 1.37)	1.04 (0.77, 1.31)
8 *df*/year	0.71 (0.52, 0.90)	0.67 (0.47, 0.87)	0.31 (0.15, 0.47)	0.26 (0.09, 0.43)	0.77 (0.49, 1.05)	0.74 (0.46, 1.02)
12 *df*/year	0.61 (0.41, 0.81)	0.58 (0.38, 0.79)	0.21 (0.05, 0.38)	0.18 (0.02, 0.35)	0.63 (0.34, 0.92)	0.58 (0.28, 0.87)
PACF	1.01 (0.80, 1.22)	0.82 (0.64, 1.01)	0.62 (0.46, 0.78)	0.40 (0.24, 0.56)	1.15 (0.72, 1.59)	1.00 (0.56, 1.43)
Controlling for O_3						
3 *df*/year	1.36 (1.06, 1.65)	1.17 (0.86, 1.47)	0.90 (0.65, 1.14)	0.71 (0.46, 0.96)		
8 *df*/year	0.93 (0.62, 1.24)	0.84 (0.53, 1.16)	0.51 (0.25, 0.76)	0.44 (0.18, 0.69)		
12 *df*/year	0.82 (0.50, 1.14)	0.80 (0.48, 1.12)	0.40 (0.15, 0.66)	0.38 (0.12, 0.64)		
PACF	1.23 (0.93, 1.53)	1.00 (0.69, 1.31)	0.74 (0.49, 0.99)	0.53 (0.27, 0.78)		

Table 38. United States: Percentage Change in Respiratory Admissions ≥ 65 Years per 10-μg/m^3 Increase in O_3[a]

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 results						
3 *df*/year	0.07 (−0.90, 1.05)	0.20 (−0.13, 0.54)	0.28 (−0.38, 0.94)	0.28 (0.01, 0.56)	−0.15 (−1.68, 1.40)	−0.24 (−1.26, 0.79)
8 *df*/year	0.18 (−0.12, 0.49)	0.30 (−0.00, 0.61)	0.26 (0.01, 0.51)	0.33 (0.08, 0.58)	0.14 (−0.39, 0.66)	0.42 (0.003, 0.84)
12 *df*/year	0.19 (−0.11, 0.48)	0.22 (−0.09, 0.52)	0.25 (0.25, 0.01)	0.26 (0.02, 0.51)	0.23 (−0.18, 0.65)	0.28 (−0.26, 0.82)
PACF	0.24 (−0.64, 1.12)	0.35 (0.02, 0.68)	0.41 (−0.14, 0.95)	0.36 (0.11, 0.61)	0.09 (−1.45, 1.65)	0.13 (−1.09, 1.37)
Controlling for PM_{10}						
3 *df*/year	0.37 (0.01, 0.72)	−0.14 (−0.66, 0.38)	0.47 (0.17, 0.76)	0.11 (−0.20, 0.42)		
8 *df*/year	0.08 (−0.26, 0.41)	0.18 (−0.17, 0.53)	0.21 (−0.06, 0.49)	0.27 (−0.01, 0.55)		
12 *df*/year	0.12 (−0.21, 0.45)	0.14 (−0.20, 0.48)	0.23 (−0.03, 0.50)	0.25 (−0.02, 0.52)		
PACF	0.36 (0.04, 0.67)	0.20 (−0.13, 0.54)	0.48 (0.22, 0.75)	0.31 (0.03, 0.58)		

[a] −0.00 indicates a value < −0.005.

Table 39. United States: Percentage Change in Cardiovascular Disease Admissions ≥ 65 Years per 10-μg/m^3 Increase in O_3

	Average of Lags 0–1		Lag 1		Distributed Lags	
Seasonality Control	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
O_3 results						
3 *df*/year	0.22 (−0.10, 0.55)	0.17 (−0.04, 0.38)	0.23 (0.05, 0.42)	0.13 (−0.04, 0.31)	0.19 (−0.20, 0.58)	0.04 (−0.24, 0.31)
8 *df*/year	0.10 (−0.11, 0.31)	0.10 (−0.11, 0.31)	0.06 (−0.11, 0.24)	0.06 (−0.11, −0.24)	−0.02 (−0.30, 0.27)	0.01 (−0.29, 0.30)
12 *df*/year	0.04 (−0.17, 0.25)	0.04 (−0.18, 0.26)	0.02 (−0.16, 0.19)	0.02 (−0.16, 0.19)	−0.10 (−0.40, 0.20)	−0.13 (−0.52, 0.27)
PACF	0.29 (0.09, 0.49)	0.19 (−0.02, 0.40)	0.22 (0.05, 0.39)	0.13 (−0.04, 0.31)	0.23 (−0.09, 0.55)	0.07 (−0.21, 0.35)
Controlling for PM_{10}						
3 *df*/year	0.09 (−0.13, 0.31)	−0.06 (−0.29, 0.18)	0.13 (−0.05, 0.31)	0.02 (−0.17, 0.21)		
8 *df*/year	−0.04 (−0.28, 0.19)	−0.03 (−0.27, 0.21)	0.01 (−0.18, 0.20)	0.01 (−0.18, 0.20)		
12 *df*/year	−0.07 (−0.31, 0.17)	−0.08 (−0.32, 0.16)	−0.02 (−0.21, 0.17)	−0.02 (−0.22, 0.17)		
PACF	0.04 (−0.19, 0.26)	0.02 (−0.22, 0.25)	0.09 (−0.10, 0.27)	0.06 (−0.13, 0.25)		

Table 40. United States: Maximum O_3 Summer-Only Analysis for Ages ≥ 65[a]

	Average of Lags 0–1		Lag 1	
	Penalized Splines % (95% CI)	Natural Splines % (95% CI)	Penalized Splines % (95% CI)	Natural Splines % (95% CI)
Respiratory admissions	0.28 (−0.07, 0.62)	0.27 (−0.08, 0.62)	0.35 (−0.01, 0.72)	0.35 (−0.002, 0.71)
Cardiovascular disease admissions	−0.02 (−0.33, 0.30)	−0.02 (−0.35, 0.31)	0.05 (−0.37, 0.47)	0.05 (−0.37, 0.47)

[a] O_3 not controlled for PM_{10}.

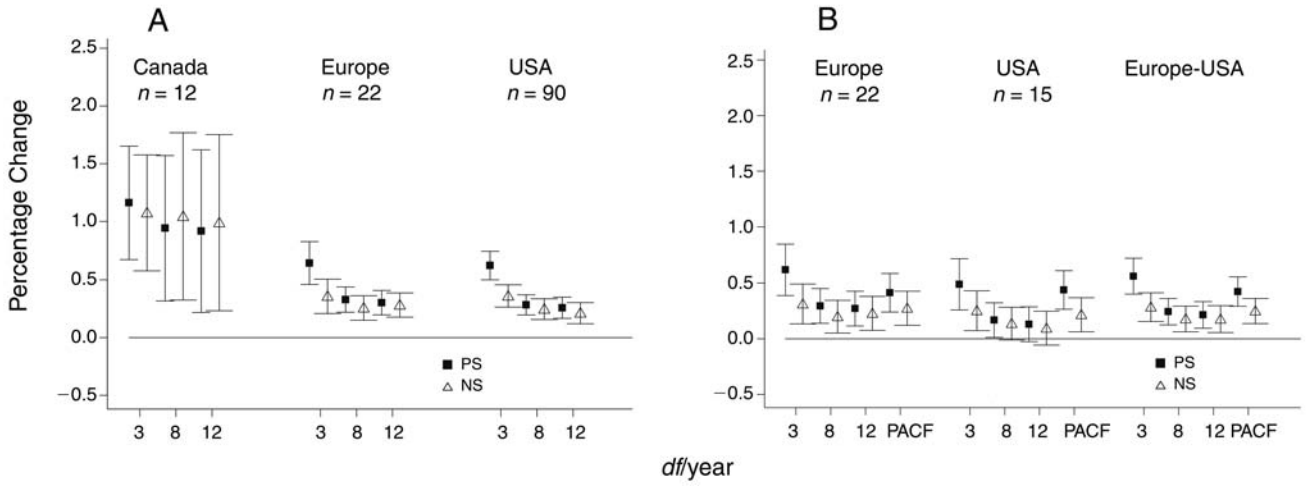


Figure 9. Percentage change in all-cause mortality, all ages, associated with a 10-μg/m^3 PM_{10} increase. A: lag 1, Canada, Europe, and United States; **B**: lag 0–1, Europe, United States, and the two centers combined.

and the model with the degrees of freedom chosen by minimizing the PACF. Adjusting for PM_{10} slightly decreased the estimates for models with 8 or 12 *df*.

For cardiovascular disease admissions, most one-pollutant models gave positive but statistically insignificant results, with the exception of lag 0–1 and lag 1 models using PS and degrees of freedom chosen by minimizing PACF. After adjusting for PM_{10}, the O_3 effects were decreased and became statistically insignificant in all models (Table 39).

Table 40 shows the summer-only analysis for O_3 effects. No effects of O_3 on cardiovascular disease admissions were observed. The estimated percentage increases in respiratory admissions were positive, but statistically insignificant for all models.

COMBINED RESULTS

Figures 9A–13A and 15A show the single-pollutant effects of a 10-μg/m^3 PM_{10} increase for lag 1 on all-cause mortality, cardiovascular mortality, and respiratory mortality for the different age groups in the three centers. The effect sizes for Europe and the United States were comparable with each other, but the effects estimated for Canadian cities were more than 2-fold higher. The CIs for the Canadian data were wider because the population size and number of cities are relatively small compared with those in Europe and the United States. Therefore the investigators concluded that it would not be appropriate to pool the Canadian estimates with those for Europe and the United States. Consequently, the pooling shown in panel B of Figures 9 through 13 involved only the European and U.S. results. All results shown are from single-pollutant models.

Figures 9A and 10A show the effects of PM_{10}, lag 1 on all-cause mortality for all ages (Figure 9A) and for people 75 years or older (Figure 10A). The effects for the older group were higher than those for all ages.

Figure 9B shows the results for all-cause mortality for all ages, and using PM_{10} lag 0–1. Lag 0–1 could be used only for cities for which daily data were collected, in other words, 15 U.S. and all 22 European cities. The estimated increases in all-cause mortality at 8 *df* were 0.25% with PS and 0.18% with NS; at 12 *df* the increases were 0.21% with PS and 0.18% with NS. Using the PACF criterion, the increases were 0.42% with PS and 0.25% with NS. The combined estimates for Europe and the United States (lag 0–1) for people 75 years or older are shown in Figure 10B.

Figure 11A and B show the corresponding increases in all-cause mortality for people younger than 75 years for all three centers, and for Europe and the United States, respectively. Although the sizes of the effects were smaller than for older age groups in all three centers, the combined European-United States effect for lag 0–1 was statistically significant.

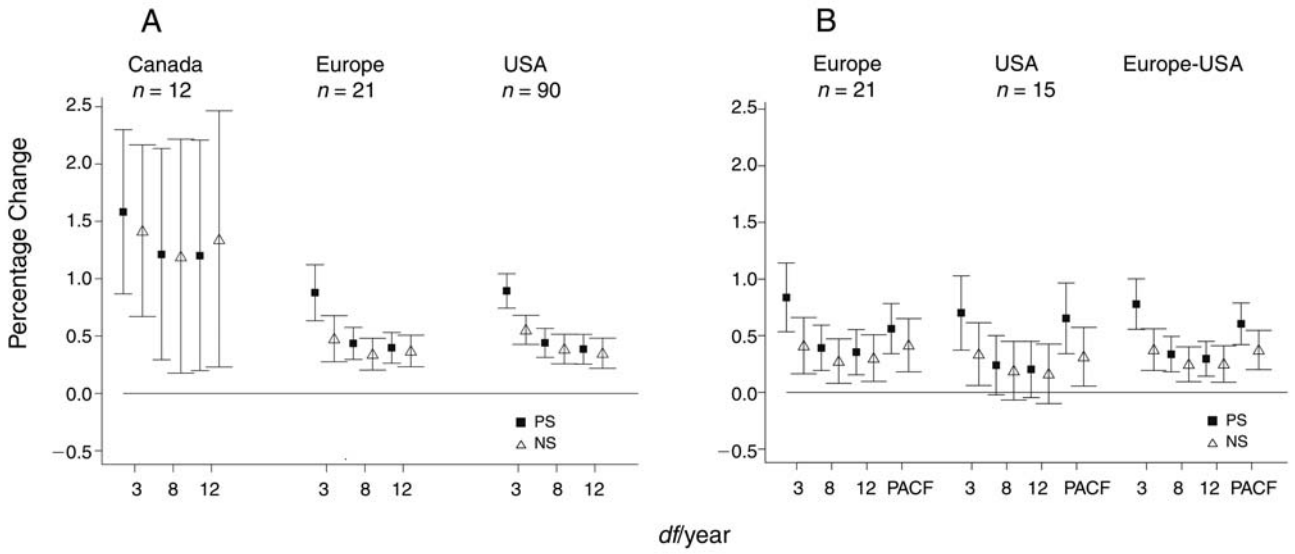


Figure 10. Percentage change in all-cause mortality, for age ≥ 75, associated with a 10-μg/m^3 PM_{10} increase. A: lag 1, Canada, Europe, and United States; **B**: lag 0–1, Europe, United States, and the two centers combined.

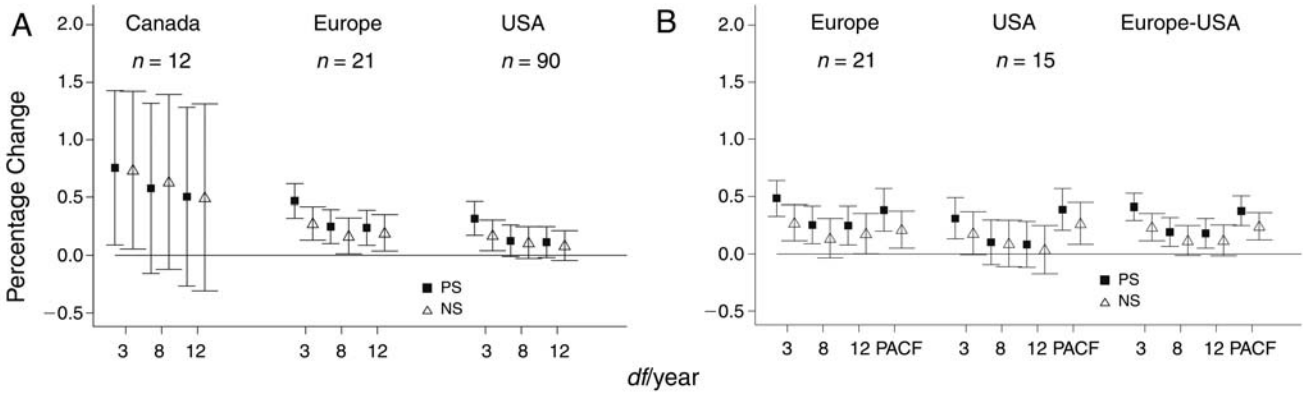


Figure 11. Percentage change in all-cause mortality, for age < 75, associated with a 10-µg/m^3 PM_{10} increase. A: lag 1, Canada, Europe, and United States; **B:** lag 0–1, Europe, United States, and the two centers combined.

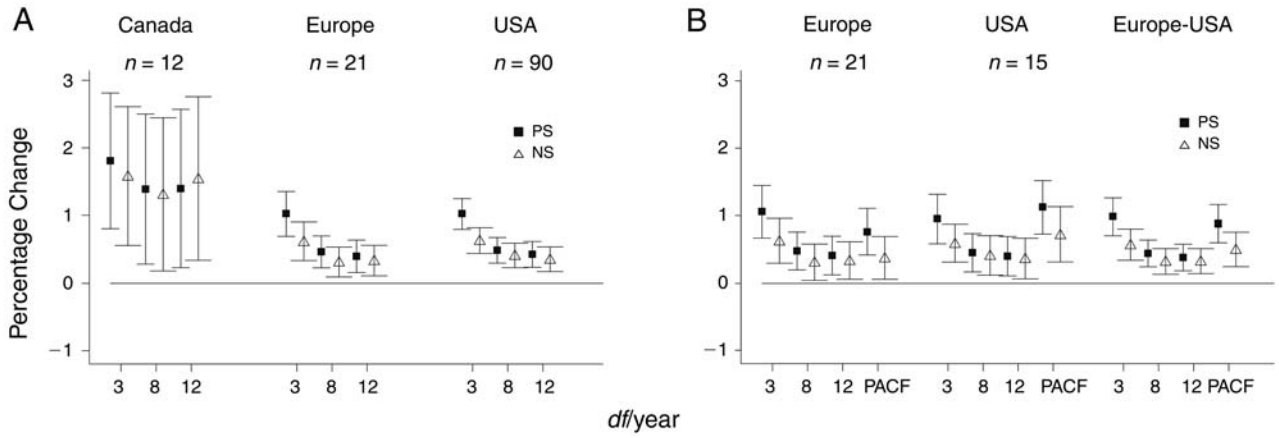


Figure 12. Percentage change in cardiovascular mortality, for age ≥ 75, associated with a 10-µg/m^3 PM_{10} increase. A: lag 1, Canada, Europe, and United States; **B**: lag 0–1, Europe, United States, and the two centers combined.

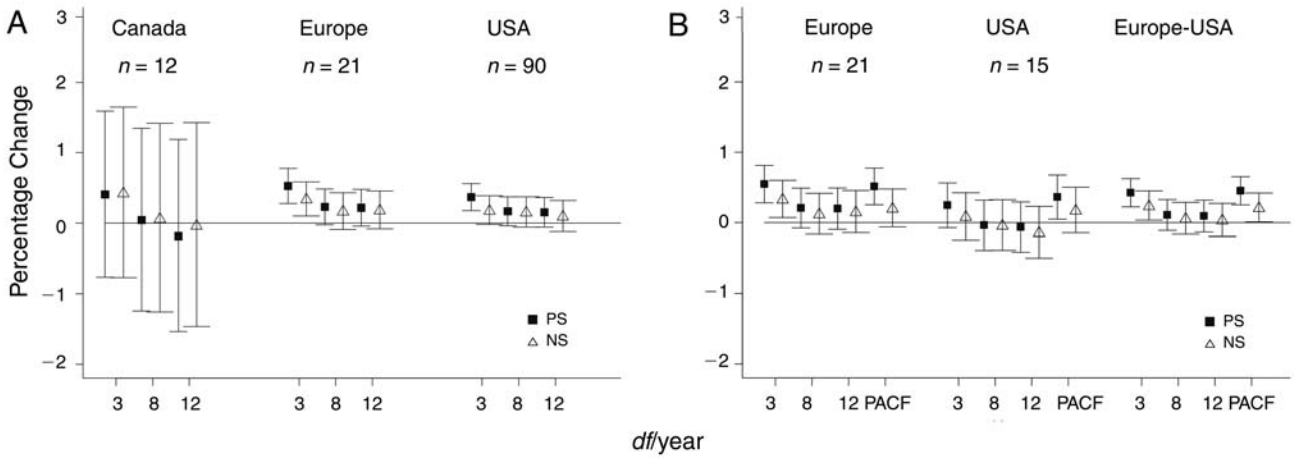


Figure 13. Percentage change in cardiovascular mortality, for age < 75, with a 10-µg/m^3 PM_{10} increase. A: lag 1, Canada, Europe, and United States; **B**: lag 0–1, Europe, United States, and the two centers combined.

Figure 12A and B shows the percent increase in cardiovascular mortality for people 75 years or older per 10-μg/m^3 increase in PM_{10}. The effects were generally higher compared with those for all-cause mortality, and all estimates (center-specific and combined) were statistically significant. In contrast, Figure 13 shows that the effects of PM_{10} on cardiovascular mortality among the younger age group (< 75 years) were close to zero for some of the models and were not statistically significant. The combined effects for Europe and the United States were statistically significant only with 3 *df* or with the PACF criterion.

Figure 14 shows the effects per 10-μg/m^3 PM_{10} lag 0–1 on respiratory mortality for all ages in Europe and the United States, and in the two centers combined. All of the combined estimates using the various models (*df* and splines) were positive and most were statistically significant. Figure 15 shows the corresponding estimates for the number of respiratory deaths among people 75 years or older. For lag 1 (Figure 15A), the effects are positive for the European and U.S. cities, but negative for the Canadian cities. For lag 0–1 (Figure 15B), most of the combined European-United States effects are positive and statistically significant.

The effect of a 10-μg/m^3 increase in O_3 with lag 0–1 on all-cause mortality for all ages is shown in Figure 16. Lag 1 results were similar, but are not shown. Canada did have daily measurements for O_3, so it is included in the effect estimates. These results display similar patterns as those for PM_{10}; the effects for Canada were generally larger with wider CIs. Estimates for Europe were positive while those

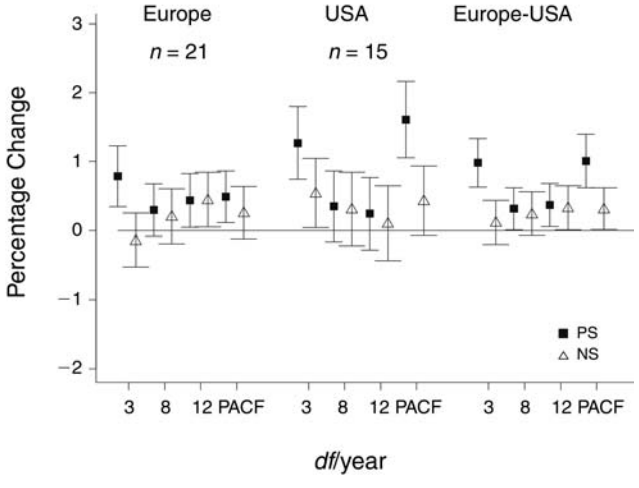


Figure 14. Percentage change in respiratory mortality, for all ages, associated with a 10-μg/m^3 PM_{10} increase (lag 0–1): Europe, United States, and the two centers combined.

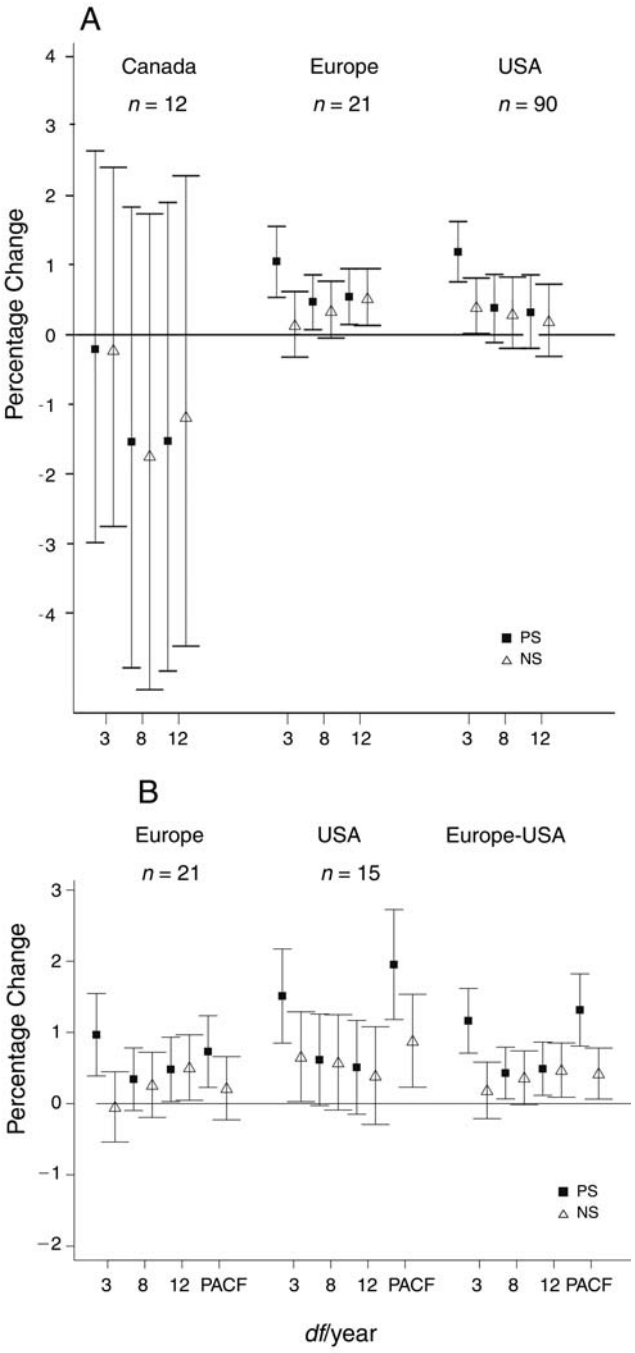


Figure 15. Percentage change in respiratory mortality, for age ≥ 75, with a 10-μg/m^3 PM_{10} increase. A: lag 1, Canada, Europe, and United States; **B**: lag 0–1, Europe, United States, and the two centers combined.

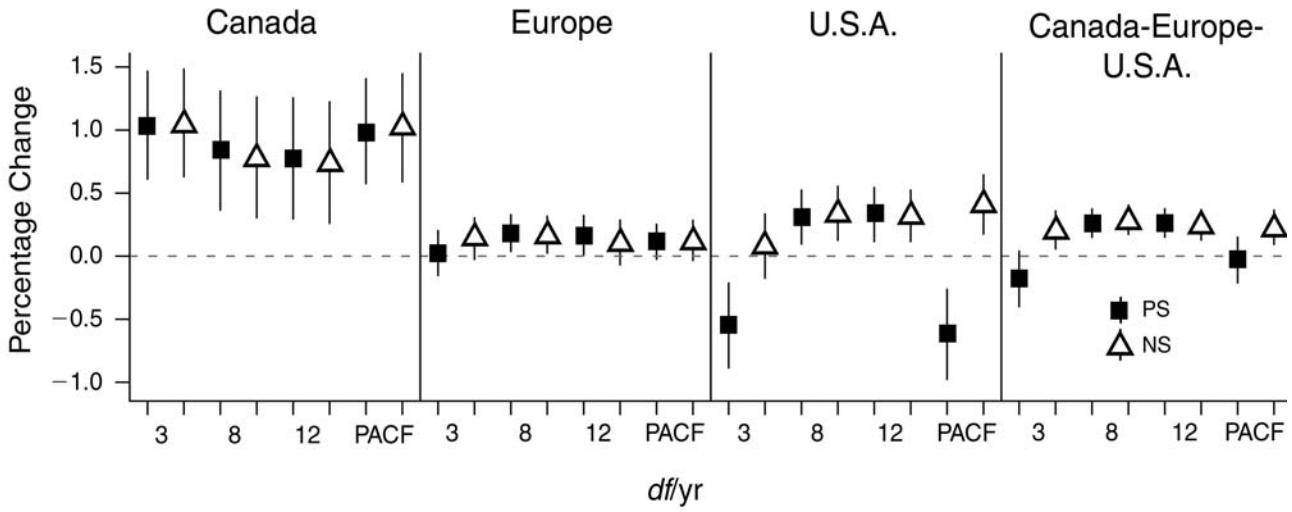


Figure 16. All three centers: Percentage change in all-cause mortality, for all ages, associated with a 10-µg/m^3 O_3 increase (lag 0–1), combined across all cities in each center and across all centers.

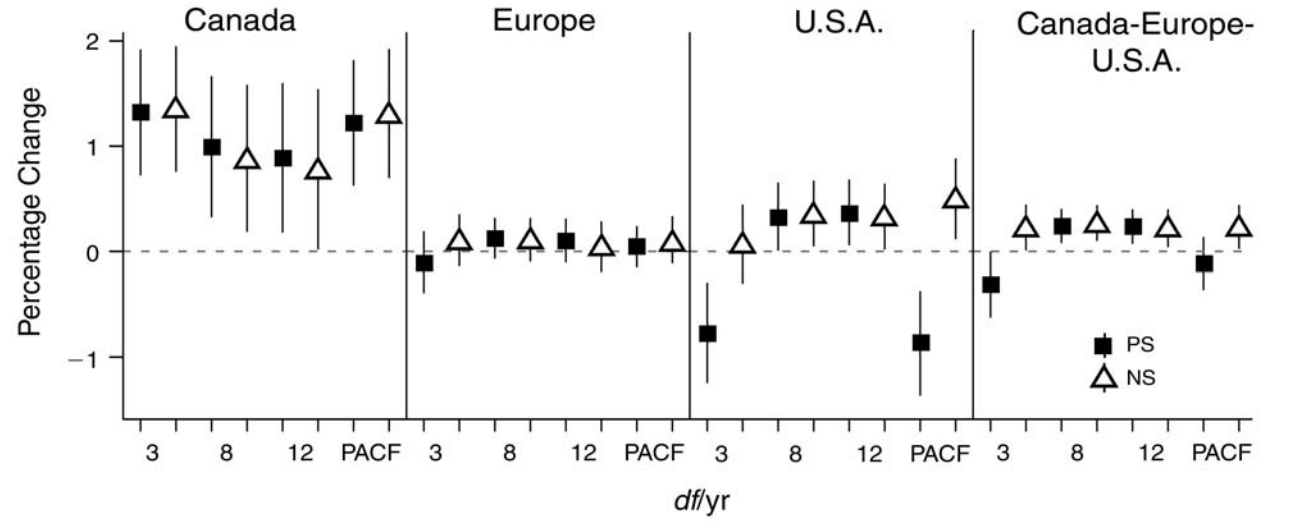


Figure 17. All three centers: Percentage change in all-cause mortality, for age ≥ 75, associated with a 10-µg/m^3 O_3 increase (lag 0–1), combined across all cities in each center and across all centers.

for the United States were mixed. In contrast to the PM_{10} analyses, the overall estimate for the O_3 analyses combined each of the mortality effects across the three centers. Because of the much higher uncertainty in the Canadian estimates, the overall effects largely reflected the contributions of the European and U.S. estimates, with the Canadian estimates contributing little or nothing.

Figure 17 shows the effect of O_3, lag 0–1, on all-cause mortality for people 75 years or older. The pattern of effect estimates across centers and across differing degrees of freedom per year was similar to that for all-cause mortality for all ages shown in the previous figure. In general, the European effect estimates in the older age group were a little lower.

Figure 18 shows the effect of O_3, lag 0–1, on respiratory mortality for all ages in each of the three centers and across all three centers combined. Generally, there was little evidence of an effect of O_3 on respiratory mortality in any

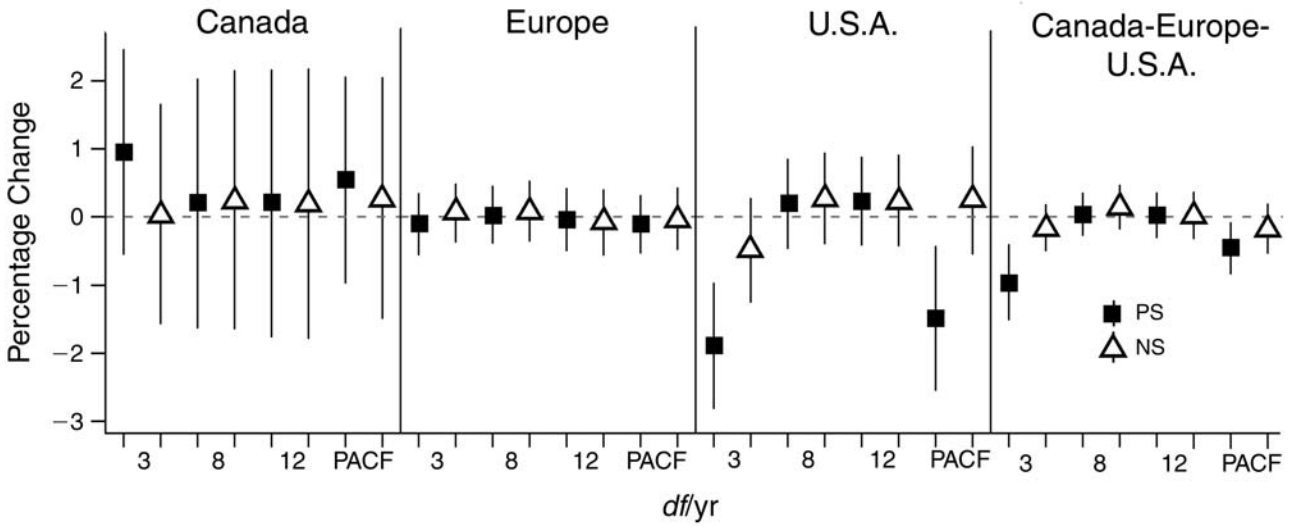


Figure 18. All three centers: Percentage change in respiratory mortality, for all ages, associated with a 10-μg/m^3 O_3 increase (lag 0–1), combined across all cities in each center and across all centers.

center. Most of the effect estimates were not statistically significant. The exceptions were mortality effects estimated for the United States and for all three centers combined using PS and either 3 *df*/year or the PACF criterion. However, the negative effect estimates for the United States may have been due to inadequate adjustment for confounding by smooth seasonal and long-term trends. As for PM_{10}, these results are from single-pollutant models only.

EXPLORATION OF EFFECT MODIFICATION

For detailed information on the effect modifiers used in this analysis for Europe, Canada, and the United States, see Appendix Tables E.1–E.3 (available on the Web).

PM_{10} Effects

Cities with Daily Data Figures 9 and 10 show the PM_{10} combined effects per 10-μg/m^3 increase for lag 0–1 on all-cause mortality for all ages and for people 75 years or older. The analysis included the 21 European cities (22 cities had data for all ages) and 15 U.S. cities with full time-series data on pollution. Although all of the models estimated statistically significant effects of PM_{10} on the outcome, most of the analytic scenarios had statistically significant heterogeneity in the city-specific estimates. As noted, increasing the degrees of freedom for control of seasonality decreased the magnitude of the effect, and consequently also decreased the amount of heterogeneity observed. The European first-stage results were more heterogeneous than those of the United States. However, the European pooled results were more consistent across analytic methods.

Table 41 presents the effect modification patterns of PM_{10} effects, lag 0–1, on all-cause mortality for all ages and for people 75 years or older, as estimated from fixed-effects models; results are for Europe and the United States only. The results are given as a percentage increase in all-cause mortality associated with an increase of 10 μg/m^3 in PM_{10}, at two values of the effect modifier, the 25th and the 75th percentiles of its distribution (i.e., the interquartile range). The estimates can be interpreted as showing the PM_{10} effects in a city characterized by a level of the effect modifier at the 25th percentile of the distribution, and in another city with a level of the effect modifier at the 75th percentile. For several effect modifiers, the distributions differ in European and U.S. cities. Consequently, the investigators chose to present the European 25th and 75th percentiles when only European cities were included; similarly, they used only U.S. data when only U.S. cities were included. The effect modifiers selected by the APHENA group collaborators are listed in Table 41; they are included in a univariate fashion in the models, but not all cities had data available for all effect modifiers.

In general, the effect modification patterns were consistent across methods, particularly for those modifiers having a significant effect. As expected, when more degrees of freedom were used for seasonality control — and consequently when the PM_{10} effects were lower and less heterogeneity was observed — the evidence for effect modification was weaker. The effect modification patterns identified in Europe and the United States were not always consistent. Table 41 summarizes the results for those effect modifiers that were significant in 4 out of 8 models for at least one center. The main results related to effect modification are

summarized below for the PM_{10} effects on the percentage increases in the total number of deaths for all ages and for people 75 years or older:

- Characteristics of exposure: in Europe, higher levels of NO_2 and a larger ratio of NO_2/PM_{10} were associated with a greater PM_{10} effect on mortality. This pattern was present but less pronounced in the United States. By contrast, in cities with higher O_3 levels, a smaller PM_{10} effect on mortality in the older age group was found, a pattern that was much more pronounced in the U.S. cities.
- Climate: Higher temperature and lower humidity were associated with larger PM_{10} effects only in Europe. No consistent pattern of effect modification with temperature was evident in the United States; and the association with humidity tended to be inverse to that in Europe, with higher humidity associated with larger PM_{10} effects.
- Age structure and health status: An increasing proportion of older persons were associated with higher PM_{10} effects in both Europe and the United States. A larger proportion of cardiorespiratory deaths among all deaths were associated with higher PM_{10} effects only in the United States, and only among the older age group. In Europe, the pattern was nonsignificant and tended to be inverse. A higher crude mortality rate was associated with a higher PM_{10} effect in the United States.
- Unemployment: A higher unemployment rate was associated with higher estimates in both regions.

O_3 Effects

Table 42 provides similar results for O_3 for all ages combined. The patterns of effect modification are inconsistent across the three groups of data. There was an indication of significant effect modification for a number of variables in the United States, but not in Europe or Canada.

The investigators compared the Berkey metaregression method with the TLNISE hierarchical model method in the effect modification analysis. Using the relative risk estimates for O_3, they applied both methods to each of the effect modifiers (Table 43). The methods were consistent with each other and produced similar estimates for each of the second stage variables.

Table 41. Percentage Change in All-Cause Mortality Associated with an Increase of 10 μg/m^3 in PM_{10} (Lag 0–1) at the 25th and 75th Percentile of the Center-Specific Distribution of Selected Effect Modifiers[a]

	Europe		United States	
Effect Modifier	25th Percentile Estimate (95% CI)	75th Percentile Estimate (95% CI)	25th Percentile Estimate (95% CI)	75th Percentile Estimate (95% CI)
All ages				
% Population ≥ 65 years	0.25 (0.12, 0.38)	0.31 (0.18, 0.45)	0.06 (−0.11, 0.24)	0.23 (0.08, 0.37)
% Population ≥ 75 years	0.25 (0.11, 0.38)	0.32 (0.18, 0.47)	0.03 (−0.17, 0.22)	0.24 (0.09, 0.39)
Crude mortality rate	0.31 (0.18, 0.44)	0.24 (0.10, 0.38)	−0.13 (−0.37, 0.11)	0.29 (0.14, 0.44)
Mean NO_2	0.17 (0.03, 0.31)	0.44 (0.28, 0.61)	0.01 (−0.26, 0.27)	0.28 (0.12, 0.45)
Mean NO_2/PM_{10}	0.19 (0.05, 0.33)	0.42 (0.25, 0.59)	0.16 (−0.05, 0.37)	0.27 (0.10, 0.44)
Mean temperature	0.15 (0.00, 0.30)	0.38 (0.24, 0.51)	0.22 (0.07, 0.36)	0.16 (−0.01, 0.34)
Mean humidity	0.38 (0.24, 0.51)	0.23 (0.11, 0.35)	−0.03 (−0.30, 0.23)	0.26 (0.11, 0.42)
Percent unemployed	0.27 (0.09, 0.46)	0.57 (0.37, 0.77)	0.11 (−0.13, 0.36)	0.23 (0.08, 0.38)
Ages ≥ 75 years				
% Population ≥ 75 years	0.31 (0.14, 0.47)	0.43 (0.24, 0.63)	0.08 (−0.19, 0.36)	0.33 (0.13, 0.54)
Crude mortality rate	0.38 (0.21, 0.55)	0.35 (0.16, 0.54)	−0.08 (−0.43, 0.26)	0.39 (0.18, 0.60)
% Cardiovascular and respiratory mortality	0.45 (0.27, 0.64)	0.25 (0.04, 0.46)	0.29 (0.08, 0.49)	0.38 (0.17, 0.60)
Mean NO_2	0.22 (0.05, 0.40)	0.60 (0.39, 0.82)	0.27 (−0.11, 0.65)	0.40 (0.18, 0.63)
Mean O_3	0.40 (0.22, 0.58)	0.33 (0.17, 0.49)	0.52 (0.25, 0.78)	0.12 (−0.20, 0.44)
Mean NO_2/PM_{10}	0.29 (0.13, 0.45)	0.56 (0.34, 0.78)	0.36 (0.05, 0.68)	0.39 (0.15, 0.62)
Mean temperature	0.23 (0.03, 0.43)	0.47 (0.29, 0.65)	0.31 (0.10, 0.51)	0.30 (0.05, 0.55)
Mean humidity	0.50 (0.32, 0.69)	0.31 (0.16, 0.46)	−0.03 (−0.42, 0.37)	0.38 (0.16, 0.59)
Percent unemployed	0.37 (0.13, 0.62)	0.69 (0.43, 0.95)	0.05 (−0.29, 0.39)	0.37 (0.15, 0.58)

[a] Lag 0–1 was estimated by using penalized splines and 8 *df*/year to control for seasonal patterns. The variables selected were those displaying significant effect modification in at least 4 out of 8 models applied in at least one center (i.e., Europe or United States).

FURTHER ANALYSES

Comparison with Results from a Case-Crossover Analysis

Although Poisson time-series have been most widely used to analyze the short-term effects of air pollution on mortality, case-crossover analyses are increasingly being used for this purpose. Although a basic case-crossover analysis is formally similar to a Poisson regression with dummy variables for months, this similarity does not hold true for more complex designs. The case-crossover approach has the advantages of familiarity to more general audiences (because of its similarity to case-control studies), and of a

Table 42. Percentage Change in All-Cause Mortality, for All Ages, Associated with an Increase of 10 μg/m^3 in O_3 (Lag 0–1) at the 25th and 75th Percentile of the Center-Specific Distribution of Selected Effect Modifiers[a]

	Canada			Europe			United States		
Effect Modifier[b]	25th Percentile Estimate (95% CI)	75th Percentile Estimate (95% CI)	*t* Value	25th Percentile Estimate (95% CI)	75th Percentile Estimate (95% CI)	*t* Value	25th Percentile Estimate (95% CI)	75th Percentile Estimate (95% CI)	*t* Value
NO_2 CV	0.39 (0.24, 0.55)	0.50 (0.30, 0.70)	1.33	0.21 (0.09, 0.33)	0.17 (−0.01, 0.35)	−0.49	0.16 (0.06, 0.25)	0.01 (−0.10, 0.12)	−2.87
Mean SO_2	0.28 (0.09, 0.48)	0.59 (0.37, 0.82)	2.16	0.20 (0.06, 0.33)	0.21 (0.05, 0.36)	0.16	0.06 (−0.06, 0.18)	0.25 (0.14, 0.37)	2.79
O_3 CV	0.36 (0.10, 0.63)	0.44 (0.27, 0.61)	0.60	0.33 (0.19, 0.47)	0.14 (0.03, 0.25)	−2.65	0.02 (−0.09, 0.14)	0.19 (0.09, 0.28)	2.68
Mean NO_2/PM_{10}	0.49 (0.32, 0.66)	0.32 (0.12, 0.52)	−1.58	0.22 (0.11, 0.34)	0.19 (0.06, 0.33)	−0.43	−0.01 (−0.13, 0.12)	0.16 (0.06, 0.26)	2.64
Mean temperature	0.36 (0.12, 0.59)	0.44 (0.28, 0.61)	0.83	0.20 (0.05, 0.36)	0.20 (0.04, 0.35)	−0.04	0.27 (0.17, 0.37)	0.00 (−0.10, 0.10)	−4.40
% ≥ 75 years	0.28 (0.10, 0.45)	0.53 (0.38, 0.69)	2.68	0.19 (0.07, 0.31)	0.23 (0.07, 0.39)	0.52	0.13 (0.03, 0.24)	0.13 (0.04, 0.22)	−0.02
Age standardized mortality	0.33 (0.10, 0.56)	0.51 (0.28, 0.73)	1.14	0.14 (−0.02, 0.30)	0.25 (0.10, 0.41)	1.07	−0.00 (−0.12, 0.11)	0.20 (0.11, 0.30)	3.81
% unemployed	0.35 (0.18, 0.51)	0.47 (0.32, 0.61)	1.88	0.18 (−0.06, 0.42)	0.17 (−0.06, 0.40)	−0.07	0.02 (−0.10, 0.15)	0.19 (0.09, 0.29)	2.45

[a] Lag 0–1 was estimated by using 8 *df*/year to control for seasonal patterns and penalized splines. The variables selected were those displaying significant effect modification in at least 4 out of 8 models applied in at least one center (i.e., Europe or United States). −0.00 indicates a value < −0.005.

[b] CV indicates coefficient of variation.

Table 43. Effect Modification Analysis: Comparison of MLE-Berkey and TLNISE Methods for O_3 Outcomes from City-Level Regressions Using 8 *df* and PS

	MLE-Berkey			TLNISE		
Effect Modifier	25th Percentile	75th Percentile	*t* Value	25th Percentile	75th Percentile	*t* Value
NO_2 coefficient of variation	0.151	0.010	−2.671	0.156	0.010	−2.870
Mean SO_2	0.056	0.252	2.534	0.058	0.253	2.785
O_3 coefficient of variation	0.027	0.180	2.473	0.024	0.185	2.683
Mean NO_2/PM_{10}	−0.005	0.155	2.330	−0.007	0.159	2.639
Mean temperature	0.269	0.001	−4.240	0.269	0.001	−4.398
% ≥ 75 years	0.119	0.120	0.016	0.132	0.132	−0.022
Age standardized mortality	−0.003	0.199	3.159	−0.136	0.138	3.811
% unemployed	0.027	0.184	2.212	0.023	0.191	2.454

more intuitive approach to handling seasonal confounding than time-series analysis. However, the case-crossover method does not account for overdispersion and may underestimate the variance of an estimator in each city. The design also induces larger estimates for heterogeneity than time-series methods. Therefore, in the discussion below, we consider the utility of the case-crossover design for addressing temporal confounding in a transparent fashion and report examples of its application.

While Poisson regression typically uses regression or PS to control for season, case-crossover approaches use matching to control for season. Matching is generally done by selecting control days in the same month of the same year as the case day. In addition to presenting a less complex analytic approach, the case-crossover design provides clearer indication of the extent to which seasonal and shorter patterns are controlled. On average, a case day would be expected to occur in the middle of the month. By choosing controls in the same month of the same year, control days are separated on average by about one week from the middle of the month, the expectation for the case day. This separation grows somewhat for days near the beginning and end of the month, but in the analyses presented below, the mean difference in time between case days and control days was 9. By presenting the temporal differences between case and control days, the analysis makes clearer the extent to which the investigator is controlling for longer-term patterns.

In contrast, even when splines are explained, the extent to which a spline controls for longer term patterns in the data are not obvious, even to a moderately sophisticated audience, who would [typically] assume that the control is similar to dummy variables for season. To illustrate the extent to which this intuition is misinformed, Figure 19 gives an example in which pneumonia deaths (light line) and cardiovascular deaths (dark line) in Detroit are analyzed with NS and 4 *df*/year and shown over a period of several seasons. The figure shows that these splines capture not only the varying shapes of the winter peaks, but also a double peak in respiratory deaths in one winter (second peak from left). The ability of a spline with 4 *df* to capture seasonal variation, as demonstrated in this figure, may not be immediately apparent, particularly to those without sophisticated understanding of time-series analysis. On the other hand, the matching on season is inherently evident in the design of the case-crossover study.

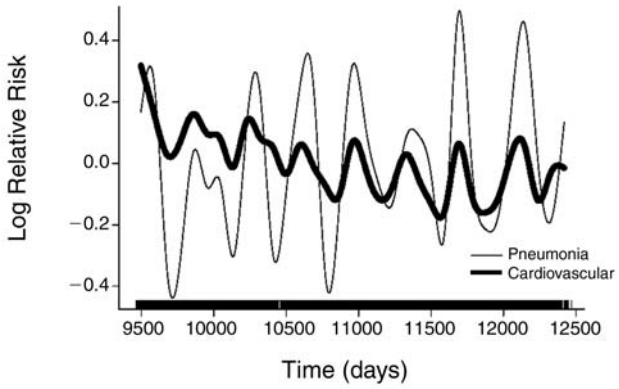


Figure 19. Temporal pattern of pneumonia and cardiovascular mortality in Detroit, Michigan. The black bar along the x-axis indicates data density.

In addition to this strength of transparency, the case-crossover analysis allows matching on other covariates besides season. For example, Schwartz (2004) matched the control days by month and by degree of temperature in a multicity study examining the effects of PM_{10} on mortality. Matching to the same degree of temperature assures that confounding is controlled, even if the association with temperature is highly nonlinear. Moreover, since both month and temperature are matched, this approach also controls for any interaction between them. Thus, any variation in the effect of temperature by month is also controlled. Such an approach provides considerable reassurance that any observed association is not confounded.

In other analyses, Schwartz subsequently applied the same approach to matching on gaseous copollutants (Schwartz 2004) and to control for temperature confounding in the association between O_3 and mortality (Schwartz 2005). Specifically, he analyzed data from the 14 cities included in the APHENA analysis using case-crossover methods. In a baseline analysis, a 10-μg/m^3 increase in PM_{10} on the previous day was associated with a 0.36% (95% CI; 0.22 to 0.50) increase in daily deaths. This estimate is comparable to that of the NMMAPS analysis. When the analysis was repeated with matching of control days on temperature (same temperature with rounded °C) while controlling for the previous day's temperature, the results were unchanged (0.39% [0.19 to 0.58] increase per 10-μg/m^3). With matching on temperature, the standard deviation of the within-strata difference in temperature on the previous day was less than 2°C, making it unlikely that the spline for the previous day's temperature was inadequate. Adding more lags for temperature in the winter (out to lag 3) similarly had little effect (0.39% [0.23 to 0.56]). Matching on CO, O_3, NO_2, or SO_2 also did not diminish the association. These results indicate that the reported associations are not due to confounding by temperature or gaseous copollutants.

As part of APHENA, Schwartz analyzed data from Athens, London, Paris, Madrid, Rome, and Stockholm using the case-crossover approach. The random effects meta-analysis for these cities indicated that a 10-μg/m^3 increase in PM_{10} was associated with a 0.60% increase in

daily deaths (95% CI; 0.28 to 0.93). This estimate is somewhat larger than the overall results for APHEA, but the six selected cities also had somewhat larger effect estimates than average in the Poisson regressions.

Investigation of Concentration–Response and Threshold Analysis for APHENA

Background Most analyses of risk to health from air pollution are based on the assumption that the concentration–response relationship is linear. As air pollution levels have declined, policymakers have sought to better understand the concentration–response relationship between air pollution and various adverse health effects, particularly questioning if there are thresholds of air pollution levels below which there is no risk. The existence of thresholds and nonlinear concentration–response relationships would have substantial implications for standard-setting and for the estimated burden of disease from air pollution and, correspondingly, for the estimated benefits of control policies.

The assumption that the concentration–response relationship is linear can result in biased estimates if a nonlinear relationship actually holds, and weaken analyses directed at a possible threshold. The multicity data sets, which have provided the most reliable and valid effect estimates, facilitate the exploration of differences among single-city concentration-response curves and make their combination possible, if appropriate. The methodology for producing and combining concentration–response curves presents difficulties and has received considerable attention recently. Several methods have been developed and proposed for the estimation and the combination of concentration–response curves. One method, developed by Schwartz and Zanobetti (2000), combines individual city-specific nonparametric smooth curves (metasmooth method). In the metasmooth approach, a smooth function of the pollutant is entered in the individual city model. The predicted values of the log-relative risk of the outcome in each city for 5 $\mu g/m^3$ increments, along with their standard errors, are computed. These predicted values at each increment are then combined across cities, using inverse variance weighting, to produce an overall concentration–response curve.

In NMMAPS, a method which combines individual city estimates of natural cubic spline (NCS)-shaped concentration–response relationships was applied (Daniels et al. 2000). Dominici and colleagues (2002a) presented an improvement to this method in which the parameters of the spline concentration–response curves and the number and location of the knots (assumed to be the same across cities within a region, but allowed to vary from region to region) were estimated jointly. The methods developed within NMMAPS used Bayesian procedures for the estimation of the parameters. All the above methods have been applied to estimate the PM_{10} daily-number-of-deaths relationship in the U.S. cities. A linear relationship without threshold was found.

In APHEA2, unrestricted cubic splines, rather than NCS, were used to fit individual city concentration–response curves. The knots were prespecified and the same for each city. This approach has the advantage of resulting in pooling of similar terms in the second stage of the analysis. The number and location of the knots were determined according to findings of exploratory graphical analysis. In the second stage of the analysis, the city-specific air pollution effect estimates that had been produced from the first stage of the analysis were regressed on city-specific covariates to obtain the overall concentration–response curve and to explore potential heterogeneity in the city-specific curves (Samoli et al. 2003, 2005). Multivariate second-stage regression models based on the method described by Berkey and associates (1998) were fitted. This latter method has the advantage of allowing the investigation of effect modification on the concentration–response shape, which is very useful given the heterogeneity observed between city curves.

In APHEA, the metasmooth method developed by Schwartz and Zanobetti (2000) was also applied. Samoli and coworkers (2003) presented a comparison of the metasmooth and the cubic spline methods used in the APHEA investigation of the concentration–response relationship between NO_2 and all-cause mortality.

Finally, another frequently applied method in the analysis of concentration–response curves (applied in both APHEA and NMMAPS) is the use of threshold models, which allow for the investigation of potential threshold levels.

Description of Threshold Analysis The APHENA investigators used threshold analysis to investigate the concentration–response relationships between PM_{10} and O_3 and all-cause mortality for all ages. They applied the threshold analysis to the models that used NS to control for the confounders and 8 *df*/year to control for seasonality.

The investigators selected a grid of hypothetical threshold values, ranging from 0 through 75 $\mu g/m^3$ (ppb in Canadian analyses of O_3) by increments of 5 (e.g., 0, 5, 10, … 75). For each threshold value h a threshold model was fit to the available cities. In the threshold model, a pollutant term was included in the model of the form $(\text{pollutant} - h)^+$, where $x^+ = x$ if $x \geq 0$ and $x^+ = 0$ if $x < 0$ and h is the threshold value. They then computed the deviance

or AIC of the fitted model (McCullagh and Neldo 1983), where:

$$\text{AIC} = \text{deviance} + 2(\text{number of parameters}). \quad (12)$$

After the AIC/deviance values were computed for all cities at a given threshold value, the average AIC/deviance for that given threshold was computed over all cities. Computations were repeated for all threshold values. Then the mean AIC/deviance versus the threshold values was plotted. A possible threshold is at the value that minimizes the mean AIC/deviance. For the model comparisons in this, computing the deviance is equivalent to computing the AIC, because all applied within-city threshold models have the same number of parameters.

Results

Canadian Results Because the Canadian PM_{10} data have measurements only for one of every six days, analyses of the concentration–response threshold model were restricted to the complete O_3 data where daily measurements are available. The results of this investigation did not appear to offer convincing evidence of a threshold effect for O_3.

They looked at all-cause, all-age mortality in both the lag 1 model and the model considering the mean of lag 0 and lag 1.

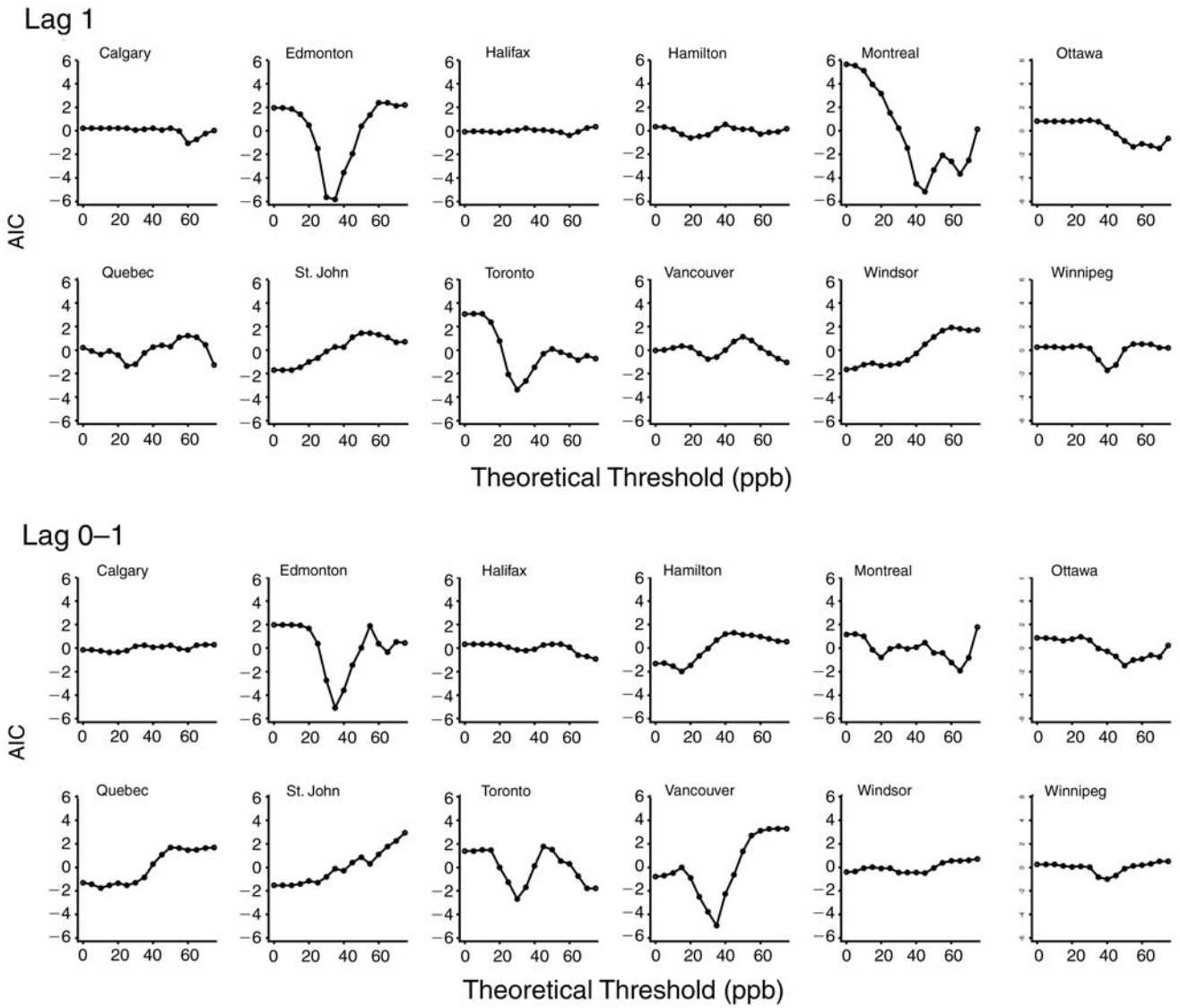


Figure 20. Canada (all ages): Plots of threshold values for O_3 versus AIC values for NS, lag 1 and lag 0–1 models and all-cause mortality in 12 Canadian cities. To facilitate comparison, the curves have been centered.

The analysis procedure consisted of fitting each of these models with a NS with 8 *df* per year for the time trend, but with a threshold. To introduce a threshold *h* into the model, we subtracted the value *h* from each pollution measurement and set all the resulting negative responses to zero. For example, if we wanted to assess a threshold of 20 ppb, the following would occur: any concentration less than 20 would become zero, and any concentration greater than 20 would be scaled by subtracting 20. We did this for each city, for 16 different values of *h*: 0, 5, 10, …, 75 ppb, and computed the AIC for each of the resulting 16 thresholds × 12 cities = 192 models. Figure 20 shows two plots of the results.

In Figure 20, the AIC values are plotted separately for each city as a function of the threshold. Since both the average AIC and the range varied greatly among the different cities, we chose to center the results by subtracting the mean AIC for a city from each of the corresponding cities' AIC values. Overall, these plots do not provide strong evidence for a threshold effect, as there is no consistent pattern over the cities. For example, in St. John the AIC increases monotonically with the threshold, while the AIC in Edmonton shows a distinct minimum at 35 ppb. In Toronto, the city with the largest population and therefore the greatest power to detect a threshold, the results are inconclusive. The Toronto AIC has two local minima, one at 30 and one at 70 ppb.

Figure 21 shows the results of the pooled threshold analysis for the 12 Canadian cities for O_3 all-age, all-cause mortality with lag 1 and lag 0–1. The plot of the AIC means against the theoretical threshold *h* indicates that there might be a threshold at around 30 or 35. The evidence is less convincing, however, with error bars (*t*-distribution, *df* = 11) around the mean AIC (Figure 21). The pattern of

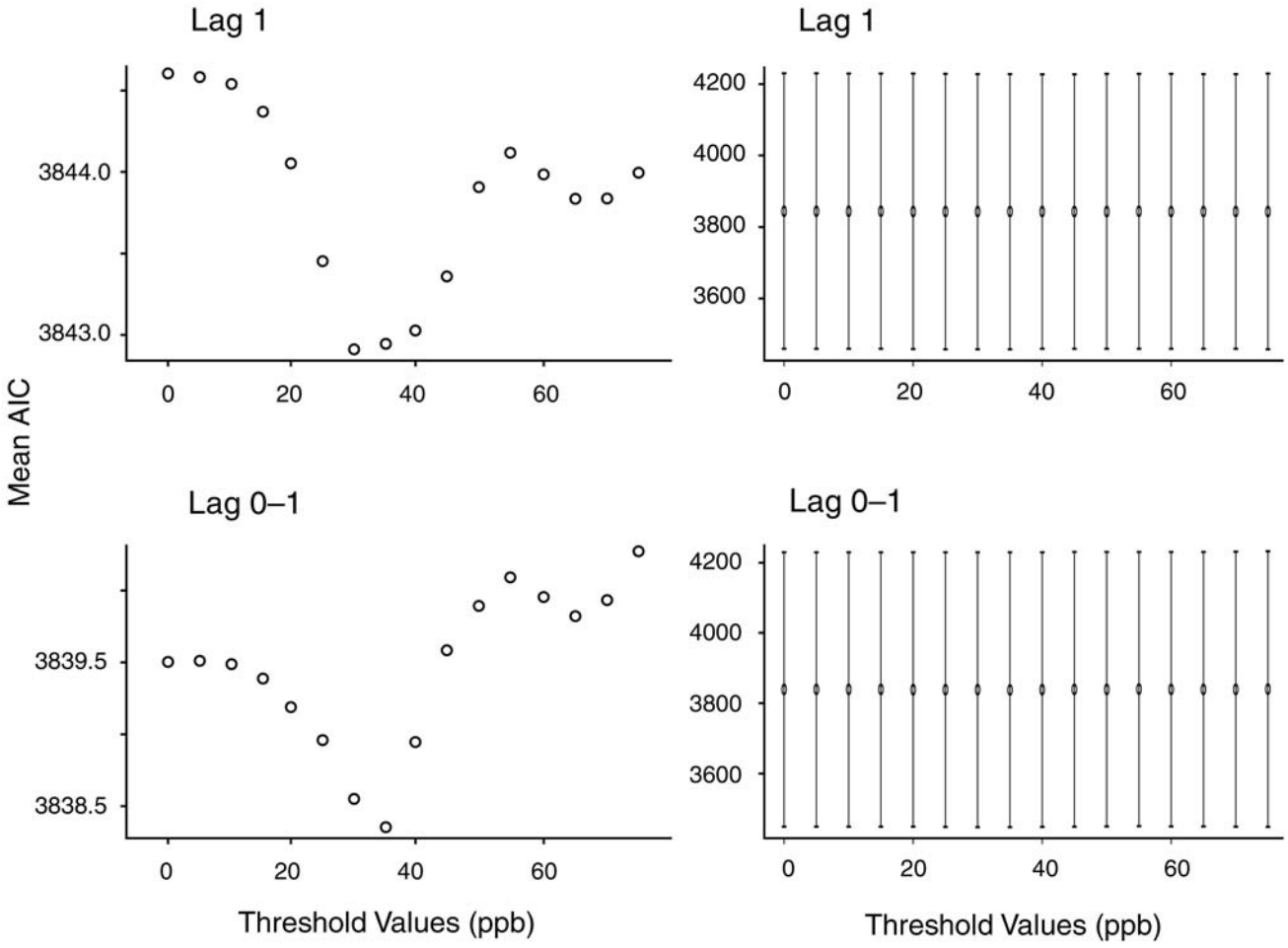


Figure 21. Canada (all ages): Plots of threshold values for O_3 versus AIC values for NS, lag 1 and lag 0–1 models and all-cause mortality in 12 Canadian cities, with and without error bars.

AIC versus threshold varies from city to city. Some cities seem to indicate a threshold, some indicate none, and some even have the lowest AIC at $h = 75$. However, the change in mean AIC is very small compared with the change in the AIC within cities. In fact, in an ANOVA analysis testing for a difference in mean across thresholds, the *P*-value is larger than 0.98 for each lag.

European Results The investigators applied the threshold analysis to investigate the forms of the concentration–response relationships of PM_{10} and O_3 (lag 1) with all-cause mortality at all ages for 22 European cities. In investigating the PM_{10}–all-cause mortality association, they excluded Stockholm from the analysis because the maximum PM_{10} concentrations in this city were at 49 μg/m^3. They ran a separate analysis that included Stockholm up to the relevant threshold values. Figure 22 presents the results of the corresponding analysis for the pooled results. All of the observed deviance values were very close. In an analysis looking for a difference in mean across thresholds, the *P*-value was larger than 0.98, giving no evidence of a possible threshold. Figure 23 presents the results from the threshold analysis investigating the O_3 all-cause mortality association for the full-year data. As with PM, the comparison of the mean deviance values did not support the hypothesis of a threshold. A similar pattern was observed for the analysis of O_3 during the summer.

United States Results The concentration–response relationship between exposure to PM_{10} and mortality was also examined in the NMMAPS data. First, a simulation study was conducted to explore the behavior of the statistical methodology for detecting a nonlinear concentration–response curve and to determine whether sufficient information was available from the data to estimate a threshold. Second, the NMMAPS database was analyzed to determine if there was any evidence in the data for a nonlinear concentration–response relationship.

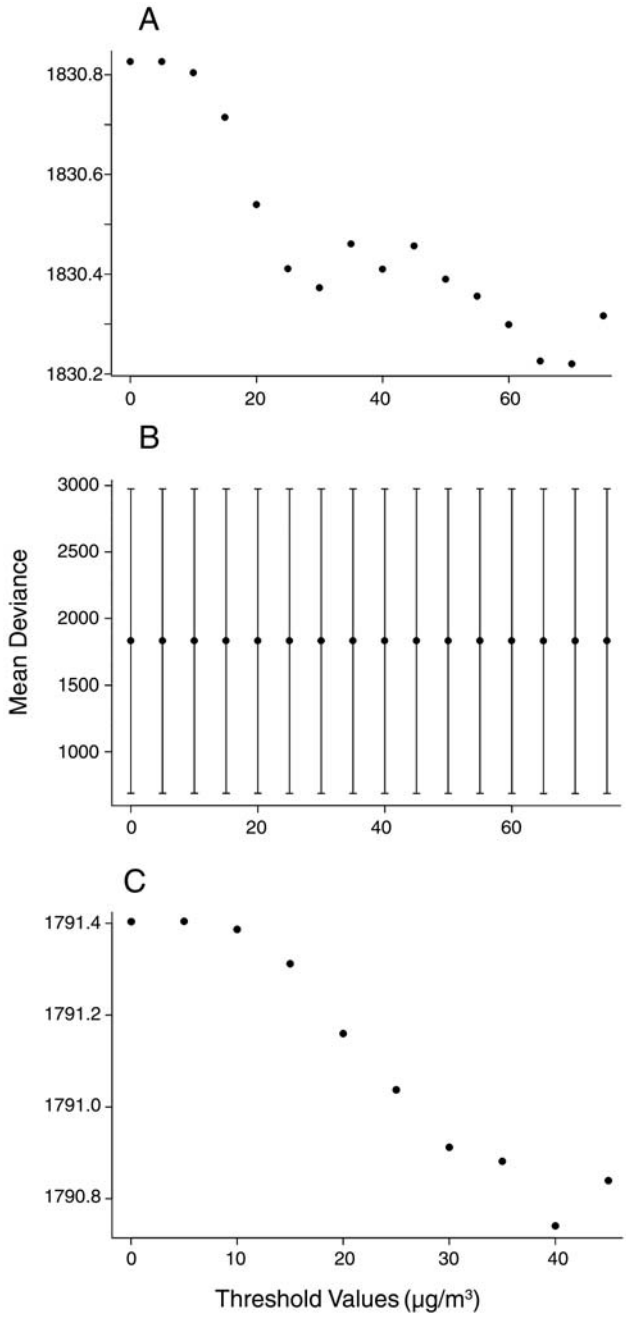


Figure 22. Europe (all ages): Plots of threshold values versus mean deviance of the fitted models for PM_{10} (lag 1) and all-cause mortality in the European cities A: excluding Stockholm; **B**: excluding Stockholm, with error bars **C**: including Stockholm.

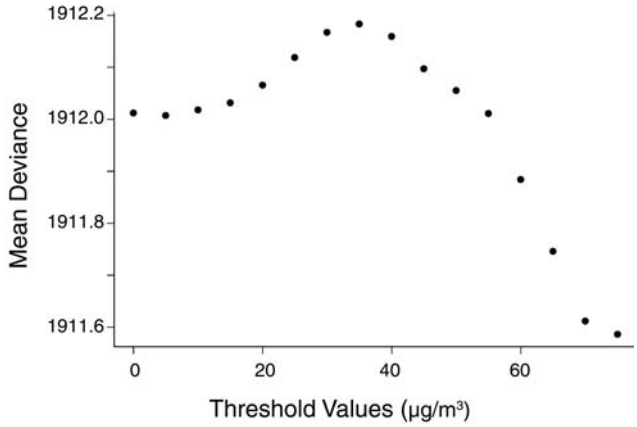


Figure 23. Europe (all ages): Plots of threshold values versus mean deviance of the fitted models for O_3 (lag 1) and all-cause mortality in the European cities.

The particular nonlinear concentration–response model is a threshold model, or broken-line model, where the effect of exposure is assumed to be zero for values of the exposure below some specified value h. For values of the exposure greater than h, the effect is assumed to be linear in the exposure. Such a model takes the outcome Y_t, which can be a daily count of mortality or hospitalization, and relates it to the exposure P_t as $Y_t = \beta_0 + \beta_1 (P_t - h)_+ +$ confounders, where $P_t - h)_+$ is zero when $P_t < h$.

For the simulation study, the investigators simulated mortality and PM_{10} data assuming four different relationships for the association between PM_{10} and all-cause mortality. They chose four values of β_1, the log-relative risk for mortality associated with PM_{10}, — 0.01, 0.005, 0.001, and 0.0005 — and simulated 250 data sets for each. In each simulated data set, the true threshold value was assumed to be zero (i.e., no threshold) and the length of the daily time-series was 10 years.

For each data set, the investigators calculated the AIC for a GLM fit to the data using a range of thresholds. The thresholds chosen were 0, 5, 10, 15, … , 75 μg/m^3. After all of the models were fitted and AIC values computed, they selected the model that minimized the AIC. This process was repeated for each of the simulated data sets.

Figure 24 shows the distribution of the thresholds estimated by the minimum AIC procedure. For $\beta_1 = 0.01$, the mass of the distribution was concentrated in the range 0–15 μg/m^3. The procedure seemed to work well, considering that the true value of h was assumed to be 0. As the true value of β_1 decreased, the distribution of the minimum AIC estimate became flatter and less informative. At a value of 0.0005, the distribution was almost completely flat, and certainly not concentrated about zero.

The results of the simulation studies indicate that detecting a threshold in a broken-line type of model would be very difficult in a scenario if the true association between PM_{10} and mortality was relatively small, as in many time-series studies. For a single-city data set, there would be insufficient evidence for discriminating among possible thresholds.

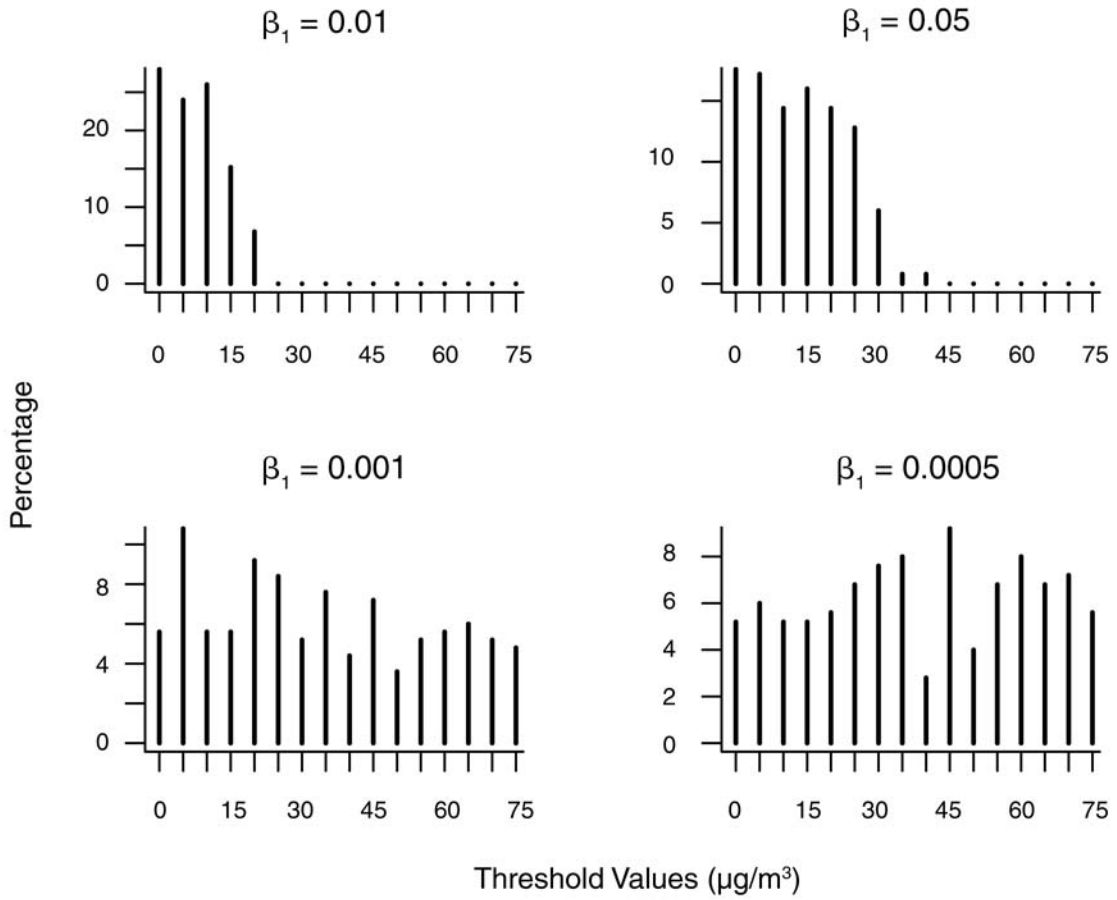


Figure 24. United States: Distribution of simulated thresholds (n = 250) for four hypothetical relationships (β_1) between PM_{10} and all-cause mortality.

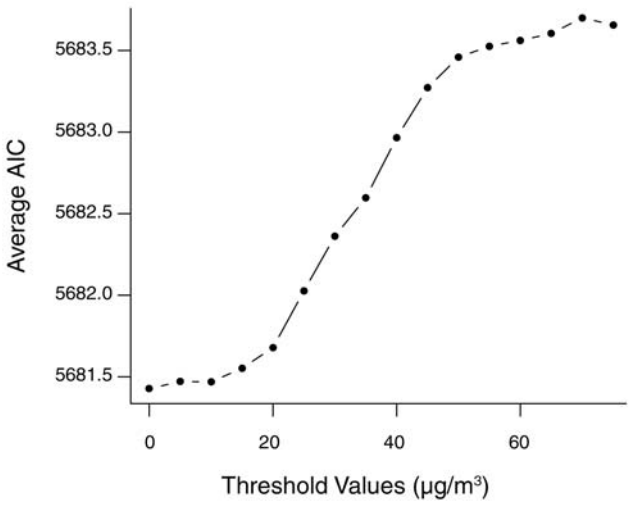


Figure 25. United States: Plot of average AIC values versus threshold values from analyses of 15 NMMAPS cities using lag 1 PM_{10} and all-cause mortality.

Evidence can be combined from multiple locations in multicity studies, a feature that may assist in an investigation of concentration–response relationships. We analyzed the NMMAPS database using the methodology of Daniels and colleagues (2000) to determine if there was evidence of a threshold in the 15 NMMAPS cities with daily PM_{10} data. For each city we analyzed the relationship between lag 1 PM_{10} and all-cause mortality in all ages using a broken-line model similar to the one used in the simulation study. In this model we also controlled for the relevant confounders specified in the first-stage protocol. This model was fit for a number of different threshold values and the AIC was computed each time. Then for a given threshold, we averaged the AIC values across cities to obtain an average AIC. We then selected the threshold to be the minimizer of this average AIC. Further details can be found in a paper by Daniels and colleagues (2000).

Figure 25 shows the average AIC values for the NMMAPS analysis of the 15 cities. In this collection of cities, the average AIC was minimized with a threshold of $h = 0$. It should also be noted that the difference between the minimum AIC value and the maximum value in this plot was approximately 2, a difference of less than 0.04%. Hence, the distribution of AIC values was actually quite flat and there was very little evidence of a threshold.

Exploration of the Effect of Systematically Missing Values on the Pollution Effect Estimates

The APHENA project included data sets with full time-series air pollution data (e.g., the European database), but also time-series with systematically missing air pollution data (e.g., Canadian PM_{10} database that had one measurement for every six days). To investigate the sensitivity of air pollution estimates to the availability of full time-series data, the investigators made an exploratory analysis.

They chose four European cities — Athens, London, Milan, and Zurich — that provided full time-series air pollution data for PM_{10} and O_3, then excluded days to obtain time-series with systematically missing data. Specifically, they started from the first day of the time series and then had measurements for days 7, 13, 19, etc., thus resulting in a new series with one measurement every six days. The selection of the four cities was based on the availability of the largest APHEA full time series for air pollution (more than five years of data).

For each of the newly created data series, the investigators applied the APHENA protocol for cities with systematically missing data; in other words, they analyzed only the previous day's pollution (lag 1) and controlled for seasonality using both smoothers (PS and NS) and 3, 8, and 12 *df* per calendar year of data. They compared the estimated effect on all-cause mortality for all ages obtained from the full versus the systematically reduced time-series data.

Figures 26 and 27 present the city-specific and pooled (under random effects) estimated percentage increase in all-cause mortality for a 10-µg/m^3 increase in PM_{10} and O_3 using the PS method, under models with 3, 8, and 12 *df* for seasonality control. For PM_{10} with all three models, the effects estimated for the series with missing data were substantially smaller than those estimated using the full series for two cities, while they were similar for the other two cities. The pooled effect was consistently smaller for all three models.

For O_3, the effects estimated when using the series with missing data were consistently smaller for all cities and all models (Figure 27). The corresponding results for PM_{10} and O_3 using NS (Figures 28 and 29) were very similar to those using PS. In summary, there was random variation in the city-specific estimates from full and missing time-series data. As expected, the standard errors for analyses using systematically missing data were higher than those for analyses using complete data, because less information was available. Although the pooled random estimate for the effect of PM_{10} on all-cause mortality was statistically significant under any choice of smoother and control of seasonality when using the full time series, it was decreased by more than 50% when the investigators used the systematically-missing time series. The comparison for the O_3 effect on all-cause mortality between estimates from full and missing time series gave similar findings.

To further investigate the sensitivity of the results, the investigators tried removing days in a different order in the

case of the analysis of PM_{10} and all-cause mortality in Athens. Rather than beginning with the first day of the series, they started with the second, third, or fourth day in the series, subsequently including data from every sixth day. Figure 30 presents the results of this sensitivity analysis for PM_{10} in Athens using 3, 8, and 12 *df* per calendar year of data for seasonality control and NS as smoothers. Although the estimates from the time series with missing data were generally lower compared with the estimates from the full time series, they also appeared to exhibit some random variation relative to the estimates from the full series.

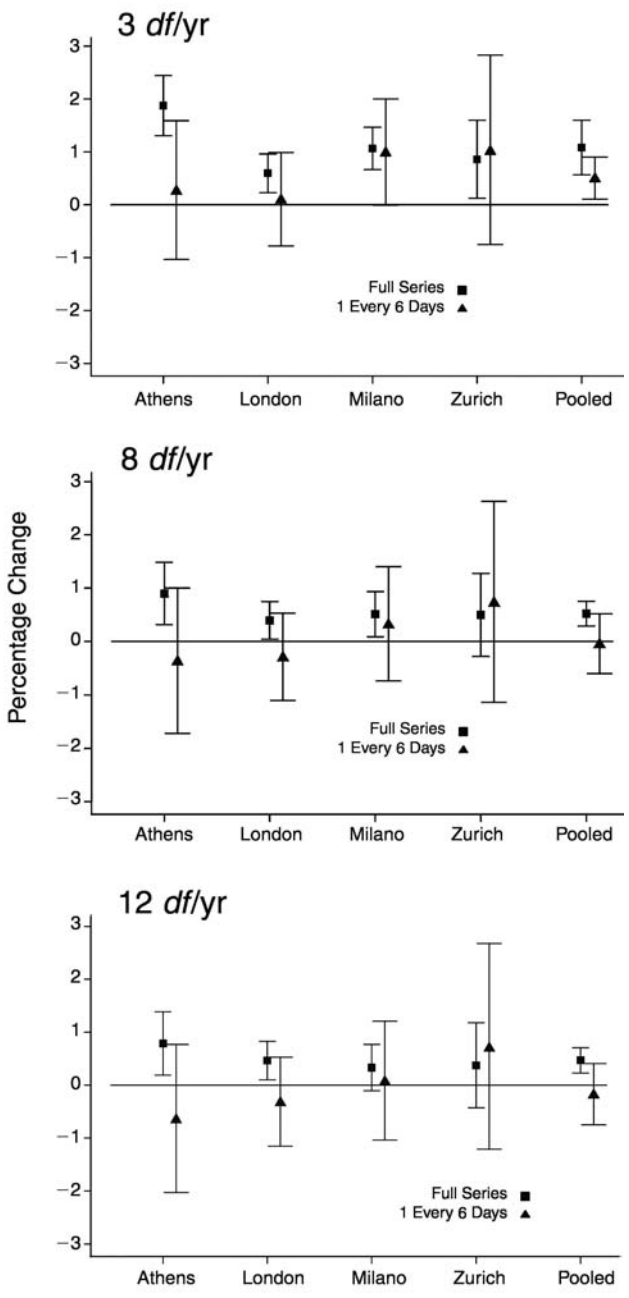


Figure 26. Four European Cities (all ages): Comparison of percentage change in all-cause mortality with a 10-μg/m^3 PM_{10} increase (lag 1) and penalized splines for time series with full or systematically missing data.

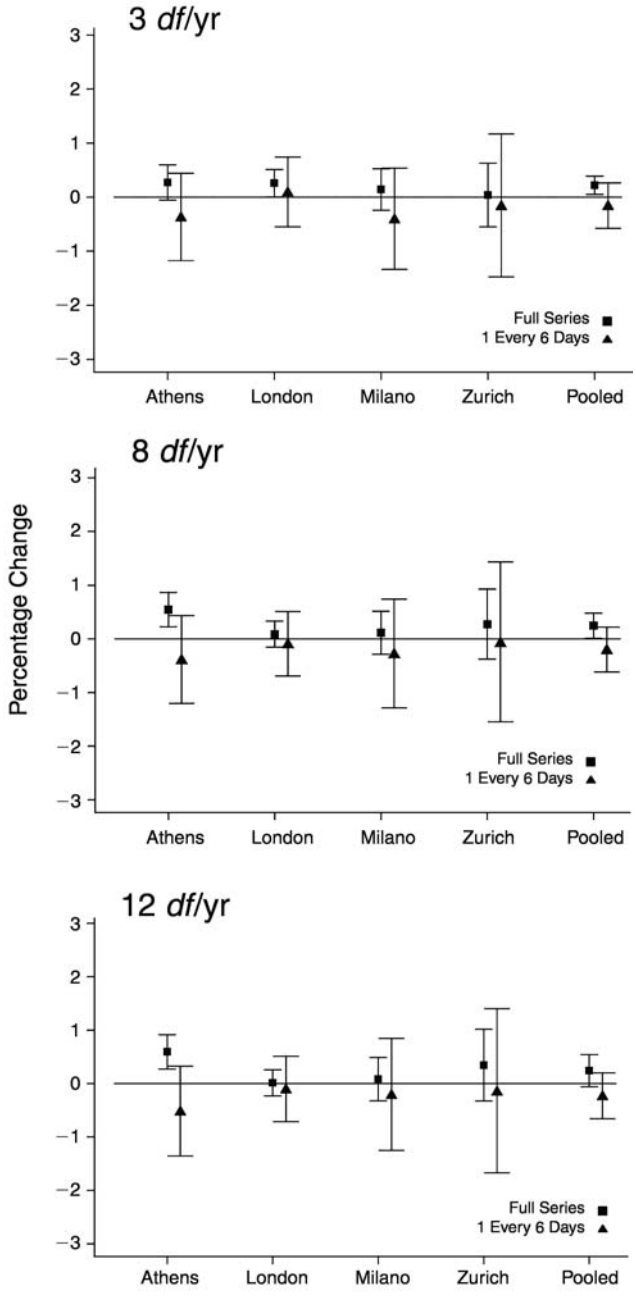


Figure 27. Four European Cities (all ages): Comparison of percentage change in all-cause mortality with a 10-μg/m^3 O_3 increase (lag 1) and penalized splines for time series with full or systematically missing data.

QUALITY CONTROL OF THE CITY-SPECIFIC ANALYSIS

Within the APHENA project the investigators carried out a quality control analysis to investigate whether the implementation of the analysis by different statisticians introduced any bias to the results. For this purpose each center (i.e., Europe, Canada, U.S.) contributed two city-specific data sets with full time-series pollution data to be analyzed by the other centers. As a result, six cities — each part of the APHENA project — were analyzed by all three statisticians involved in the mortality analysis. These cities were Athens and London from the European database, Halifax and Toronto from the Canadian database, and Chicago and Detroit from the U.S. database.

The investigators analyzed only all-cause mortality for the all-age category. They applied the APHENA methodology

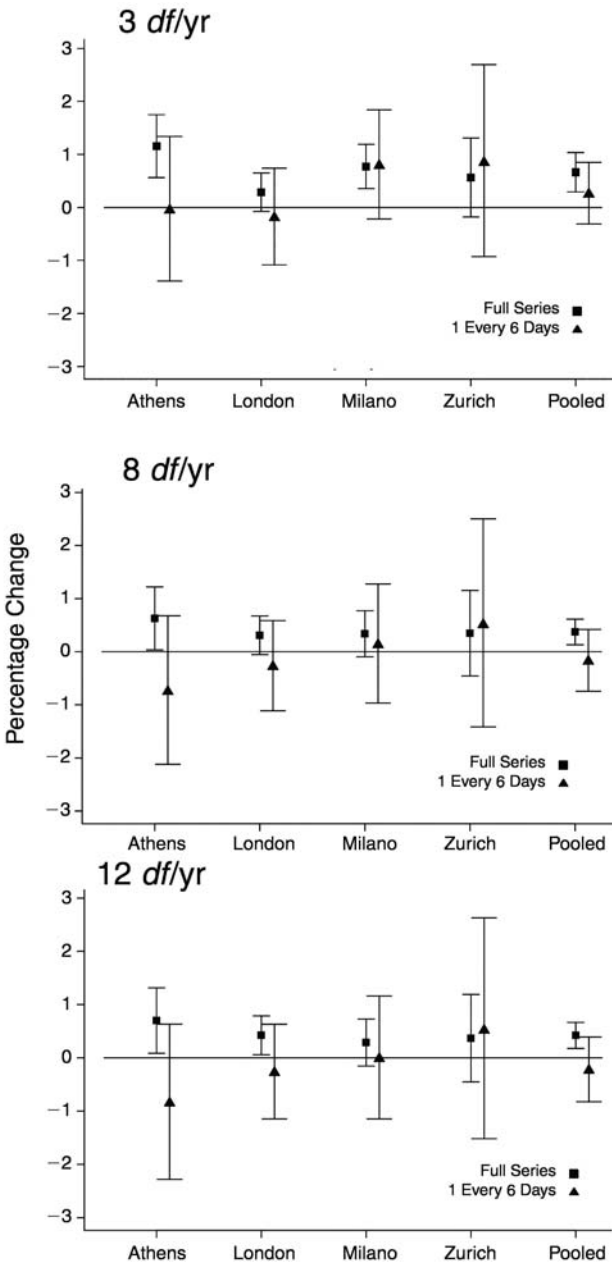


Figure 28. Four European Cities (all ages): Comparison of percentage change in all-cause mortality with a 10-μg/m^3 PM_{10} increase (lag 1) and natural splines for time series with full or systematically missing data.

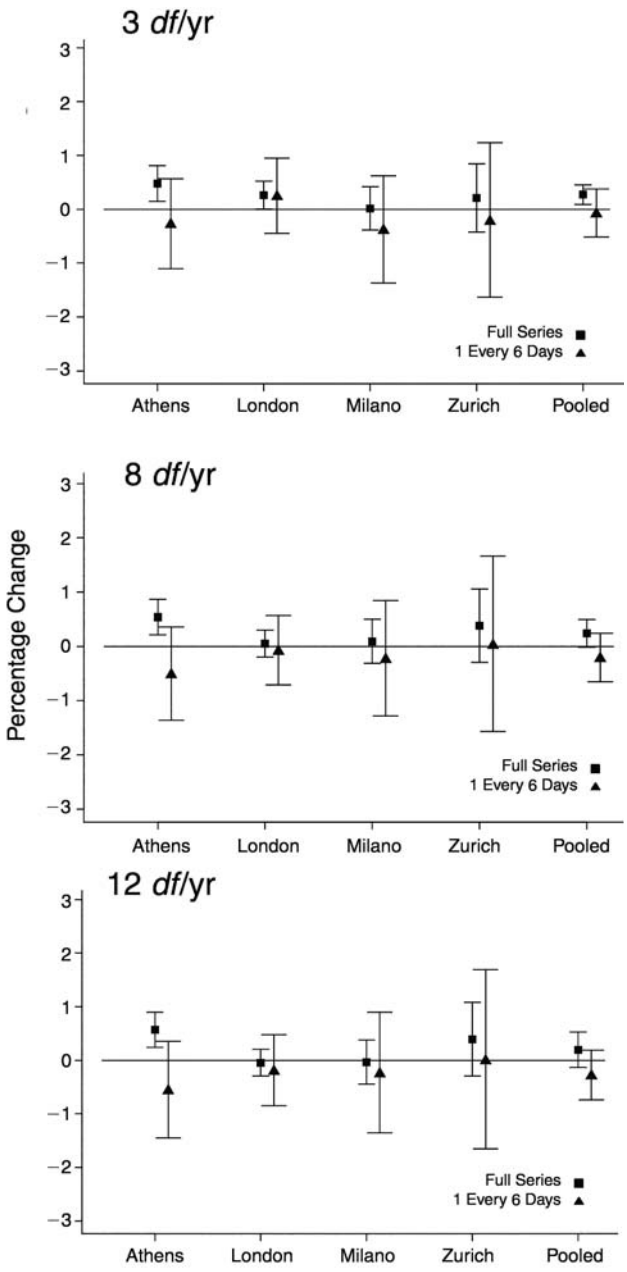


Figure 29. Four European Cities (all ages): Comparison of percentage change in all-cause mortality with a 10-μg/m^3 O_3 increase (lag 1) and natural splines for time series with full or systematically missing data.

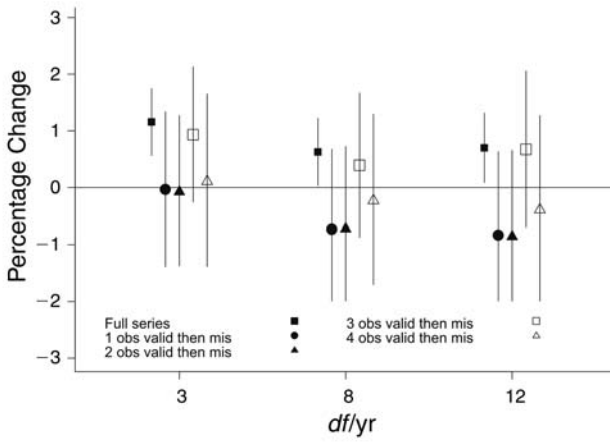


Figure 30. Athens (all ages): Sensitivity of percentage change in all-cause mortality with a 10-μg/m^3 PM_{10} increase (lag 1) using NS to systematically missing data.

protocol for single-pollutant models for both pollutants (PM_{10} and O_3). They analyzed both the previous day's pollution and, when possible, the average of the previous and same day (lags 1 and 0–1, respectively). Analyses were done using 3, 8, and 12 *df* per year, and for cities with full time series data, degrees of freedom were determined using the PACF criterion. NS and PS were evaluated for both pollutants.

Given differences in the data available from the six cities included in this quality control exercise, not all cities were included in each analysis. Canadian cities had systematically missing data for PM_{10}, and thus were included only in the analysis of lag 1 using 3, 8, and 12 *df* per year for the PM_{10} analyses. Detroit had O_3 data only for the summer period, so it was excluded from the O_3 analyses.

Figures 31–34 and 35–38 present the various estimates for O_3 and PM_{10}, respectively, obtained by the European,

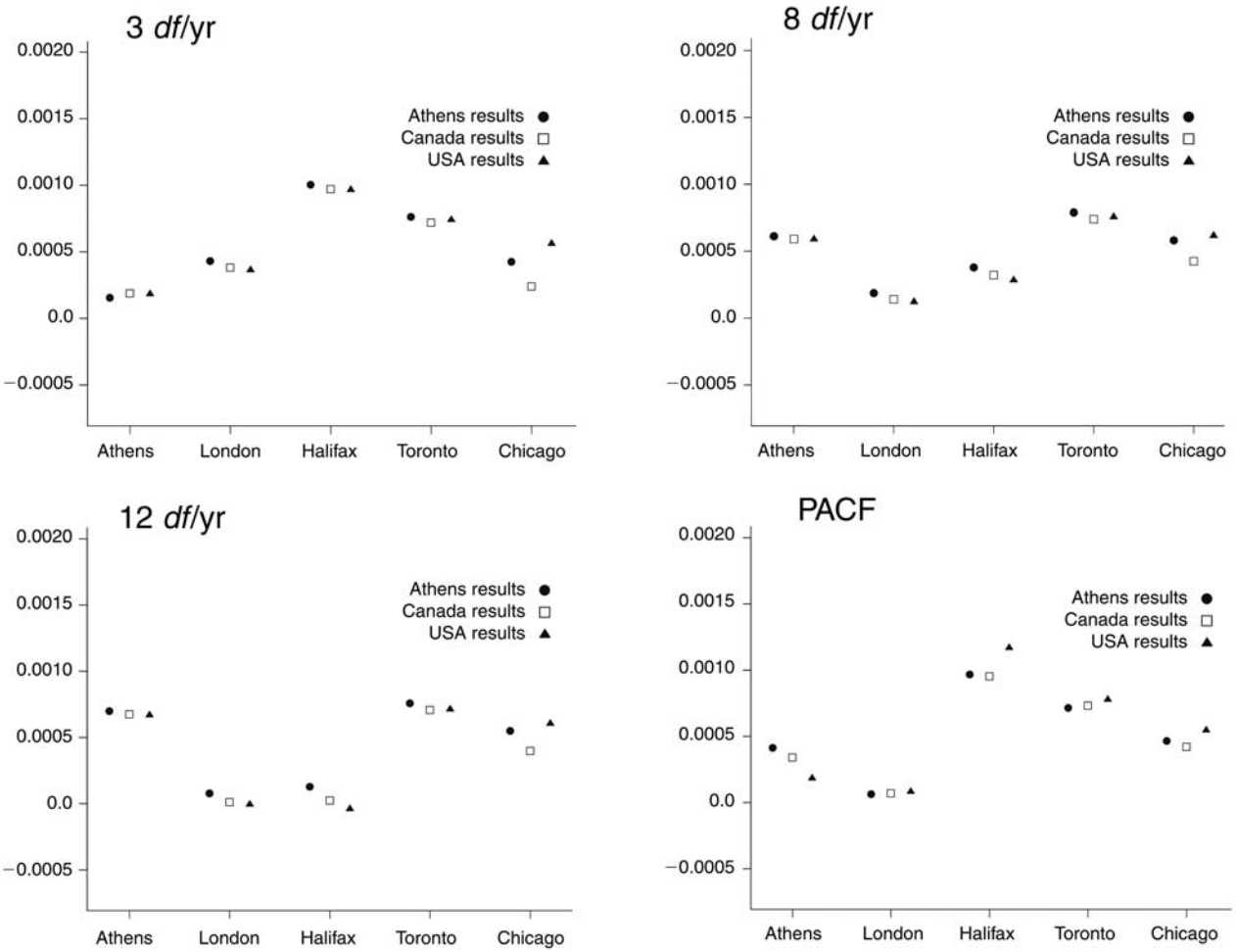


Figure 31. Coefficient for estimated O_3 effect (per 10-μg/m^3 at lag 0–1) with penalized splines for all-cause mortality (all ages) in five cities as analyzed by three statisticians (Athens, Canada, U.S.).

Canadian, and U.S. analysts for each city. Variables are: O_3 or PM_{10}; 3, 8, or 12 *df* or PACF; PS or NS; lag 0 or lag 0–1.

In summary, the city- and method-specific estimates provided from all three centers matched closely. Some minor differences could be easily explained by standard numerical variability and differences in computer hardware. In particular, the great majority of results obtained using the NS smoother were identical, while the minor differences were observed mainly when the PS smoother was applied. Where small differences in the estimates obtained by the PACF method were observed, they were due to the introduction of autoregressive terms in the final model. Standard errors for all estimates were generally identical.

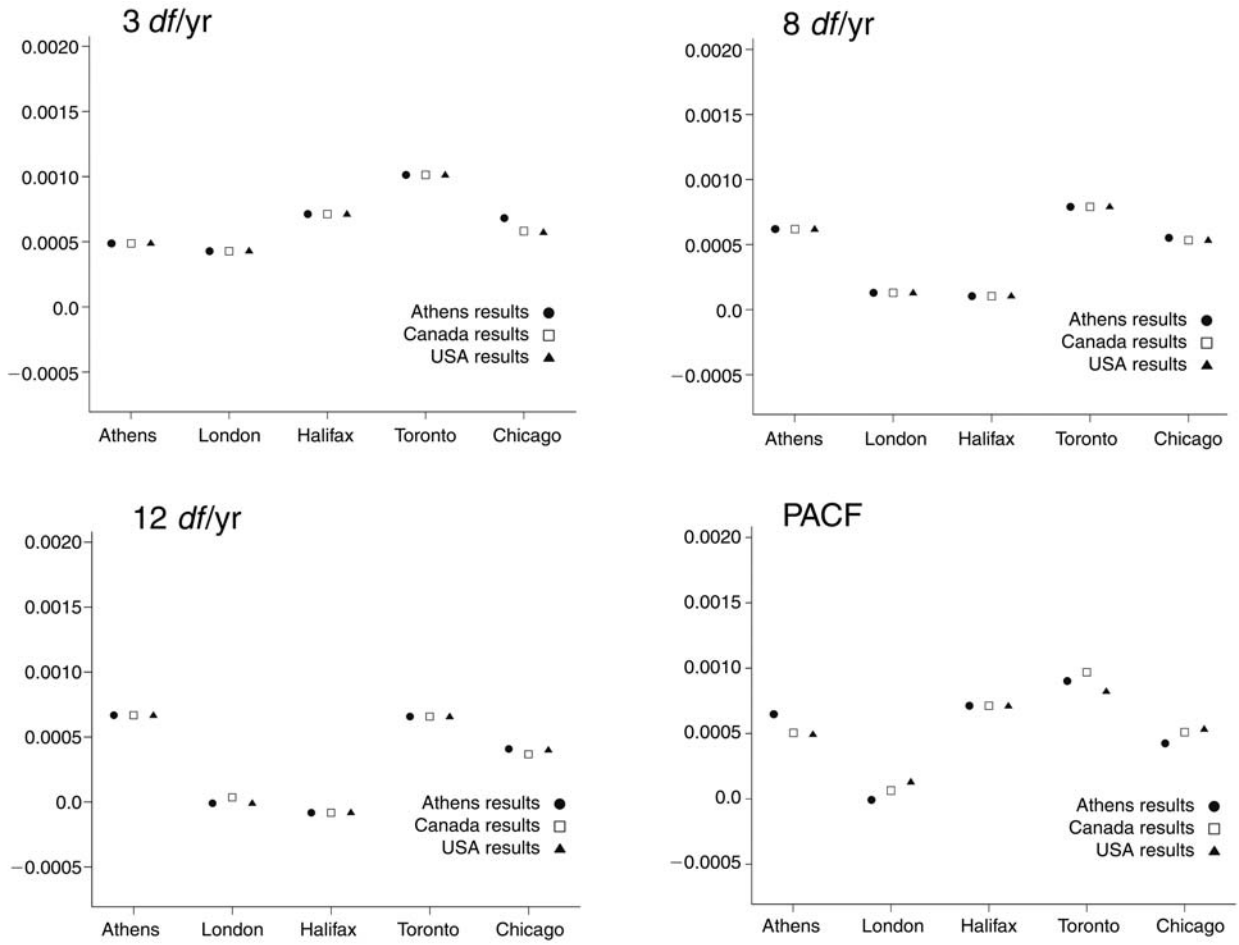


Figure 32. Coefficient for estimated O_3 effect (per 10-µg/m^3 at lag 0–1) with naturalized splines for all-cause mortality (all ages) in five cities as analyzed by three statisticians (Athens, Canada, U.S.).

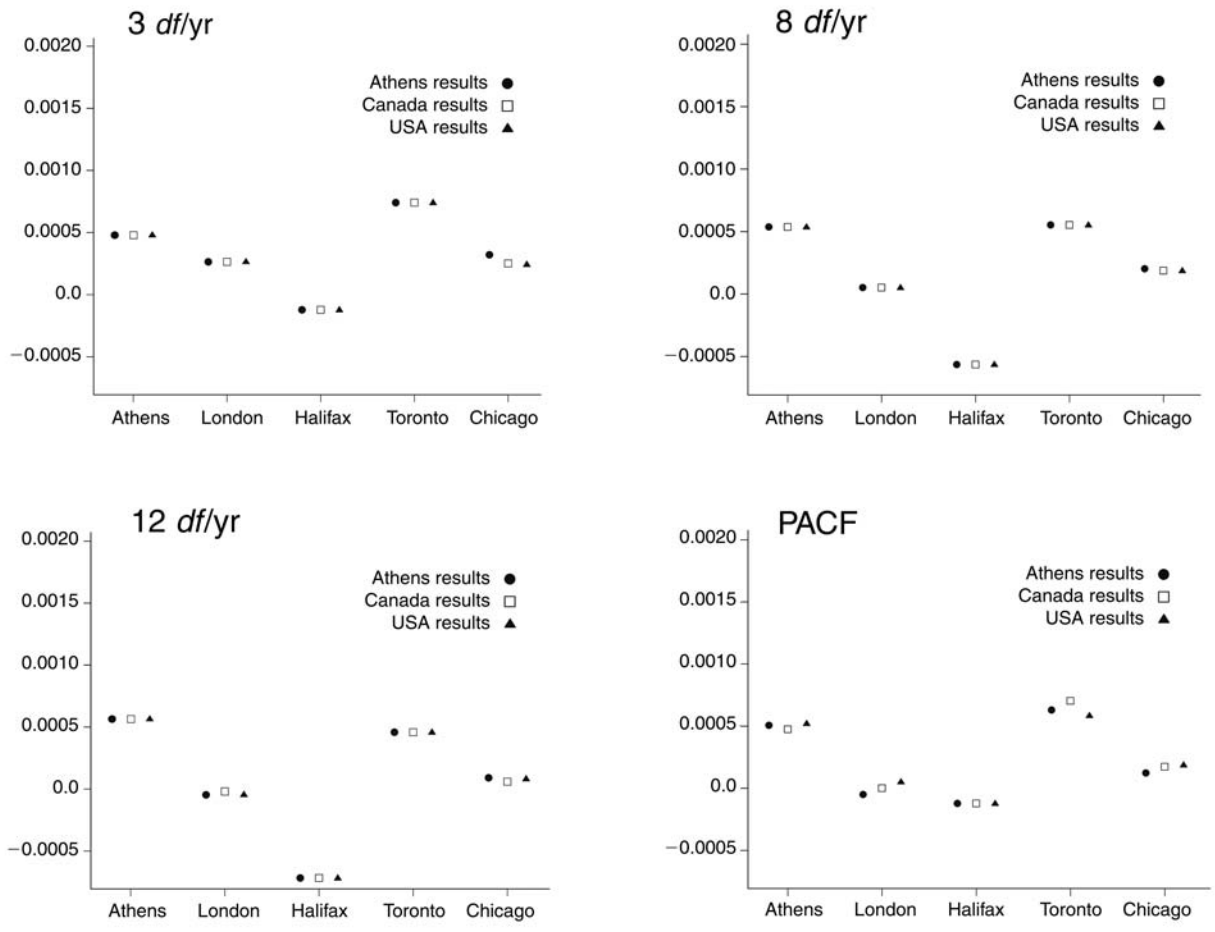


Figure 33. Effects per 10-µg/m^3 O_3 increase (lag 1) and penalized splines on all-cause mortality (all ages) in five cities as analyzed by three statisticians (Athens, Canada, U.S.).

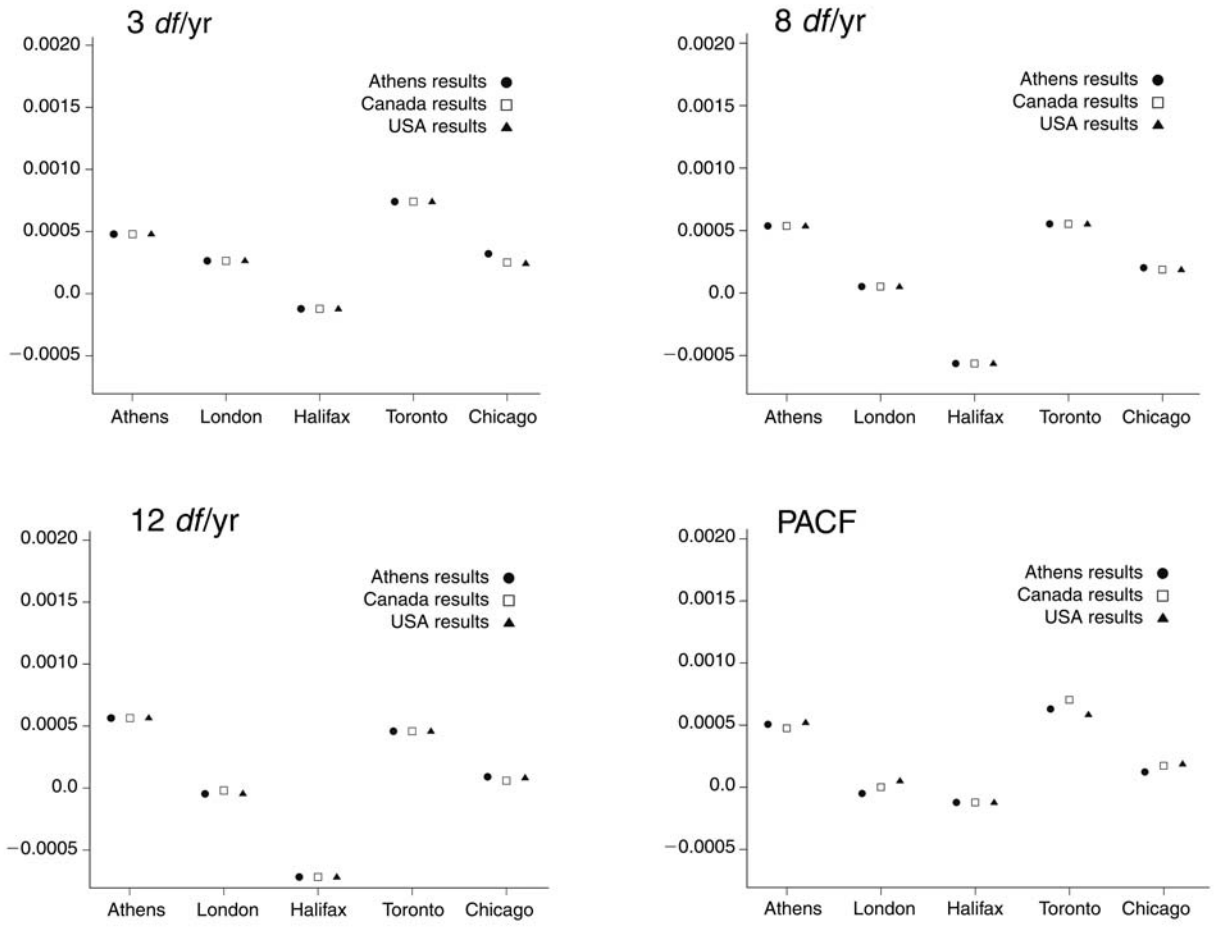


Figure 34. Effects per 10-μg/m^3 O_3 increase (lag 1) and natural splines on all-cause mortality (all ages) in five cities as analyzed by three statisticians (Athens, Canada, U.S.).

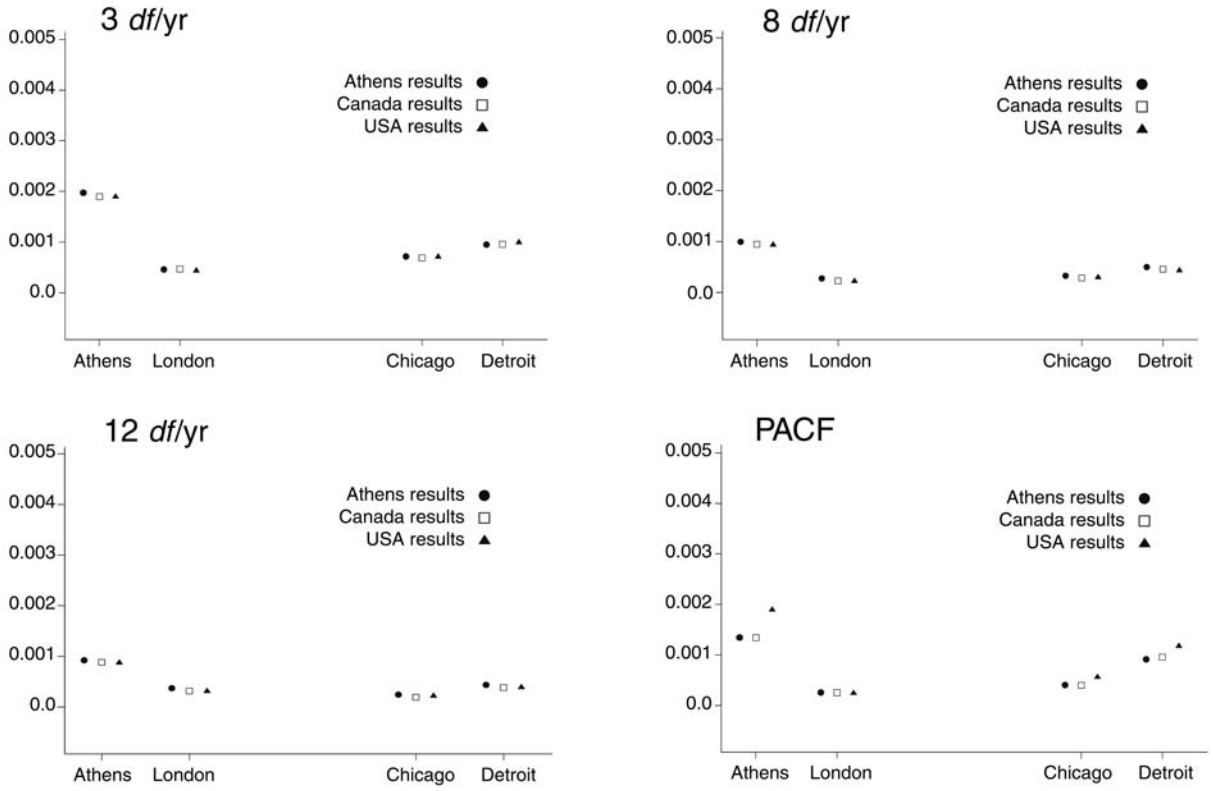


Figure 35. Effects per 10-μg/m^3 PM_{10} increase (lag 0–1) and penalized splines on all-cause mortality (all ages) in four cities as analyzed by three statisticians (Athens, Canada, U.S.).

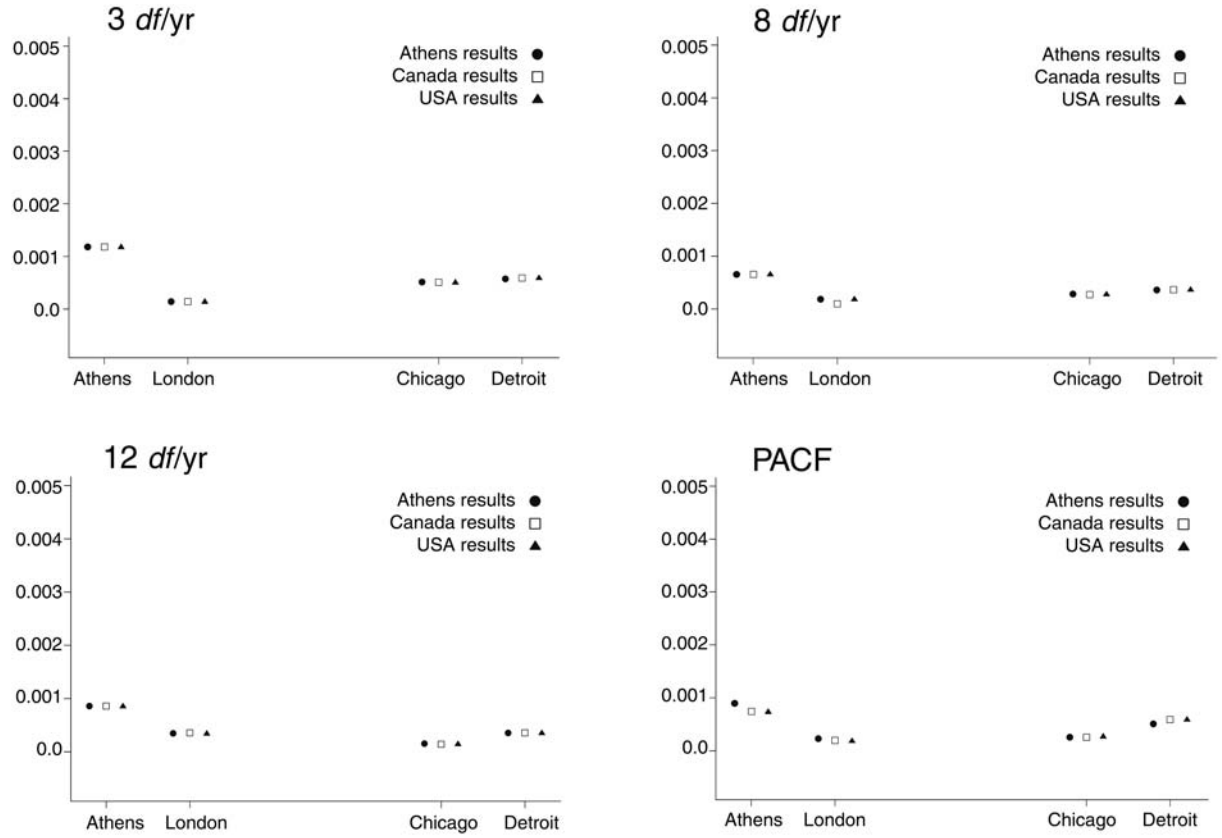


Figure 36. Effects per 10-μg/m^3 PM_{10} increase (lag 0–1) and natural splines on all-cause mortality (all ages) in four cities as analyzed by three statisticians (Athens, Canada, U.S.).

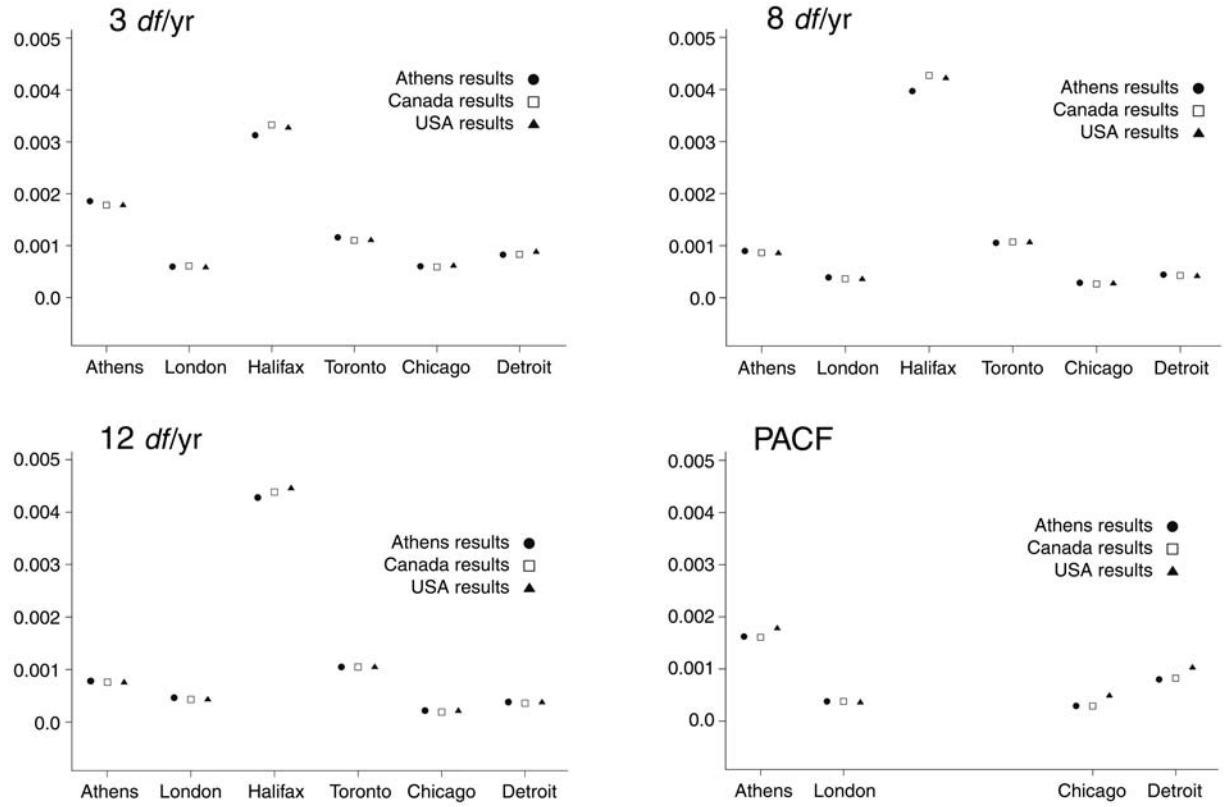


Figure 37. Effects per 10-μg/m^3 PM_{10} increase (lag 1) and penalized splines on all-cause mortality (all ages) in six cities (for 3, 8, and 12 *df*) or four cities (for PACF) as analyzed by three statisticians (Athens, Canada, U.S.).

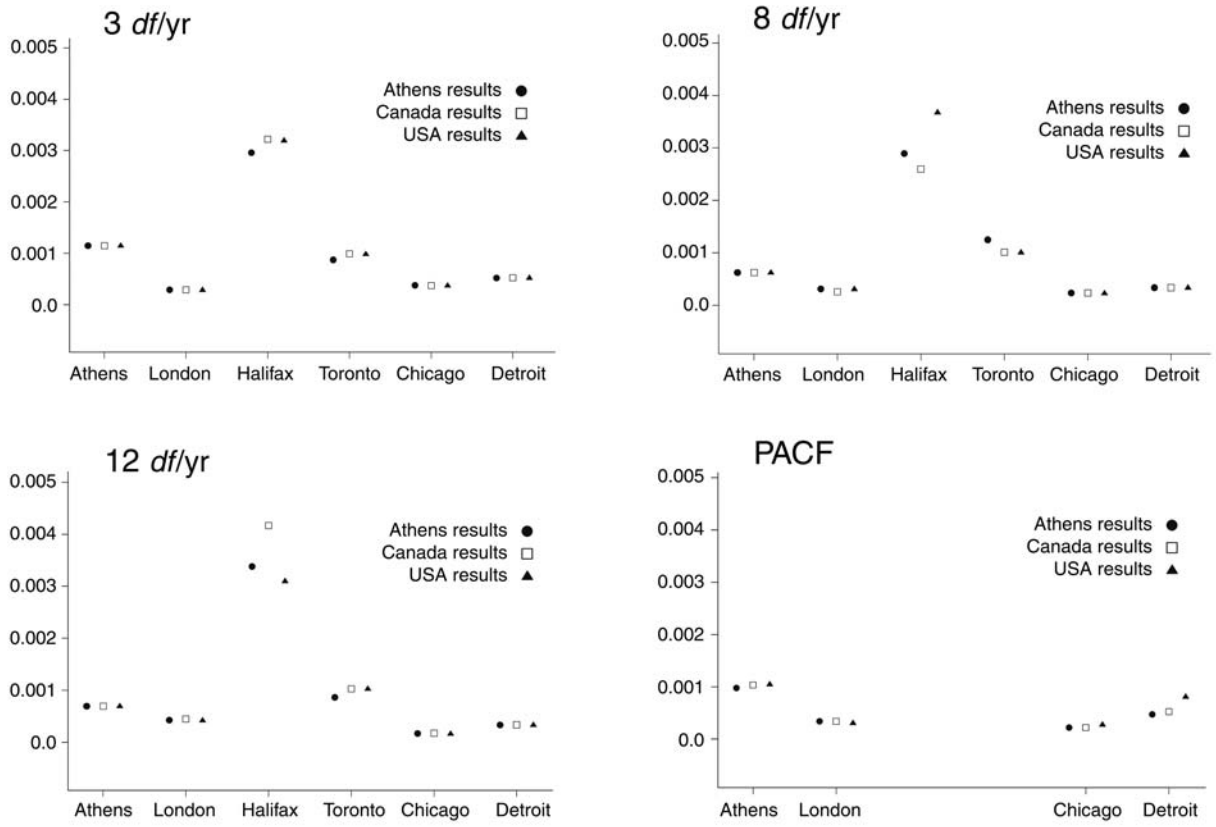


Figure 38. Effects per 10-μg/m^3 PM_{10} increase (lag 1) and natural splines on all-cause mortality (all ages) in six cities (for 3, 8, and 12 *df*) or four cities (for PACF) as analyzed by three statisticians (Athens, Canada, U.S.).

DISCUSSION

BACKGROUND

The first of the contemporary wave of time-series studies of air pollution and mortality and morbidity was published in the late 1980s and early 1990s (e.g., Hatzakis et al. 1986; Schwartz and Dockery 1992); and others as reviewed by Bell and colleagues (2004b). The initial studies were based on data from single cities and used a variety of analytic approaches, particularly in studies in Europe, to address potential confounding by time-varying covariates, such as season and temperature. Initial criticisms of these studies focused on whether findings were model-dependent. The difficulty of comparing results among studies was also noted. By the mid-1990s, the need for well-documented analytic approaches and sensitivity analyses was recognized, and multicity studies were initiated to gain precision and to explore variation in estimates of the effects of air pollution across locations.

The APHEA1 project was initiated in 1993 to address these methodological concerns and to provide estimates of the health effects, particularly mortality and hospital admissions outcomes, of air pollution for Europe. Its protocol called for a network of collaborators to obtain data for cities that had the requisite monitoring and data-capture systems in place, and then to analyze the data using a common analytical approach (Katsouyanni et al. 1996). In the initial analyses, time-varying factors were addressed through use of sinusoidal terms, and models were developed on a city-specific basis. In the subsequent APHEA2 project, a new protocol elaborated on the APHEA1 protocol, addressing time-varying confounders by applying GAM Poisson regression models that involved use of LOESS smoothers (Touloumi et al. 2004). Time-series models were developed and fit to each individual city's data according to a common protocol, but to accommodate potential local differences, the models were not required to be identical.

The National Morbidity and Mortality Study (NMMAPS) was initiated several years after the APHEA1 project, as an outgrowth of the Particle Epidemiology Evaluation Project carried out from 1994 through 1996 by the Johns Hopkins investigators with funding by HEI (Samet et al. 1997, 2000b,c). The design of NMMAPS was intended to address the same methodological challenges and policy needs that motivated APHEA1 and APHEA2, but for the United States. The availability of uniformly collected and reported data for the United States facilitated the conduct of NMMAPS. In the NMMAPS analyses, GAMs were used to control potential temporal confounding. The same model was applied in each city, and sensitivity analyses were carried out by varying the degrees of freedom in the GAM. NMMAPS also included a hospitalization component based on Medicare data for 14 U.S. cities.

In Canada, data on mortality and hospitalization were available through Health Canada. Multiple time-series analyses of the effects of air pollution on mortality and morbidity were carried out, with the number of cities varying among the analyses. One key publication on PM, O_3, and mortality involved data from eight Canadian cities for the 11-year period from 1986 through 1996 (Burnett et al. 2000; Burnett and Goldberg 2003). A 1998 report on O_3 and mortality included 11 cities (Burnett et al. 1998). In reanalyses of the mortality data undertaken for HEI, a LOESS smoother for day of study was used with a 90-day span for all cities (Burnett and Goldberg 2003). A number of analyses of air pollution and hospitalization were also reported, some in multiple cities and some in single cities or provinces (Burnett et al. 1997a,b; Chen et al. 2004).

After issues in using S-Plus for fitting GAMs to air pollution time-series data were identified, the investigators involved in the European, Canadian, and U.S. studies reanalyzed the most critical data sets using more stringent convergence criteria with the S-Plus GAM function as well as with alternative models. The estimates from the revised analyses changed little in the APHEA data, were highly sensitive to model choice in the Canadian data, and dropped substantially in the NMMAPS data (Health Effects Institute 2003). The mortality effect estimates for the cities included in the APHENA study from these revised analyses are summarized in Table 44.

In spite of the differing methodological approaches, the published estimates from the studies in Europe and the United States were generally close in value, while the mortality values from the studies in Canada tended to be higher than those from Europe and the United States (Table 44). The APHENA project was motivated by the need to compare findings among these major time-series studies of air pollution with mortality and hospitalization and by the possibility of exploring the basis for any heterogeneity among the risk estimates in the three data sets. Its conduct required initial methodological work to establish a common analytic protocol for both first-stage and second-stage analyses that would replace the differing approaches of the original analyses. In addition, the investigators developed a set of variables across the three databases for consideration as potential effect modifiers in a second-stage analysis that was intended to explore determinants of heterogeneity. In APHENA, the first-stage results generally replicated the previous independent

Table 44. Results of Prior Mortality Analysis for the APHENA Cities[a]

Location	Data Set	Estimated %	Comment
PM_{10}			
Canada	12 cities, 1981–1999	0.47[b]	NS, separate models
Europe	21 cities	0.42[b]	NS, separate models
United States	90 cities, 1987–1994	0.21[b]	NS, common model
O_3			
Canada	12 cities, 1981–1999	0.89[c]	NS, separate models
Europe	21 cities, 1990s	0.31[b]	GAM, LOESS, separate models by city pooled with random effects, 8-hr O_3, summer, lags 0–1
United States	95 cities, 1987–2000	0.52[c]	NS, distributed lag, all year

[a] Canada results are from Table 2 of Burnett (2004), rescaled to 10-μg/m^3 intervals. Europe results are from Gryparis and colleagues (2004). United States results are from Bell and colleagues (2004a,b).

[b] Per 10 μg/m^3.

[c] Per 10 ppb., 2-day moving average of the daily 1-hour maximum O_3.

analyses conducted by the three groups of investigators. For mortality and PM, risk estimates from the APHEA2 and NMMAPS databases were relatively close, while estimates from the Canadian studies were substantially higher. For hospitalization, results were more variable without discernable patterns of variation among the three data sets.

The findings on the impact of O_3 on mortality varied depending on whether results were considered for the full year or only for the summer months. The effects tended to be larger for the summer months and, in the U.S. cities, to be diminished by control for PM_{10}. The estimated effect of O_3 on mortality varied with the number of degrees of freedom and across the three geographic regions. The effect of O_3 on mortality was larger in Canada.

METHODOLOGICAL ISSUES

Overview

A major objective of the APHENA project was to develop a common approach for the first-stage analyses of the time-series data; applied research was carried out to meet this objective. Similarly, the investigators also evaluated the approaches previously used for combining data in the second stage, as well as for assessing concentration–response relationships. Additional methodological research compared findings of time-series analyses with the case-crossover approach. The consequences of missing data, a problem for the U.S. and Canadian cities, were also addressed.

A critical and difficult component of the APHENA project was the development of a common statistical framework for the analyses of the multisite time-series data. To reconcile previous statistical approaches that had been used previously by the three different teams of investigators, the APHENA investigators compared existing statistical approaches and their implementation, and then developed new statistical methods based on these comparisons.

When APHENA was initiated, there was already widespread recognition of methodological issues with analyses of daily time-series data. In fact, since the early 1990s, these same methodological issues have been raised as one explanation for positive findings. The recognition of widespread use of insufficiently stringent convergence criteria in the S-Plus GAM function sparked substantial methodological work on time-series analyses of air pollution data (Dominici et al. 2002b). Additional methodological issues were identified at the time, related to the approximate estimation of standard errors in the S-Plus GAM function, concurvity, and model sensitivity (Ramsay et al. 2003). Many of these issues were explored in the reanalyses that were carried out after the identification of the GAM issues, and were covered in the HEI Report on the revised analyses (Health Effects Institute 2003).

Consequently, the APHENA investigators were aware of the need to justify the methods selected and to explore sensitivity of findings to modeling decisions. One major issue is the approach used to control for unmeasured temporal confounding, caused by factors with a temporal pattern variation similar to that of air pollution concentration that affects the same outcomes measured for the study. In this study, those outcomes are all-cause mortality or hospital admissions. Temporal confounding can occur if

models incompletely control for the effects of temporal factors or if potential confounders are not considered (Zeger et al. 2006).

First-Stage Protocol

In developing the first-stage protocol, the APHENA investigators faced a series of decisions about how to establish a uniform analytic approach: (1) the class of models to be used for analysis; (2) the approach for selecting the appropriate number of degrees of freedom for smoothing temporal confounders; and (3) the suite of variables to be included in the model. To guide the development of the protocol, simulation studies were carried out by the APHEA and NMMAPS teams. Consideration was given to GLM with NS, and also to PS. For the degree of smoothing, the simulation study compared the PACF, AIC, fixed degrees of freedom, and GCV (Appendix A).

In recent daily time-series studies, control of confounding has been carried out by including functions of time in the log-linear regression model (McCullagh and Nelder 1989; Hastie and Tibshirani 1990). These include SS (McCullagh and Nelder 1989; Hastie and Tibshirani 1990), NCS (Green and Silverman 1994), or PS (Wood 2000). As an alternative, temporal confounding can be accounted for by matching in a case-crossover study (Maclure 1991). When a smooth function of time is included in a time-series model, critical decisions include: (1) specification of the degrees of freedom; (2) whether the degrees of freedom should be common across all locations; and (3) whether the degrees of freedom should be fixed a priori or should be estimated from the data. The APHENA investigators conducted theoretical investigations and simulation studies to provide a solid basis for making these decisions.

Dominici and colleagues (2004) showed, theoretically, that for parametric smoothers the degrees of freedom should be at least as large as needed to best predict the pollution time series (Speckman 1988). Peng and associates (2006a) extended this approach to nonparametric smoothers, finding that the approach of Dominici and colleagues performs well in simulated data and produces estimates of the relative rate that are nearly unbiased under all scenarios assumed, even in the presence of severe confounding. Alternatively, selecting the degrees of freedom for the smooth function of time by minimizing the PACF is a data-based approach that is widely used in studies of air pollution and health. Schwartz (1994) suggested that the presence of residual autocorrelation may lead to underestimation of standard errors and result in biased hypothesis tests of the effects of air pollutant variables. Therefore minimizing autocorrelation would appear to be an appropriate goal of the analysis approach. Touloumi and coworkers (2005) evaluated this approach in combination with using parametric and nonparametric smoothers. Simulation studies showed that minimizing the PACF of the residuals worked relatively well when applied to nonparametric or semiparametric smoothers (LOESS and PS), but not as well with NCS.

These studies and the further work by the APHENA investigators showed that no single method alone is adequate or preferred for selecting the underlying model and the degrees of freedom for confounder control. The APHENA investigators learned that most of the alternative methods perform similarly, although there were subtle variations in results that were data-dependent in sensitivity analyses. The simulation studies, not surprisingly, showed some sensitivity of the estimates to each element of model specification. In selecting a core model, emphasis was placed on stability of findings across the three sets of data. Sufficient sensitivity analyses were carried out to confirm that findings were not strongly dependent on model specification. This work led to finalization of the first-stage protocol, and to decisions to have standard sensitivity analyses and to fully report the findings.

The potential temporal confounding factors include day-of-week, season, temperature, and epidemics of respiratory infection. Day-of-week can be readily included in models, and daily temperature data are routinely available. Adjusting for epidemics of disease that increase hospitalization and mortality counts requires either information on their occurrence or diagnosis-specific counts, so that extreme events consistent with an epidemic can be identified. Touloumi and associates (2005) conducted a study involving seven European cities that controlled for the potential confounding effect of influenza by including influenza data or other indirect estimates into the regression model, instead of using smooth functions of time. They found that the association between air pollution and mortality was not weakened by adjustment for influenza epidemics.

Temperature is also associated with mortality; in many cities, temperatures above and below some optimum level are associated with increased mortality. The APHENA investigators used temperature itself in the models, rather than indicators of weather patterns. Previous analyses, including one carried out at the request of HEI, showed that using indicators of synoptic weather patterns did not provide greater control of the effect of temperature on mortality (Pope and Kalkstein 1996; Samet et al. 1998).

Model specification for temperature control has also been explored. Welty and Zeger (2005) developed several

models for controlling for temperature in log-linear regression models of air pollution and health. They found that the pooled relative rates of mortality associated with short-term exposure to PM_{10} were robust to more complex adjustment for temperature effects. For risks associated with O_3, controlling for temperature is more complicated; higher temperatures lead to more O_3 production, and O_3 levels are typically much higher in warmer than colder seasons. Statistical modeling of temperature can affect the association of O_3 with mortality because of this strong temporal correlation between O_3 concentration and temperature (Bell et al. 2004a; Gryparis et al. 2004; Medina-Ramon et al. 2006). Consequently, the APHENA investigators decided to estimate the association between O_3 and mortality with stratification by season.

Concentration–Response Relationships The APHENA investigators also considered issues related to characterizing concentration–response relationships: specifically, whether to use a nonlinear concentration–response curve, and whether current statistical methods are appropriate for providing evidence for a potential threshold in the concentration–response curve. Dominici and colleagues (2000a), Schwartz (2000b), and Dominici and associates (2002a) have developed hierarchical models for estimating a pooled concentration–response curve. Dominici and associates (2002a) modeled the concentration–response curves as NCS with a fixed number of knots, but estimated the knots' locations from the data. With application of this approach to the NMMAPS data, they did not find evidence for a nonlinear concentration–response curve. Samoli and coworkers (2003, 2005) also developed a hierarchical model for estimating a pooled concentration–response curve across 22 European cities included in the APHEA2 project. They modeled the curves using regression splines with two knots and concluded that the concentration–response curve for PM_{10} with mortality can be reasonably approximated with a linear model.

The fits to the data with alternative concentration–response models (e.g., linear vs. nonlinear) are generally compared by using the AIC. In the APHENA project, the investigators compared the fit of several possible threshold models and concluded that the pollutant–mortality association is essentially linear. They also conducted simulation studies to investigate whether sufficient power existed to identify a change point in the concentration–response relationship. They found that given the extent of the available data including the length of the time-series data and the daily average number of deaths, the power to detect a change-point varied greatly with the size of the true pollutant effect; in general, with small effects typical of these studies, limited power existed to detect thresholds.

Distributed Lag Models Consideration was also given to the use of DL models. Such models have the advantage of estimating the DL function, which describes the change over time in the relative risk associated with a given day's air pollution (Almon 1965). The DL coefficients give insight into the total effect of air pollution on a particular day, as that day's exposure contributes to the effect of air pollution on multiple subsequent days. Information about the shape of the DL function provides useful evidence concerning the time course of risk for the outcome, and may give clues about mechanisms by which air pollution causes disease.

Distributed lag models can also give insights into mortality displacement (Zeger et al. 1999; Schwartz 2000a; Schwartz 2001; Dominici et al. 2002b; Zanobetti et al. 2002). In other words, they can be used to investigate whether an increase in the number of deaths during and immediately after exposure is counterbalanced by a decrease in the number of deaths a few days later, when some of the earlier deaths would otherwise have occurred. This pattern is indicative of mortality displacement. Recent approaches to the estimation of the DL function include smoothing the lag-specific coefficients by using polynomials or SS (Zanobetti et al. 2000). Alternatively, unconstrained DL models have also been fitted for estimating the time course of respiratory and heart disease mortality in response to an air pollution episode (Zanobetti et al. 2003). Welty and colleagues (2005) proposed a Bayesian model for the DL function that uses an informative prior that constrains the function by allowing effects corresponding to early lags to take any value, while effects at more distant lags are constrained to be near zero and relatively smooth. Peng and associates (2007) extended this Bayesian approach in a hierarchical fashion and estimated the shape of DL functions between $PM_{2.5}$ and hospital admissions for cardiovascular and respiratory diseases pooled across 100 U.S. counties. While these methodological studies show the utility of DL approaches, the application of DL models in the APHENA project was limited because most locations did not have daily data. Regardless, as part of the APHENA protocol the investigators fit unconstrained DL models to all locations with daily data and estimated cumulative effects. They did not explore mortality displacement.

Second-Stage Analyses One feature of the multisite time-series design is the possibility of combining information across locations (DerSimonian and Laird 1986) and using the combined information to explore effect modification. From a statistical standpoint, the combination step can be accomplished using either Bayesian hierarchical regression models (Gelman et al. 2003) fitted by use of

Monte Carlo Markov Chain (TLNISE) (Everson and Morris 2000), or using random-effect models fitted by maximizing the likelihood function (MLE-Berkey) (Berkey et al. 1998). Dominici and associates (2000b, 2002a) implemented Bayesian hierarchical models for pooling relative rates of mortality across locations and for investigating whether county-specific characteristics would modify the effects. Strengths and limitations of the Bayesian approach have been discussed elsewhere (Dominici 2002). Katsouyanni and associates (2001), Touloumi and colleagues (2004), and La Tetre and coworkers (2005) implemented random-effects models with weighted regression to compare maximum likelihood and empirical Bayes (shrunken) estimates of the location-specific relative rates.

These two statistical approaches were compared in the APHENA project. The hierarchical models formulated by the APHEA and the NMMAPS investigators were almost identical, but the fitting algorithms were different. Simulation studies were conducted for several scenarios involving the degree of heterogeneity of the true effects across locations. As expected, the efficiency of the two computational approaches was similar. However, TLNISE was slightly more efficient than MLE-Berkey (smaller MSE) when the degree of heterogeneity was very small.

In summary, in carrying out the APHENA analyses, the participating statisticians addressed several statistical aspects of the modeling of multisite time-series data and resolved issues by theoretical development and simulation studies. This work has led to several contributions in the statistical literature (Peng et al. 2006; Touloumi et al. 2006; Samoli et al 2008). When simulation studies did not provide a clear indication of a preferable method, the APHENA investigators decided to conduct statistical analyses under all possible approaches to assess sensitivity of the results.

OVERALL RESULTS

Mortality

The APHENA analyses focused on the effects of PM_{10} and O_3; data on other pollutants was not uniformly available across the data sets and consideration had been given to potential confounding by these pollutants in prior reports by the three groups. Overall, PM_{10} was associated with increased all-cause mortality (Figure 9), particularly for people 75 years or older (Figures 10 and 11). Positive associations were generally present for cardiovascular and respiratory mortality, although there were some inconsistencies by health outcome and region (Figures 12–15). Estimates tended to drop with increasing degrees of freedom and were generally lower for the average of lags 0–1 compared with lag 1 concentrations alone. With regard to qualitative findings, there was little indication of model sensitivity, except for respiratory mortality, the category with the smallest numbers of events. The new estimates based on the APHENA protocol were close to those from the prior reports.

As noted in the prior analyses summarized in Table 44, the European and North American estimates were quite similar, but the estimates from the Canadian data were higher than those of the other two centers. Because the data have been analyzed uniformly in APHENA, the higher values observed in Canada cannot be attributed to analytic approaches. Bias leading to higher estimates could come from having more accurate exposure and outcome data in Canada compared with the European countries and the United States. Validation data are not available for either exposure or outcomes to explore this possibility. For example, data on the relationships between ambient concentrations and personal exposures in the various communities would be useful, as would data on the comparative accuracy of death certificate diagnosis across the participating centers. While the effect of PM_{10} on mortality may have been greater in Canada in comparison with the other regions, the investigators could not immediately identify any specific source-mix differences among the APHENA regions that could introduce such effect modification.

Several authors have assembled estimates of the effect of PM_{10} on mortality (both from individual studies and meta-analyses) (Stieb et al. 2002, 2003; Anderson et al. 2004; Pope and Dockery 2006). The values reported from meta-analyses tend to be higher when based on the single-city studies, and not all estimates have been revised subsequent to the identification of the S-Plus issue (Anderson et al. 2005). Summary estimates from single-city studies ranged from about 0.4% to 0.8% per 10-$\mu g/m^3$ PM_{10} increase (Pope and Dockery 2006). Prior European and U.S. estimates were just below this range while the Canadian estimates were at the upper end. For the European and U.S. estimates, the generally lower values in comparison with these meta-analytic summary estimates are likely to reflect aspects of city selection and model specification that may have led to upwardly biased estimates in the meta-analyses, as well as to publication bias, for which there is evidence (Anderson et al. 2005; Bell et al. 2005).

Increased mortality was also observed in association with higher O_3 concentrations, although results were more variable in relation to model specification. They were also more seasonally dependent in Europe, with effect estimates being much greater in the summer. As for PM_{10}, the effect estimates for O_3 concentrations were greater for Canada than for Europe and the United States. When compared by

cause of death, the estimates tended to be greater for cardiovascular mortality. The magnitude of the effect estimates from the APHENA study corresponded to those in preceding reports (Table 44).

The findings confirm that O_3 is associated with increased mortality in North America and Europe. The higher mortality rates in summer have been noted previously (Gryparis et al. 2004) and may reflect the higher concentrations of O_3 during that season, greater personal exposure during the summer, and differing characteristics of the air pollution mixture by season.

Notably, the estimates were relatively robust in the sensitivity analyses for both PM_{10} and O_3. Neither the model used nor the specified degrees of freedom led to substantial variation in the estimates.

Hospitalization

Numerous time-series analyses of air pollution and hospital admissions have been carried out. Analyses of time-series data on hospital admissions must control for temporal confounding from day-of-week and from holidays, which affect both admission patterns and pollution levels. Patterns of potential temporal confounding also vary from place to place, depending on the specific interrelationships of day-of-week and holidays with air pollution and hospital admissions. When analysing the hospitalization data in APHENA, the investigators used a uniform approach to controlling for such temporal confounding, but found that results across the three data sets were variable.

PM_{10} was not associated with risk of respiratory or cardiovascular hospitalizations in Canada. In Europe, PM_{10} was associated with both respiratory and cardiovascular disease hospital admissions, although the quantitative estimates were sensitive to the number of degrees of freedom and to the underlying model (NS vs. PS). Even greater model sensitivity was observed for respiratory admissions in the United States, with positive and significant estimates at 3 *df*, and negative and nonsignificant estimates at 8 and 12 *df*. For cardiovascular disease admissions, the estimates were more stable and generally statistically significant and positive. The sensitivity of estimates to model specification may reflect the multiplicity of temporal factors affecting hospital admissions; also, the respiratory admissions series had much smaller numbers than the cardiovascular disease admissions series and were more unstable as a result.

Prior analyses of APHENA's component databases have shown associations of PM with risk for hospital admissions. For Canada, parallel analyses have not been reported at the region level, although positive associations have been identified in single-city and province-level analyses (Burnett et al. 1995, 1999; Chen et al. 2004), and in an analysis of 11 of 16 cities using soiling index as the index of PM (Burnett et al. 1997a). For APHEA2, Atkinson and colleagues (2001) reported analyses based on 10 European cities for PM_{10} and respiratory admissions. Positive and significant estimates were found for major categories of respiratory diseases. The summary estimate for all-respiratory disease categories for people 65 years or older was 0.9% per 10-$\mu g/m^3$ PM_{10} increase, similar to the APHENA estimates with 3 *df*/year. For cardiovascular admissions, Le Tertre and colleagues (2002) reported findings for cardiovascular disease in eight cities in Europe. The pooled estimate of 0.7% per 10-$\mu g/m^3$ PM_{10} increase was within the range spanned by the various APHENA estimates. The APHEA findings proved robust to reanalysis with alternative methods and more stringent convergence criteria (Atkinson et al. 2003; Le Tertre et al. 2003). For the United States, Zanobetti and Schwartz (2003) presented revised analyses of hospital admissions using NS and PS. For respiratory and cardiovascular disease admissions, the estimates were around 1.0% per 10-$\mu g/m^3$ PM_{10} increase for lag 0–1. At higher degrees of freedom in the APHENA analyses, the estimates tended to be lower for both categories of disease.

For O_3, estimates of the effects on both cardiovascular and respiratory disease admissions were generally positive across the three data sets, and model sensitivity was variable. For example, estimates for both categories were lower at higher degrees of freedom in the European data and in the U.S. cardiovascular admissions data, but not in the U.S. respiratory admissions data. The relative magnitude of the two associations with O_3 varied across the three data sets.

Previous reports exist for the associations between O_3 and hospital admissions for Canada and for APHEA1, but not for APHEA2 or NMMAPS. Burnett and colleagues (1997a) assessed the relationship between the maximum one-hour concentration of O_3 and hospitalization for respiratory diseases in 16 Canadian cities, 1981–1991. With control for other pollutants, they found a positive association from April to December but not in the remaining months. In APHEA1, O_3 was associated with hospitalizations for respiratory diseases (Anderson et al. 1997; Spix et al. 1998). The new findings confirm that O_3 remains associated with risk for hospitalization.

HETEROGENEITY AND EFFECT MODIFICATION

One objective of APHENA was to explore patterns of effect modification across a wide range of geographic regions with air pollution coming from differing source mixtures and with populations having differing underlying sociodemographic characteristics. In general, the

European cities tend to have a higher prevalence of diesel vehicles, particularly passenger cars, than the cities in North America. Although not characterized, source inventories related to power generation and industry likely also vary, both across the regions and within the regions. The United States has well-characterized differences in pollution sources between the eastern and western states. Coal-fired power plants substantially contribute to particulate mass in the eastern cities. Consequently, particles contain a higher sulphate mass than in the western cities. Motor vehicle emissions tend to dominate in the western cities, so the nitrate component of particle mass is greater there than in the eastern cities (U.S. EPA 2004b). There is also variation in distributions of the levels of PM and O_3 that could contribute to heterogeneity in concentration–response relationships, particularly if the underlying curves are nonlinear. Concentrations of copollutants also vary across the cities.

In prior analyses of single-city and multicity data, a number of potential modifiers of the risk of exposure to air pollution were identified; some were related to sociodemographic and health characteristics of the population (e.g., age, disease prevalence, and socioeconomic status), while others reflected individual disease status (O'Neill et al. 2003; National Research Council and Committee on Research Priorities for Airborne Particulate Matter 2004; Pope and Dockery 2006). One hypothesis is that a lower socioeconomic status increases vulnerability to air pollution, potentially through diverse mechanisms (O'Neill et al. 2003). Effects of socioeconomic status on risks associated with air pollution have been found in studies of acute and chronic effects of air pollution (Krewski et al. 2000; Pope and Dockery 2006). Housing characteristics may also act as effect modifiers by affecting the relationships between concentrations at ambient monitors and personal exposures. In an extension of NMMAPS analyses, Janssen and colleagues (2002) found that the proportion of homes with air conditioning modified the association of PM_{10} with mortality.

Within the APHEA project, prior analyses have identified significant PM_{10} effect modification patterns for both mortality and admissions outcomes (Katsouyanni et al. 1997; Atkinson et al. 2001; Katsouyanni et al. 2001; Le Tertre et al. 2002; Aga et al. 2003; Samoli et al. 2005; Analitis et al. 2006). Among exposure variables explored as potential effect modifiers, higher NO_2 concentrations were found to be associated with larger PM mortality effect estimates. The APHEA group also found that PM mortality effects were higher in warmer and drier sites, as well as in places with a higher proportion of older people in the population and longer life expectancy. For hospital admissions, some effect modification was identified in relation to O_3 concentration and to the proportion of older people in the population, and to various specific causes of death.

In NMMAPS, effect modification was explored extensively in the original analyses of the 90 cities mortality data — analyses which were not repeated after the identification of the S-Plus issue (Samet et al. 2000a). Exploratory analyses identified several potential modifiers; the most consistent of these was the level of PM_{10} and the percentage of residents not completing high school. The effect of PM_{10} on mortality decreased with increasing level of PM_{10} and increased with the percentage of residents not finishing high school. There was also an indication of variation by geographic region, with the highest mortality effect estimate coming from the northeastern United States. For hospital admissions, neither underlying rate of hospitalization nor sociodemographic factors were modifiers. Two-stage analyses similar to those conducted in Europe and in the United States were not carried out on the Canadian studies.

In APHENA, effect modification was approached in the second-stage analysis; variation in effect estimates was assessed for variables indicative of characteristics of the air pollution mixture, climate, age structure and health status, and socioeconomic status (SES) determinants. Because of the difficulty of finding SES variables that can be compared within European countries as well as between Europe and North America, only unemployment rate was used. The PM_{10} effect modification patterns, explored only for cities with daily data (i.e., 22 European and 15 U.S. cities), were not entirely consistent across centers, and varied somewhat with the underlying model and geographic location. Thus, the concentrations of pollutants modified the effects differently for cities in Europe and the United States. Climatic variables were important only in Europe. The most consistent evidence of effect modification was found for age in both the United States and Europe; an increasing proportion of older people in the population was associated with a greater effect of PM_{10}. A higher unemployment rate was also associated with greater risk in both regions (Appendix Tables E.2 and E.3).

The differences observed in how air pollutant concentrations affect the size of the mortality effect estimates in Europe and the United States may be attributed to differences in the pollution mixture. Thus, in Europe, higher annual average concentrations of NO_2 were associated with higher PM effects. While the same pattern was noted in the United States, it was less pronounced and the effect modification was not significant. This result may be related to the fact that, in contrast to the United States, a

large proportion of PM in Europe originates from diesel vehicles with emissions that raise NO_2 concentrations (European Commission 2005; Green Car Congress 2008). Temperature and humidity levels also modify PM effects in Europe; effects were estimated to be higher in warmer and drier locations. This pattern of effect modification by temperature and humidity was not evident in the U.S. cities. This difference may be due to climatic differences or to the much more extensive use of air conditioning in warm parts of the United States.

No consistent effect modification pattern could be identified for O_3. A more detailed exploration of effect modification was not possible because the necessary data were not uniformly available from all of the cities.

LIMITATIONS

The APHENA project was initiated to develop an analytic strategy to be applied to three major databases on air pollution and health. Using a two-stage process, the APHENA researchers were successful in meeting this goal. They explored a reasonable set of alternative models for the first-stage time-series analyses before finalizing the protocol. The second-stage model was based on a standard approach. The findings largely replicated earlier analyses and provided some indication of consistent effect modification, but they also showed that there may be different modification patterns due to different environmental, health, and socioeconomic conditions.

The principal limitations in interpreting the APHENA findings lie with the data available to the investigators. The data came from multiple countries and had not been collected according to a uniform protocol, but all had been edited, extensively analyzed, and had undergone a quality assessment audit. For the United States and Canada, nationwide reference methods for air pollution data collection were in place and deaths were captured and coded uniformly. The APHEA2 data came from 24 countries, and methods for their collection varied among the countries. Nonetheless, well-documented approaches were used to develop air pollution concentration values that would be as complete and uniform as possible. Given the nature of the APHENA project, the decision to begin with the existing databases, and the potential magnitude of the task in reconstructing all of the time series, particularly those for the APHEA cities, it was not possible to assemble a fully-harmonized database for the first-stage analysis. Consequently, differences in the time-series data as well as the possibility of differing measurement error structures across the three sets of data remain as potential contributors to heterogeneity.

A number of potential sources of measurement error may have affected the APHENA findings (Zeger et al. 2000). In general, bias toward the null would be anticipated, but the degree of bias could vary across the component databases. The relationships between ambient concentration and personal exposure could be different because of variations in time–activity patterns and housing stock. In addition, although PM_{10} concentrations were estimated for the majority of APHENA cities, the approach to estimation was location-specific.

Finally, gaps exist in the PM_{10} data because data had been collected according to regulatory monitoring protocols, which specified the routine collection of data only on every sixth day. These gaps reduced precision of estimates and also precluded the use of DL models in many of the U.S. and Canadian cities.

A main objective of the study was to explore the influence of determinants of variation in the effects of air pollution across the APHENA cities. While a number of potential effect modifiers have been identified in previous studies, the APHENA analyses were limited in exploring effect modification because the number of variables that were measured with sufficient commonality across the full data set was small. Basic demographic data were available, but unemployment rate was the only socioeconomic status indicator that extended across the full data set. Information was not available on the health status of the populations or on the frequency of specific diseases, such as diabetes, that have been postulated as effect modifiers. Similarly, data were not available on housing characteristics or the mix of sources across the APHENA cities.

SYNTHESIS AND FUTURE RESEARCH

This multiyear project led to the development of a standardized protocol for analyses of daily time-series data on mortality and hospitalization. Data were pooled from studies that had been carried out in multiple cities, a total of 126 in Europe and North America. In this report, the APHENA investigators summarize the common approach they developed for the analysis of multicity time-series data. The APHENA findings confirm the acute, adverse effects of PM_{10} and O_3 on mortality and hospitalization and provide risk estimates that have been developed with fully transparent methods. The analytic protocol should be of value for researchers carrying out single-city analyses and for future multicity studies.

Overall, the findings document that PM and O_3 adversely affect public health; APHENA results document that both pollutants are associated with increased mortality and hospital admissions in Europe and North America. While the data were collected as far back as 20 years, the findings

remain relevant to contemporary urban air pollution. The findings for mortality and hospitalization complement one another and are consistent in linking the air pollutants to cardiovascular and respiratory outcomes, and in showing greater risk for older persons. These findings draw biologic plausibility from a wide range of toxicologic studies on the mechanistic basis of adverse effects of PM and O_3, discussed in several recent reviews (U.S. EPA 2004a, 2006; Pope and Dockery 2006).

APHENA's extensive exploration of the sensitivity of findings to model selection and specification showed that the results were quite robust across a reasonable range of choices. The Canadian, APHEA, and NMMAPS data sets have now been extensively analyzed: in the original analyses, in the reanalyses with more rigid convergence criteria conducted in response to identification of the S-Plus GAM issue, and now in APHENA. Notably, while the sensitivity of the quantitative estimates to reanalysis has varied among the three data sets (Dominici et al. 2003), the overall weight of evidence from these key data sets has remained unchanged.

The APHENA investigators worked together for more than five years, from the initial planning to the preparation of the report. They encountered several barriers in carrying out the project, particularly related to coordinating multiple groups without a single coordinating center where all analyses could be carried out. Concerns about honoring agreements related to restricted data access led to the decision not to create a centralized database. In addition, second-stage analyses were limited by the extent of common data on potential effect modifiers across the participating regions; nonetheless, the findings were largely confirmatory. If resources had allowed the investigators to collect additional data, perhaps this limitation could have been addressed. Their experience indicates that greater emphasis needs to be placed on new data collection for effect modifiers if this limitation in APHENA is to be overcome by future combined analyses.

Finally, future research should consider whether APHENA is a useful model for carrying out combined analyses of air pollution data. Several factors suggest an ongoing need for appropriate methods for synthesizing and learning from air pollution research. The extent of research on air pollution and health has grown remarkably over the last several decades as the importance of air pollution as a public health problem has been reconfirmed. In addition, many developing countries face worsening air pollution, driven by industrialization and by an increasing number of motor vehicles. Periodic syntheses of the literature are necessary to inform decisions about air pollution control and to assess whether resultant declines in air pollution concentrations and changes in sources have led to changes in the risk. They represent an important type of accountability research (HEI Accountability Working Group 2003).

Active tracking of published reports and cataloguing their findings represents one approach to ongoing collection and synthesis of estimates on the health risks of air pollution. The database assembled by Anderson and colleagues (WHO 2004), funded by the United Kingdom Department of Health, exemplifies this approach. This resource has proved valuable for carrying out meta-analyses, and it was an important element of the evidence considered in developing the WHO Air Quality Guidelines (WHO 2006). Anderson and colleagues have used the database to document the publication bias that affects reports of single-city studies.

In carrying out APHENA, the investigators emphasized gaining an understanding of the heterogeneity of the effect of air pollution. Some initiatives assume that such heterogeneity may be relevant to public health and decision-making. In the United States, the Center for Disease Control's Environmental Public Health Tracking Program supports local- and state-level analyses of the effects of air pollution on health (McGeehin et al. 2004). For daily time-series analyses, APHENA provides methods and a context for carrying out and interpreting findings from this or other smaller-scale tracking efforts.

The APHENA approach makes possible more informative analyses than can be achieved with a meta-analytic summary. The data can be analyzed with multiple models to assess the sensitivity of findings to model specification, and effect modification can be assessed using risk estimates coming from a common analytic approach. Given the strengths of the approach, the investigators recommend that periodic pooling exercises like APHENA be conducted. Other alternatives include the ongoing collection of time-series data sets into a single database so that analyses can be carried out periodically, or the data can be made generally available, as has been done with NMMAPS. Peng and colleagues (2006b) have proposed that reproducible research be fostered by making data generally available to the scientific community. Ultimately, robust findings that are relevant to decision making should be repeatable.

ACKNOWLEDGMENTS

The investigators thank Debra Moffitt, M.S. for project administration and Charlotte Gerczak, M.L.A. and Athena Foong for editorial assistance.

REFERENCES

Aga E, Samoli E, Touloumi G, Anderson HR, Cadum E, Forsberg B, Goodman P, Goren A, Kotesovec F, Kriz B, Macarok-Hiti M, Medina S, Paldy A, Schindler C, Sunyer J, Tittanen P, Wojtyniak B, Zmirou D, Schwartz J, Katsouyanni K. 2003. Short-term effects of ambient particles on mortality in the elderly: Results from 28 cities in the APHEA2 project. Eur Respir J Suppl 40:28s–33s.

Ackermann-Liebrich U, Leuenberger P, Schwartz J, Schindler C, Monn C, Bolognini G, Bongard JP, Brändli O, Domenighetti G, Elsasser S, Grize L, Karrer W, Keller R, Keller-Wossidlo H, Künzli N, Martin BW, Medici TC, Perruchoud AP, Schöni MH, Tschopp JM, Villiger B, Wüthrich B, Zellweger JP, Zemp E. 1997. Lung function and long term exposure to air pollutants in Switzerland. Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) Team. Am J Respir Crit Care Med 155(1): 122–129.

Almon S. 1965. The distributed lag between capital appropriations and expenditures. Econometrica 33:178–196.

Analitis A, Katsouyanni K, Dimakopoulou K, Samoli E, Nikoloulopoulos AK, Petasakis Y, Touloumi G, Schwartz J, Anderson HR, Cambra K, Forastiere F, Zmirou D, Vonk JM, Clancy L, Kriz B, Bobvos J, Pekkanen J. 2006. Short-term effects of ambient particles on cardiovascular and respiratory mortality. Epidemiology 17(2):230–233

Anderson HR, Atkinson RW, Peacock JL, Sweeting MJ, Marston L. 2005. Ambient particulate matter and health effects: Publication bias in studies of short-term associations. Epidemiology 16:155–163.

Anderson HR, Spix C, Medina S, Schouten J, Castellsague J, Rossi G, Zmirou D, Touloumi G, Wojtyniak B, Pönkä A, Bachárová L, Schwartz J, Katsouyanni K. 1997. Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: Results from the APHEA project. Eur Respir J 10:1064–1071.

Atkinson RW, Anderson HR, Sunyer J, Ayres J, Baccini M, Vonk JM, Boumghar A, Forastiere F, Forsberg B, Touloumi G, Schwartz J, Katsouyanni K. 2001. Acute effects of particulate air pollution on respiratory admissions: Results from APHEA 2 project. Air Pollution and Health: A European Approach. Am J Respir Crit Care Med 164:1860–1866.

Atkinson RW, Anderson HR, Sunyer J, Ayres J, Baccini M, Vonk JM, Boumghar A, Forastiere F, Forsberg B, Touloumi G, Schwartz J, Katsouyanni K. 2003. Acute effects of particulate air pollution on respiratory admissions. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report. pp 81–84. Health Effects Institute, Boston, MA.

Bell ML, Dominici F, Samet JM. 2005. A meta-analysis of time-series studies of ozone and mortality with comparison to the national morbidity, mortality, and air pollution study. Epidemiology 16:436–445.

Bell ML, McDermott A, Zeger SL, Samet JM, Dominici F. 2004a. Ozone and short-term mortality in 95 US urban communities, 1987–2000. JAMA 292:2372–2378.

Bell ML, Samet JM, Dominici F. 2004b. Time-series studies of particulate matter. Annu Rev Public Health 25:247–280.

Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. 1998. Meta-analysis of multiple outcomes by regression with random effects. Stat Med 17:2537–2550.

Burnett RT, Brook J, Dann T, Delocla C, Philips O, Cakmak S, Vincent R, Goldberg MS, Krewski D. 2000. Association between particulate- and gas-phase components of urban air pollution and daily mortality in eight Canadian cities. Inhal Toxicol 4:15–39.

Burnett RT, Brook JR, Yung WT, Dales RE, Krewski D. 1997a. Association between ozone and hospitalization for respiratory diseases in 16 Canadian cities. Environ Res 72:24–31.

Burnett RT, Cakmak S, Brook JR. 1998. The effect of the urban ambient air pollution mix on daily mortality rates in 11 Canadian cities. Can J Public Health 89:152–156.

Burnett RT, Cakmak S, Brook JR, Krewski D. 1997b. The role of particulate size and chemistry in the association between summertime ambient air pollution and hospitalization for cardiorespiratory diseases. Environ Health Perspect 105:614–620.

Burnett RT, Dales R, Krewski D, Vincent R, Dann T, Brook JR. 1995. Associations between ambient particulate sulfate and admissions to Ontario hospitals for cardiac and respiratory diseases. Am J Epidemiol 142:15–22.

Burnett RT, Goldberg MS. 2003. Size-fractionated particulate mass and daily mortality in eight Canadian cities. Revised analyses of time-series studies of air pollution and health. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report. pp 85–89. Health Effects Institute, Boston, MA.

Burnett RT, Smith-Doiron M, Stieb D, Cakmak S, Brook JR. 1999. Effects of particulate and gaseous air pollution on

cardiorespiratory hospitalizations. Arch Environ Health 54:130–139.

Burnett RT, Stieb D, Brook JR, Cakmak S, Dales R, Raizenne M, Vincent R, Dann T. 2004. Associations between short-term changes in nitrogen dioxide and mortality in Canadian cities. Arch Environ Health 59(5):228-236.

Cadum E, Rossi G, Mirabelli D, Vigotti MA, Natale P, Albano L, Marchi G, Di Meo V, Cristofani R, Costa G. 1999. Air pollution and daily mortality in Turin, 1991–1996. Epidemiol Prev 23:268–276.

Chen Y, Yang Q, Krewski D, Shi Y, Burnett RT, McGrail K. 2004. Influence of relatively low level of particulate air pollution on hospitalization for COPD in elderly people. Inhal Toxicol 16:21–25.

Commission of European Communities. 1999. Council Directive 1999/30/EC relating to limit values for sulphur dioxide, oxides of nitrogen, particulate matter and lead in the ambient air. Official Journal of the European Communities 163:41–60.

Committee on Research Priorities for Airborne Particulate Matter, National Research Council. 2004. Research Priorities for Airborne Particulate Matter: IV. Continuing Research Progress. National Academies Press, Washington, D.C.

D'Innocenzio F, Di Filippo P, Lepore L, Marconi A. 1998. PM_{10} and $PM_{2,5}$ concentrations in urban air and size fraction distribution of toxic metals. Annali di chimica 88:281–289.

Daniels M, Dominici F, Samet JM, Zeger SL. 2000. Estimating particulate matter-mortality dose–response curves and threshold levels: An analysis of daily time-series for the 20 largest US cities. Am J Epidemiol 152:397–406.

DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. Control ClinTrials 7:177–188.

Dominici F. 2002. Invited commentary: Air pollution and health-what can we learn from a hierarchical approach? Am J Epidemiol 155:11–15.

Dominici F, Daniels M, Zeger SL, Samet JM. 2000a. National models for estimating the effect of particulate matter on mortality in US cities. (Technical Report). Johns Hopkins University, Baltimore, MD.

Dominici F, Daniels M, Zeger SL, Samet JM. 2002a. Air pollution and mortality: Estimating regional and national dose-response relationships. J Am Stat Assoc 97:100–111.

Dominici F, McDermott A, Daniels M, Zeger SL, Samet JM. 2003. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report. pp 9–24. Health Effects Institute, Boston, MA.

Dominici F, McDermott A, Hastie TJ. 2004. Improved semiparametric time series models of air pollution and mortality. J Am Stat Assoc 99:938–948.

Dominici F, McDermott A, Zeger SL, Samet JM. 2002b. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol 156:193–203.

Dominici F, Samet J, Zeger SL. 2000b. Combining evidence on air pollution and daily mortality from the largest 20 U.S. cities: A hierarchical modeling strategy (with Discussion). J R Stat Soc [Ser A] 163:263–302.

European Commission. 2005. Impact Assessment. Commission Staff Working Paper. Annex to: The Communication on Thematic Strategy on Air Pollution and the Directive on "Ambient Air Quality and Cleaner Air for Europe" COM(2005) 446 final. European Commission (last updated 4/25/08). *http://ec.europa.eu/environment/archives/air/cafe/index.htm.* Accessed 5/13/08.

Everson P, Morris C. 2000. Inference for multivariate normal hierarchical models. J Royal Stat Soc (B) 62:399–412.

Fairley D. 1990. The relationship of daily mortality to suspended particulates in Santa Clara County, 1980-1986. Environ Health Perspect 89:159–168.

Gehrig R, Hofer P. Dübendorf: EMPA; 1999. Vergleich von TSP-, PM_{10}- und $PM_{2.5}$-Schwebestaubmessungen im NABEL 1997 und 1998. Nr. 200'039/168'107.8600.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2003. Bayesian Data Analysis. Chapman and Hall Inc., New York, NY.

Gianelle V, Ravasini G, Ventura A. 1991. Misure di aerosol atmosferico nell'area urbana Milanese. Aggiornamento al 31 Marzo, Comune di Milano.

Green Car Congress. 2008. European automobile production grows by 5.3% in 2007; Diesel accounts for 53.5% of new car registrations (last updated 2/20/2008). *www.greencarcongress.com/2008/02/european-automo.html*. Accessed 5/12/08.

Green PJ, Silverman BW. 1994. Nonparametric Regression and Generalized Linear Models: A roughness penalty approach. Chapman and Hall Inc., New York, NY.

Gryparis A, Forsberg B, Katsouyanni K, Analitis A, Touloumi G, Schwartz J, Samoli E, Medina S, Anderson HR,

Niciu EM, Wichmann HE, Kriz B, Kosnik M, Skorkovsky J, Vonk JM, Dörtbudak Z. 2004. Acute effects of ozone on mortality from the "Air pollution and health: A European approach" project. Am J Respir Crit Care Med 170:1080–1087.

Hastie TJ, Tibshirani RJ. 1990. Generalized Additive Models (Monographs on Statistics and Applied Probability). Chapman and Hall Inc., New York, NY.

Hatzakis A, Katsouyanni K, Kalandidi A, Day N, Trichopoulus D. 1986. Short-term effects of air pollution on mortality in Athens. Int J Epidemiol 15:73–81.

HEI Accountability Working Group. 2003. Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research. Communication 11. Health Effects Institute, Boston, MA.

Health Effects Institute. 2000. 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. Health Effects Institute, Cambridge, MA.

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

Janssen NA, Schwartz J, Zanobetti A, Suh HH. 2002. Air conditioning and source-specific particles as modifiers of the effect of PM(10) on hospital admissions for heart and lung disease. Environ Health Perspect 110:43–49.

Katsouyanni K, Touloumi G, Spix C, Schwartz J, Balducci F, Medina S, Rossi G, Wojtyniak B, Sunyer J, Bachárová L, Schouten JP, Pönkä A, Anderson HR. 1997. Short-term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: Results from time series data from the APHEA project. Air Pollution and Health: A European Approach. BMJ 314:1658–1663.

Katsouyanni K, Schwartz J, Spix C, Touloumi G, Zmirou D, Zanobetti A, Wojtyniak B, Vonk JM, Tobias A, Pönkä A, Medina S, Bachárová L, Anderson HR. 1996. Short term effects of air pollution on health: A European approach using epidemiologic time series data: The APHEA protocol. J Epidemiol Community Health 50 Suppl 1:S12–S18.

Katsouyanni K, Touloumi G, Samoli E, Gryparis A, Le Tertre A, Monopolis Y, Rossi G, Zmirou D, Ballester F, Boumghar A, Anderson HR, Wojtyniak B, Paldy A, Brainstein R, Pekkanen J, Schindler C, Schwartz J. 2001. Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 European cities within the APHEA2 project. Epidemiology 12:521–531.

Kelsall JE, Samet JM, Zeger SL, Xu J. 1997. Air pollution and mortality in Philadelphia, 1974–1988. Am J Epidemiology 146:750–762.

Le Tertre A, Medina S, Samoli E, Forsberg B, Michelozzi P, Boumghar A, Vonk JM, Bellini A, Atkinson R, Ayres JG, Sunyer J, Schwartz J, Katsouyanni K. 2002. Short-term effects of particulate air pollution on cardiovascular diseases in eight European cities. J Epidemiol Community Health 56:773–779.

Le Tertre A, Medina S, Samoli E, Forsberg B, Michelozzi P, Boumghar A, Vonk JM, Bellini A, Atkinson R, Ayres JG, Sunyer J, Schwartz J, Katsouyanni K. 2003. Short-term effects of particulate air pollution on cardiovascular diseases in eight European cities. Revised analyses of time-series studies of air pollution and health. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report. pp 173–178. Health Effects Institute, Boston, MA.

Le Tertre A, Schwartz J, Touloumi G. 2005. Empirical Bayes and adjusted estimates approach to estimating the relation of mortality to exposure of PM(10). Risk Anal 25:711–718.

Maclure M. 1991. The case-crossover design: A method for studying the transient effects of risk of acute events. Am J Epidemiol 133:144–153.

McCullagh P, Nelder JA. 1983. Generalized linear models. Chapman and Hall, Inc., London.

McCullagh P, Nelder JA. 1989. Generalized linear models. Chapman and Hall, Inc., New York, NY.

McGeehin MA, Qualters JR, Niskar AS. 2004. National environmental public health tracking program: Bridging the information gap. Environ Health Perspect 112:1409–1413.

Medina-Ramón M, Zanobetti A, Schwartz J. 2006. The effect of ozone and PM_{10} on hospital admissions for pneumonia and chronic obstructive pulmonary disease: A national multicity study. Am J Epidemiol 163:579–588.

Morris CN, Normand S-L. 1992. Hierarchical models for combining information and for meta-analysis. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, eds. Bayesian Statistics 4. Proceeding of the Fourth Valencia International Meeting. Oxford University Press, New York, NY.

O'Neill MS, Jerrett M, Kawachi I, Levy JI, Cohen AJ, Gouveia N, Wilkinson P, Fletcher T, Cifuentes L, Schwartz J; Workshop on Air Pollution and Socioeconomic Conditions. 2003. Health, wealth, and air pollution: Advancing theory and methods. Environ Health Perspect 111:1861–1870.

Peng RD, Dominici F, Louis T. 2006a. Model choice in multi-site time series studies of air pollution and mortality. J R Stat Soc Ser A 169:179–203.

Peng RD, Dominici F, Welty LJ. 2007. A Bayesian hierarchical model for constrained distributed lag functions: Estimating the time course of hospitalization associated with air pollution exposure. Johns Hopkins University, Department of Biostatistics Working Papers. Working Paper 128. Available from *www.bepress.com/jhubiostat/paper128*.

Peng RD, Dominici F, Zeger SL. 2006b. Reproducible epidemiologic research. Am J Epidemiol 163:783–789.

Pope CA III, Dockery DW. 2006. Health effects of fine particulate air pollution: Lines that connect. J Air Waste Manag Assoc 56:709–742.

Pope CA III, Kalkstein LS. 1996. Synoptic weather modeling and estimates of the exposure-response relationship between daily mortality and particulate air pollution. Environ Health Perspect 104:414–420.

Ramsay TO, Burnett RT, Krewski D. 2003. The effect of concurvity in generalized additive models linking mortality and ambient air pollution. Epidemiology 14:18–23.

Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL. 2000a. Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. N Engl J Med 343:1742–1749.

Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000b. The National Morbidity, Mortality, and Air Pollution Study. Part I. Methods and Methodologic Issues. Research Report 94. Health Effects Institute, Cambridge, MA.

Samet JM, Zeger SL, Dominici F, Curriero F, Coursac I, Dockery DW, Schwartz J, Zanobetti A. 2000c. 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.

Samet JM, Zeger SL, Kelsall JE, Xu J, Kalkstein LS. 1997. Particulate Air Pollution and Daily Mortality: Analyses of the Effects of Weather and Multiple Air Pollutants. The Phase I.B Report of the Particle Epidemiology Evaluation Project. Special Report. Health Effects Institute, Cambridge, MA.

Samet JM, Zeger S, Kelsall JE, Xu J, Kalkstein LS. 1998. Does weather confound or modify the association of particulate air pollution with mortality? An analysis of the Philadelphia data, 1973–1980. Environ Res 77:9–19.

Samoli E, Analitis A, Touloumi G, Schwartz J, Anderson HR, Sunyer J, Bisanti L, Zmirou D, Vonk JM, Pekkanen J, Goodman P, Paldy A, Schindler C, Katsouyanni K. 2005. Estimating the exposure-response relationships between particulate matter and mortality within the APHEA multicity project. Environ Health Perspect 113:88–95.

Samoli E, Peng R, Ramsay T, Pipikou M, Touloumi G, Dominici F, Burnett R, Cohen A, Krewski D, Samet J, Katsouyanni K. 2008. Acute Effects of Ambient Particulate Matter on Mortality in Europe and North America: Results from the APHENA Study. Environ Health Perspect 116(11):1480–1486.

Samoli E, Touloumi G, Zanobetti A, Le Tertre A, Schindler C, Atkinson R, Vonk J, Rossi G, Saez M, Rabczenko D, Schwartz J, Katsouyanni K. 2003. Investigating the dose-response relation between air pollution and total mortality in the APHEA-2 multicity project. Occup Environ Med 60:977–982.

Schwartz J. 1994. Total suspended particulate matter and daily mortality in Cincinnati, Ohio. Environ Health Perspect 102:186–189.

Schwartz J. 2000a. Harvesting and long term exposure effects in the relationship between air pollution and mortality. Am J Epidemiol 151:440–448.

Schwartz J. 2000b. The distributed lag between air pollution and daily deaths. Epidemiology 11:320–326.

Schwartz J. 2001. Is there harvesting in the association of airborne particles with daily deaths and hospital admissions? Epidemiology 12:55–61.

Schwartz J. 2004. Is the association of airborne particles with daily deaths confounded by gaseous air pollutants? An approach to control by matching. Environ Health Perspect 112:557–561.

Schwartz J. 2005. How sensitive is the association between ozone and daily deaths to control for temperature? Am J Respir Crit Care Med 171:627–631.

Schwartz J. Dockery DW. 1992. Increased mortality in Philadelphia associated with daily air pollution concentrations. Am Rev Respir Dis 145(3):600–604.

Schwartz J, Marcus A. 1990. Mortality and air pollution in London: A time series analysis. Am J Epidemiol 131:185–194.

Schwartz J, Spix C, Touloumi G, Bachárová L, Barumamdzadeh T, Le Tertre A, Piekarksi T, Ponce de Leon A, Pönkä A, Rossi G, Saez M, Schouten JP. 1996. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. J Epidemiol Community Health 50:S3–S11.

Schwartz J, Zanobetti A. 2000. Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. Epidemiology 11:666–672.

Speckman P. 1988. Kernel smoothing in partial linear models. J Royal Stat Ser B 50:413–436.

Spix C, Anderson HR, Schwartz J, Vigotti MA, Le Tertre A, Vonk JM, Touloumi G, Balducci F, Piekarski T, Bachárová L, Tobias A, Pönkä A, Katsouyanni K. 1998. Short-term effects of air pollution on hospital admissions of respiratory diseases in Europe: A quantitative summary of APHEA study results. Air Pollution and Health: A European Approach. Arch Environ Health 53:54–64.

Stieb DM, Judek S, Burnett RT. 2002. Meta-analysis of time-series studies of air pollution and mortality: Effects of gases and particles and the influence of cause of death, age, and season. J Air Waste Manag Assoc 52:470–484.

Stieb DM, Judek S, Burnett RT. 2003. Meta-analysis of time-series studies of air pollution and mortality: Update in relation to the use of generalized additive models. J Air Waste Manag Assoc 53:258–261.

Touloumi G, Atkinson R, Le Tertre A, Samoli E, Schwartz J, Schindler C, Vonk JM, Rossi G, Saez M, Rabszenko D, Katsouyanni K. 2004. Analysis of health outcome time series data in epidemiological studies. Environmetrics 15:101–117.

Touloumi G, Samoli E, Pipikou M, Le Tertre A, Atkinson R, Katsouyanni K. 2006. Seasonal confounding in air pollution and health time-series studies: Effect on air pollution effect estimates. Stat Med 25:4164–4178.

Touloumi G, Samoli E, Pipikou M, Le Tertre A, Katsouyanni K, and on behalf of the APHEA-2 Group. 2003. Confounding factors in time-series studies of air pollution and health: Effects of different adjustment methods. In: Proceeding of the 18th International Workshop on Statistical Modeling. July 11–13, 2003. Leuven, Belgium.

Touloumi G, Samoli E, Quenel P, Paldy A, Anderson RH, Zmirou D, Galan I, Forsberg B, Schindler C, Schwartz J, Katsouyanni K. 2005. Short-term effects of air pollution on total and cardiovascular mortality: The confounding effect of influenza epidemics. Epidemiology 16:49–57.

U.S. Environmental Protection Agency. 2004a. Air quality criteria for particulate matter. EPA/600/p-99/022aD and bD. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC.

U.S. Environmental Protection Agency. 2004b. The Particle Pollution Report: Current Understanding of Air Quality and Emissions through 2003 (last updated October 1, 2007). Available from *http://epa.gov/airtrends/aqtrnd04/pm.html*.

U.S. Environmental Protection Agency. 2006. Air Quality Criteria for Ozone and Other Photochemical Oxidants. EPA/600/R-05/004aF-cF. National Center for Environment Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

Welty LJ, Zeger SL. 2005. Are the acute effects of particulate matter on mortality in the National Morbidity, Mortality, and Air Pollution Study the result of inadequate control for weather and season? A sensitivity analysis using flexible distributed lag models. Am J Epidemiol 162:80–88.

Welty LJ, Zeger SL, Dominici F. 2005. Bayesian distributed lag models: Estimating effects of particulate matter air pollution on daily mortality. Johns Hopkins University, Department of Biostatistics Working Papers. Working Paper 96. Available from *http://www.bepress.com/jhubiostat/paper96*.

Wood SN. 2000. Modeling and smoothing parameter estimation with multiple quadratic penalties. J Royal Stat Ser B 62:413–428.

World Health Organization. 1979. International Statistical Classification of Diseases and Related Health Problems, 9th revision. World Health Organization, Geneva, Switzerland.

World Health Organization. 1992. International Classification of Disease and Related Health Problems, 10th revision. World Health Organization, Geneva, Switzerland.

World Health Organization. 2000. Air Quality Guidelines for Europe. WHO Regional Publications, European Series, No. 91. World Health Organization Regional Office for Europe, Copenhagen, Denmark.

World Health Organization. 2004. Meta-analysis of time series studies and panel studies of particulate matter (PM) and ozone (O_3). Report of a WHO task group. World Health Organization, Copenhagen, Denmark.

World Health Organization. 2006. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide. Global update 2005. Summary of risk assessment. WHO/SDE/PHE/OEH/06.02. World Health Organization, Geneva, Switzerland.

Zanobetti A, Schwartz J. 2003. Airborne particles and hospital admissions for heart and lung disease. Revised analyses of time-series studies of air pollution and health. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report. pp 241–248. Health Effects Institute, Boston, MA.

Zanobetti A, Schwartz J, Samoli E, Gryparis A, Touloumi G, Atkinson R, Le Tertre A, Bobros J, Celko M, Goren A, Forsberg B, Michelozzi P, Rabczenko D, Aranguez RE, Katsouyanni K. 2002. The temporal pattern of mortality responses to air pollution: A multicity assessment of mortality displacement. Epidemiology 13:87–93.

Zanobetti A, Schwartz J, Samoli E, Gryparis A, Touloumi G, Peacock J, Anderson RH, Le Tertre A, Bobros J, Celko M, Goren A, Forsberg B, Michelozzi P, Rabczenko D, Hoyos SP, Wichmann HE, Katsouyanni K. 2003. The temporal pattern of respiratory and heart disease mortality in response to air pollution. Environ Health Perspect 111:1188–1193.

Zanobetti A, Wand M, Schwartz J. 2000. Generalized additive distributed lag models: Quantifying mortality displacement. Biostatistics 1:279–292.

Zeger SL, Dominici F, Samet J. 1999. Harvesting-resistant estimates of air pollution effects on mortality. Epidemiology 10:171–175.

Zeger SL, Irizarry R, Peng RD. 2006. On time series analysis of public health and biomedical data. Annu Rev Public Health 27:57–79.

Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, Cohen A. 2000. Exposure measurement error in time-series studies of air pollution: concepts and consequences. Environ Health Perspect 108:419–426.

APPENDIX A. Simulation Study

Table A.1. Results from 250 Replications[a]

	Fixed Effects		Random Effects	
Method	% (95%CI)	% Bias	% (95%CI)	% Bias
LOESS (PACF/AIC)[b]	0.6537 (0.558, 0.750)	5.64	0.6686 (0.392, 0.946)	8.05
LOESS (BIC)	0.7141 (0.618, 0.810)	15.37	0.7276 (0.464, 0.992)	17.55
LOESS (AIC)	0.6679 (0.572, 0.764)	7.93	0.6776 (0.413, 0.943)	9.49
NS (PACF/AIC)[b]	0.4688 (0.352, 0.586)	−24.17	0.5248 (0.255, 0.795)	−15.13
NS (BIC)	0.4293 (0.313, 0.546)	−30.54	0.4949 (0.226, 0.765)	−19.96
NS (AIC)	0.4413 (0.323, 0.559)	−28.61	0.4913 (0.226, 0.757)	−20.54
NS (fixed *df*)[c]	0.5729 (0.458, 0.688)	−7.37	0.5885 (0.307, 0.871)	−4.86
SS (fixed *df*)[c]	0.5767 (0.481, 0.673)	−6.76	0.6139 (0.317, 0.912)	−0.77
PS (GCV)[d]	0.5636 (0.450, 0.677)	−8.87	0.6456 (0.324, 0.968)	4.33
PS (fixed *df*)[b]	0.5957 (0.481, 0.710)	−3.70	0.6121 (0.321, 0.904)	−1.06
PS (PACF)	0.6113 (0.497, 0.726)	−1.19	0.6616 (0.367, 0.957)	6.92

[a] Percent change and 95% confidence intervals (CI) in all-cause mortality per 10-μg/m^3 increase in PM_{10} by the smoothing method. The percentage of bias is also shown. True percentage change was 0.619 per 10-μg/m^3 increase in PM_{10}.

[b] The PACF criterion was used to select the average *df* for time and the AIC for temperature and humidity.

[c] The total *df* were 7/year for time, 6 for temperature, and 3 for humidity.

[d] For the GCV, the number of basis functions was set equal to 30 for time and to 10 for temperature and humidity.

APPENDIX B. APHENA Methodology Protocol

This document contains a description of outcome and exposure variables and models used for the city-specific analysis of the APHENA project.

OUTCOMES

Mortality Data

All nonaccidental (ICD-9 < 800) causes of death for all ages, for people < 75 years, and for people ≥ 75 years.

Cardiovascular (ICD-9 390–459) causes of death for people < 75 years, and for people ≥ 75 years.

Respiratory causes of death (ICD-9 460–519) for all ages, for people < 75 years, and for people ≥ 75 years.

Hospital Admissions Data

Cardiovascular hospital admissions without stroke (ICD-9 390–429) and respiratory hospital admissions (ICD-9 460–519) for people < 65 years, and for people ≥ 65 years.

EXPOSURE

PM_{10} (24-hour average) and ground level O_3 (daily 1-hour maximum). Models will be run using PM_{10} alone, O_3 alone, and PM_{10} + O_3. Additionally, summer-only models will be run for O_3.

DATA SETS

It was decided to follow two kinds of analyses depending on the availability of measurements in each city:

- Part 1 concerns cities with complete time-series (i.e., daily data, possibly with a few nonsystematic missing data. Note: all cities have daily O_3 data and will be included in this category for O_3 analyses, but data availability for PM_{10} varies.)
- Part 2 concerns all cities, regardless of data availability.

MODEL SPECIFICATION

Mortality and Hospital Admissions

- Exposure lags, Part 1.
  - The exposure for the average lags of days 0 and 1.
  - DL models: 3 lags (i.e., day 0, 1, 2) for mortality and 5 lags for respiratory and cardiovascular disease admissions, using unconstrained DL models.
- Exposure lags, Part 2.
  - Lag 1 effects will be analyzed.
- Smoothing: NS and PS will be used as smooth functions for trend, seasonality and temperature control.
- Missing values: We delete the days with missing values BEFORE we fit the model (to ensure correct placement of the knots). This is particularly important in S+. In R it does not matter, because it does it on its own.
- Degrees of freedom for seasonality: In the analysis of complete and incomplete time-series we use 3, 8, and 12 *df*/year. Thus, the NS models will be fit with 3, 8, and 12 *df*/year. For the PS method, for each city we will apply a model with $k = 50 \times$ number of years (k is the number of basis functions) and use the smoothing parameter option to get to the degrees of freedom described above for the NS model. Additionally for complete time-series analysis (Part 1), we use the criterion of minimization of the PACF to choose the optimal degrees of freedom when we apply the PS method for control of seasonality. The PACF will be minimized for lags 1 through 30 for mortality. We will use sum of the absolute value PACF. We decided to incorporate a test for white noise. The degrees of freedom chosen by PACF in the PS method are then applied to a model using NS for seasonality control.

  We will complement this by an extensive sensitivity analysis, presenting results for 2, 6, 8, 10, 12, 14, 16, 18, and 22 *df*/year for both NS and PS models. The sensitivity analysis for mortality will be done for all-cause mortality > 75 years and for respiratory mortality, all ages, for lag 1.
- Covariates in the model
  - Temperature: Smooth term for lag 0. Smooth term for lag 1. When we do DL analyses for the pollutants, we include the same DL term for temperature, as we do for the pollutants.
  - Degrees of freedom for the temperature smooth terms: 3 for each, the same for all cities. This decision is backed up by the sensitivity analyses results.
  - Relative humidity or dew point temperature will NOT be included in the models, based on all preliminary results.
  - No control for influenza periods will be included in the models, based on sensitivity results.
  - Dummy variables for day of the week and holiday effect will be included.
- Age groups
  - Separate analyses will be applied for each age group.

- Autoregressive terms
  - For models based on minimization of PACF criterion, after looking at the PACF, autoregressive terms will be introduced if appropriate (i.e., if there is still autocorrelation).

Hospital Admissions

For admissions analyses we decided on a few additional features:

- It will be possible to use different degrees of freedom for different periods within a year, for example, possible smooth terms for the summer, especially if we have general (not emergency) admissions; or respiratory infections in the winter.
- We will control for influenza through the separate smoother for winter respiratory epidemics if appropriate.
- We confirmed that we will not control for humidity.
- The sum of the PACF for admissions will be for lags 8 through 30.

O_3 Seasonal Analysis

In the model with only half-year data, we will use dummy variables for the months (by year) instead of splines. Summer (or rather warm season) will be April through September. Temperature lags for summer-only models: for NS we use 2 *df* for lag 0 and linear term for lag 1. For PS we use 3 *df* for lag 0 and linear term for lag 1. NOTE: For annual O_3 models, we use the same temperature terms as for PM_{10}.

PACF

We minimized the PACF of the residuals from the city-specific models using the following procedure. Given a specific degree of freedom for the time trend, we fit a GLM or GAM to the data and obtained the residuals. We then estimated the PACF from the residuals and summed the absolute value of the coefficients corresponding to lags 1 through 30. This sum was then minimized with respect to the degrees of freedom for the time trend.

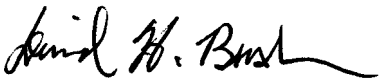
APPENDIX C. HEI Quality Assurance Statement

The conduct of this study was subjected to independent audits by Mr. David Bush of T&B Systems, Inc. Mr. Bush is an expert in quality assurance for air quality monitoring studies and data management. The audits included on-site reviews of study activities for conformance to the study protocol and operating procedures. The dates of the audits are listed below with the phase of the study examined.

QUALITY ASSURANCE AUDITS

November 2006:	The auditor conducted on-site audits at the three principal APHENA analysis centers in Baltimore, Ottawa, and Athens. The audit consisted of a review of the existing processing programs, management activities, and documentation, and interviews with key personnel. No problems significantly affecting study data were noted, though some recommendations were presented for clarifying inherent differences between the U.S., Canadian, and European data sets.
April 2009:	The auditor reviewed the study final report. Minor comments were provided for clarifying presentation of some of the data.

Written reports of each inspection were provided to the HEI project manager, who transmitted the findings to the Principle Investigators. These quality assurance audits demonstrated that the study was conducted by an experienced team with a high concern for data quality. The report appears to be an accurate representation of the study.



David H. Bush, Quality Assurance Officer

APPENDICES AVAILABLE ON THE WEB

Appendices D and E contain supplemental material not included in the printed report. They are available on the HEI Web site *http://pubs.healtheffects.org*. They may also be requested by contacting the Health Effects Institute at 101 Federal Street, Suite 500, Boston, MA 02110, +1-617-488-2300, fax +1-488-2335, or e-mail (*pubs@healthef fects.org*). Please give (1) the first author, full title, and number of the Research Report and (2) the title of the appendix requested.

Appendix D. Mortality and Hospital Admission Tables Describing the Canada, Europe, and United States Data Sets

Appendix E. Effect Modifier Tables for Canada, Europe, and the United States

ABOUT THE AUTHORS

H. R. Anderson qualified in medicine in Australia in 1964. From 1966 to 1972, he worked in Papua, New Guinea, where he investigated chronic lung disease and asthma and their relation to indoor air pollution and other factors. He moved to Britain in 1972 and spent two years at the Medical Research Council's Pneumoconiosis Unit in South Wales. He later received an M.Sc. in Social Medicine at the London School of Hygiene and Tropical Medicine. In 1976 he was appointed to St. George's, University of London, and became professor of epidemiology and public health in 1985. His main research is in the epidemiology of asthma and the health effects of air pollution. He is a member of the steering group of the European Union-funded multicity European study of the acute effects of air pollution on health (APHEA project), and the International Study of Asthma and Allergies in Childhood. He is a member of the United Kingdom Committee on the Medical Effects of Air Pollution, the Expert Panel on Air Quality Standards, the HEI Review Committee, and a number of WHO working groups.

Richard Atkinson studied mathematics at Bradford University from 1982 to 1985 before pursuing a career in industry with British Gas. During his nine years with the company he obtained an M.Sc. in statistics and operational research. In 1994 Richard joined St. George's in Tooting, South West London as a statistician where he worked initially with Professor Arthur Crisp in mental health sciences. He then joined the Department of Community Health Sciences where he is currently a senior research fellow. His first project was a study of risk factors for cardiovascular diseases in the local community in Wandsworth. Since then he has worked on a number of projects investigating short-term temporal associations between air pollution and health. Initial work funded by the United Kingdom led to his participation in the European-funded project APHEA. In 2000 he was awarded a COLT Foundation Ph.D. studentship to study the effects of airborne fungal spores on asthma exacerbations. This led to the awarding of his Ph.D. in 2006. He is currently a member of the Department of Health subcommittee undertaking a review of the evidence for health effects of air pollution, and has undertaken systematic reviews for the Department of Health, the WHO, and HEI.

Richard Burnett received his Ph.D. from Queen's University in 1982 in mathematical statistics. He is a senior research scientist with the Healthy Environments and Consumer Safety Branch of Health Canada, where he has been working since 1983 on issues relating to the health effects of outdoor air pollution. He is an adjunct professor in the Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa. Dr. Burnett's work has focused on the use of administrative health and environmental information to determine the public health impacts of combustion-related pollution using nonlinear random-effects models, time series, and spatial analytical techniques. Dr. Burnett is currently on the editorial board of *Risk Analysis* and is an external reviewer for numerous scientific journals, the U.S. EPA, and WHO.

Francesca Dominici is a professor of biostatistics at the Johns Hopkins Bloomberg School of Public Health. Dr. Dominici received her doctoral degree from the University of Padua in 1997. Her research interests include time-series analysis in environmental epidemiology, semiparametric Bayesian models for concentration–response data, hierarchical models for combining information and for meta-analysis, and missing data. She is the first recipient of the HEI Walter A. Rosenblith New Investigator Award. Her work on air pollution and mortality was honored by an invitation to read her paper on this topic to the Royal Statistical Society in London.

Klea Katsouyanni was born in Thessaloniki, Greece. She obtained her first degree in mathematics in 1976 from the University of Patras, Greece, an M.Sc. in statistics from Brunel University, United Kingdom in 1978, and a doctoral degree in medical sciences from the University of Athens Medical School in 1982. That same year she joined the Department of Hygiene and Epidemiology of the University of Athens Medical School as a lecturer and is now a professor in medical statistics and epidemiology in the same department. She has worked in environmental epidemiology for more than 20 years and is the Coordinator of the APHEA network, an European Commission-funded project established in 1992 that produced results contributing to the understanding of the air pollution health effects in Europe and the establishment of legislation for the management of air quality. She has been a member of several advisory committees, including for the European Commission and WHO, in environmental health topics.

Daniel Krewski has been professor in the Department of Epidemiology and Community Medicine in the Faculty of Medicine at the University of Ottawa since 1988, and Director of the R. Samuel McLaughlin Centre, where he holds the NSERC/SSHRC/McLaughlin Chair in population health risk assessment. He is also the scientific director of the PAHO/WHO Collaborating Centre in Population Health Risk Assessment at the University of Ottawa. His professional interests include epidemiology, biostatistics, risk assessment, and risk management. Dr. Krewski

obtained his Ph.D. in statistics from Carleton University and subsequently completed an MHA at the University of Ottawa. Dr. Krewski is a fellow of the American Statistical Association and the Society of Risk Analysis.

Alain Le Tertre is head of statistical and information systems of the Environmental Health Department at the National Institute of Public Health Surveillance (InVS) in France. He received his doctoral degree in biomathematics in 2005. His research interests include time-series analysis and mixed-effect models. Since 1992 he has worked on the short term effects of air pollution on health, first at the regional level and then moving to the national agency.

Sylvia Medina received her medical degree from Malaga University and her Ph.D. from Rene Descartes University in Paris. She has served as coordinator of European programs in the Department of Environmental Health at the French Institute for Public Health Surveillance since 2005, and as coordinator of the APHEIS European surveillance program on air pollution and health since 1999. Her research interests include air pollution related health effects, and she has participated in the French watch warning system on heat and cold waves since 2003.

Roger D. Peng is assistant professor of biostatistics at the Johns Hopkins Bloomberg School of Public Health. He received his Ph.D. in statistics from the University of California, Los Angeles. His research interests include statistical methods for environmental epidemiology, methods for spatial-temporal data, statistical computing, and software development.

Timothy Ramsay received a Ph.D. in statistics from Queen's University in Ontario and is currently an assistant professor in the University of Ottawa's Department of Epidemiology and Community Medicine, as well as a scientist with the clinical epidemiology unit of the Ottawa Health Research Institute. His primary interest focuses on methodology and biostatistics, particularly in the field of epidemiology. He spent five years with the McLaughlin Centre for Population Health Risk Assessment, where he held an NSERC/SSHRC/McLaughlin Junior Chair in quantitative risk assessment and conducted research on a variety of methodological issues affecting the science of risk assessment. He has been particularly active in the area of time-series methodology for assessing the health effects associated with short-term exposure to ambient air pollution, and discovered in 2002 that the standard implementation of the generalized additive model was overestimating the precision of estimated pollution effects.

Jonathan M. Samet is professor and Flora L. Thornton chair of the Department of Preventive Medicine, Keck School of Medicine, and director of the Institute for Global Health, University of Southern California. During the APHENA study he was professor and chairman of the Department of Epidemiology at the Johns Hopkins Bloomberg School of Public Health. He received his medical degree in 1970 from the University of Rochester School of Medicine and Dentistry and a master's degree in epidemiology from the Harvard School of Public Health. His research has focused on the effects of environmental and occupational agents. He was chair of the National Research Council's Board on Environmental Studies and Toxicology. He currently chairs the Clean Air Scientific Advisory Committee of the U.S. EPA.

Evangelia Samoli is a research scientist at the University of Athens. She received her doctoral degree from that institution in 2005. She is involved in teaching biostatistics to undergraduate and graduate students at the Medical School, University of Athens, and participates in related workshops. Her research interests include environmental epidemiology, biostatistics, and cancer epidemiology.

Joel Schwartz, professor of environmental epidemiology at Harvard School of Public Health and Harvard Medical School, received his doctorate from Brandeis University in theoretical physics and was given the John D. and Catherine T. MacArthur Fellowship award. He has held appointments as a visiting scientist at the University of Basel, Switzerland and at the University of Wupperal, Germany. He has also served as a senior scientist at the U.S. EPA. His research focuses on the effects of antioxidants on respiratory health, the health effects of air and water pollution and of lead, and methodological questions regarding the modeling of covariates in epidemiologic studies. For more than a decade, his studies have been instrumental in precipitating a renewed interest in airborne particles.

Giota Touloumi was born in Chalkis Evias, Greece. She obtained her first degree in mathematics in 1985 from the University of Ioannina, Greece, an M.Sc. in Statistics from the London School of Hygiene and Tropical Medicine, United Kingdom in 1992, a doctoral degree in medical sciences from the University of Athens Medical School in 1995, and a Ph.D. in medical statistics from the London School of Hygiene and Tropical Medicine, United Kingdom in 2001. In 1998 she joined the Department of Hygiene and Epidemiology of the University of Athens Medical School as a lecturer, and is now assistant professor in medical statistics in the same university. Dr. Touloumi has worked in medical statistics and environmental epidemiology for more than 10 years. She is a member of several international

statistical societies, including IBS, RSS, SM, and ISCB, and participates in the APHEA network, an E.C.-funded project established in 1992.

Antonella Zanobetti received her Ph.D. in 1999 in applied statistics from the University of Florence. She is currently a research scientist in the Exposure Epidemiology and Risk Program at Harvard School of Public Health. Her field of research includes statistical analysis of environmental epidemiology data, with particular interest in adverse effects of air pollution on health. Her expertise is analyzing the short-term effect of air pollution on mortality and morbidity by means of time-series analysis. Dr. Zanobetti's other research interests include mechanisms linking inhalation of ambient particles to acute exacerbation of cardiovascular or respiratory disease, health effects of air pollution and temperature extremes, and socioeconomic influences on health. She is also interested in developing innovative statistical methodologies in environmental epidemiology.

OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

Samoli E, Peng R, Ramsay T, Pipikou M, Touloumi G, Dominici F, Burnett R, Cohen A, Krewski D, Samet J, Katsouyanni K. 2008. Acute Effects of Ambient Particulate Matter on Mortality in Europe and North America: Results from the APHENA Study. Environ Health Perspect 116(11):1480–1486.

Peng RD, Dominici F, Louis T. 2006. Model choice in multi-site time-series studies of air pollution and mortality, J R Stat Soc [Ser A] 169:179–203.

Touloumi G, Samoli E, Pipikou M, Le Tertre A, Atkinson R, Katsouyanni K. 2006. Seasonal confounding in air pollution and health time-series studies: Effect on air pollution effect estimates. Stat Med 25:4164–4178.

ABBREVIATIONS AND OTHER TERMS

AIC	Akaike Information Criterion
APHEA1	Air Pollution and Health: A European Approach, Phase 1
APHEA2	Air Pollution and Health: A European Approach, Phase 2
APHENA	Air Pollution and Health: A European and North American Approach
CI	confidence interval
CIHI	Canadian Institute for Health Information
CO	carbon monoxide
COPD	chronic obstructive pulmonary disease
df	degrees of freedom
DL	distributed lag; the cumulative effects of lags 0, 1, and 2.
U.S. EPA	U.S. Environmental Protection Agency
FIPS code	Federal information processing standards codes
GAM	generalized additive models
GCV	generalized cross validation criterion
GLM	generalized linear model
ICD	International Classification of Diseases
iHAPSS	internet-based Health and Air Pollution Surveillance System
lag 0–1	average of lags 0 and 1
LOESS	locally-weighted scatterplot smoothing
MLE-Berkey	Berkey-maximum likelihood estimator
MSE	mean squared error
NAPS	National Air Pollution Surveillance
NCS	natural cubic spline
NMMAPS	National Morbidity, Mortality, and Air Pollution Study
NO_2	nitrogen dioxide
NS	natural spline
O_3	ozone
PACF	partial autocorrelation function; in tables and figures PACF indicates the degrees of freedom chosen by minimizing that function.
PM	particulate matter
PM_{10}	particulate matter $\leq$ 10 μm in aerodynamic diameter
PS	penalized spline
SAPALDIA	Swiss Study on Air Pollution and Lung Disease in Adults
SO_2	sulfur dioxide
SES	socioeconomic status
SS	smoothing spline
TLNISE	two-level normal independent sampling estimation
TSP	total suspended particles
WHO	World Health Organization

Research Report 142, *Air Pollution and Health: A European and North American Approach (APHENA)*, K. Katsouyanni and J. Samet et al.

INTRODUCTION

For nearly two decades, scientists seeking to understand the role that air pollution might play in population health effects have relied heavily on epidemiologic studies known as time-series studies. Such time-series studies use information on daily changes in air pollutant concentrations and daily counts of morbidity and mortality, initially at the level of a single city. However, the wide range of methods used to assemble and analyze data from individual cities has made the findings difficult to interpret and has led to progressive efforts to combine information across multiple cities and, ultimately, across geographic regions. The goal of these larger analyses has been to develop more reliable estimates of the potential effects of air pollution on human health, as well as to increase the ability to discriminate between health effects that may truly be related to air pollution and those that may be attributable to other factors.

The Air Pollution and Health: A European and North American Approach (APHENA*) project was a natural extension of these efforts. APHENA was a collaboration among investigators from Europe, the United States, and Canada, led by co-principal investigators Klea Katsouyanni, University of Athens and Jonathan Samet, then at the Johns Hopkins Bloomberg School of Public Health. It was launched in 2002 with joint funding from HEI and the European Commission. The APHENA investigators' basic goals were: 1) to develop standardized approaches to the analysis of time-series data at the city and country (regional) level, and 2) to apply these common approaches to each of the three study regions, and thereby to develop clearer insights into geographic differences in the effects of pollution on mortality and morbidity and into potential sources of or explanations for the variation observed.

The study by Drs. Katsouyanni and Samet, "Air Pollution and Health: A Combined European and North American Approach", began in June 2002. Total HEI expenditures were $617,000. The draft Investigators' Report from Katsouyanni, Samet, and colleagues was received for review in March 2007. A revised report, received in May 2008, was accepted for publication in October 2008. During the review process, the HEI Health Review Committee and the investigators had the opportunity to exchange comments and to clarify issues in both the Investigators' Report and in the Review Committee's Commentary.

* A list of abbreviations and other terms appears at the end of the Investigators' Report.

SCIENTIFIC BACKGROUND

In air pollution epidemiology, a commonly exploited dimension of exposure variability is short-term temporal variability. Application of time-series statistical techniques to the analysis of daily variations in mortality and air pollution concentrations has become particularly common. Relatively simple and inexpensive to conduct when the relevant data are available, time-series studies have been conducted in cities all around the world (Anderson 2009).

The results of several daily time-series studies were first reported in the early 1990s for individual cities or countries (Hatzakis et al. 1986; Derrienic et al. 1989; Fairlay 1990; Katsouyanni et al. 1990; Schwartz and Marcus 1990; Schwartz 1991; Dockery et al. 1992; Pope et al. 1992; Schwartz and Dockery 1992a,b; Schwartz 1993; Sunyer et al 1993; Touloumi et al. 1994). These studies used time-series data and Poisson regression models to estimate the association between daily changes in pollution and daily changes in mortality while controlling for other time-dependent covariates that were potential confounders. These studies found small, but statistically significant effects of air pollution on daily mortality rates even at relatively low pollutant concentrations. The original research was largely replicated (Samet et al. 1995), and comparable associations were observed in other cities with different climates, alternative modeling approaches for weather conditions, different pollution mixes, and different demographics (Pope and Kalkstein 1996; Samet et al. 1998; Pope 1999; Bell et al. 2004b).

Studies of short-term exposure have found associations between concentrations of airborne particulate matter (PM) and a large range of outcomes; these findings have been reviewed more extensively elsewhere (Pope 1999; Bell et al. 2004b, Pope and Dockery 2006, Anderson 2009). PM pollution has been associated with daily mortality (from all-causes [nonaccidental or all natural causes], respiratory causes, or cardiovascular causes), hospital admissions for respiratory diseases (from all-causes, chronic obstructive pulmonary disease, asthma, or pneumonia),

and hospital admissions for cardiovascular diseases (from acute myocardial infarction or congestive heart failure).

At the same time, numerous concerns were being voiced regarding sources of uncertainty in the findings of individual studies that might undermine the hypothesis of a cause and effect relationship between PM and mortality. One was that the magnitude of the relative-risk estimates from time-series studies of daily mortality depends on the approach used to model both the temporal pattern of exposure (Braga et al. 2001) and potential confounders that vary with time (such as season and weather) (HEI 2003). Other concerns included the role of uncontrolled confounding or effect modification by the pollutants not classified as PM (Moolgavkar et al. 1995; Gamble and Lewis 1996), variations in statistical approaches to modeling of time-series data in individual cities (Thurston and Kinney 1995), exposure measurement error (Lipfert and Wyzga 1997), mortality displacement or "harvesting" (Lipfert and Wyzga 1995; McMichael et al. 1998), and publication bias. When the APHENA study began, many of these issues were unresolved.

Investigators have attempted to address these concerns in a variety of ways. One of the most common approaches has been to conduct larger studies that rely on using uniform methods for assembling and analyzing data from multiple cities. Several multicity studies have now been conducted, but three in particular became the foundation for the APHENA study, representing different regions of the world as well as different methodological approaches to the multicity analyses.

One of the largest multicity daily time-series studies to be included was the National Morbidity, Mortality, and Air Pollution Study (NMMAPS). NMMAPS was getting under way as the APHENA project was being contemplated. This study began with efforts to replicate several early single-city time-series studies (Samet et al. 1995) and ultimately developed multistage hierarchical methods to assess the relationship between air pollution and mortality and morbidity in the 90 largest U.S. cities. NMMAPS Part I (Samet et al. 2000a) was designed to address methodological issues including uncertainties about exposure measurement error, mortality displacement, and the analysis of multisite data. NMMAPS Part II addressed questions about bias in selecting locations to study, differences in the statistical techniques applied, and adequacy of control for the effects of other pollutants on the associations between PM_{10} ($PM \leq 10$ μg/m^3 in aerodynamic diameter) and morbidity and mortality. Results of various analyses of these data have been reported (Dominici et al. 2000a,b; Samet et al. 2000b; Dominici et al. 2003; Peng et al. 2005). The PM-mortality effect estimates were somewhat sensitive to various modeling and city-selection choices. However, relatively small but statistically significant PM-mortality associations were consistently observed, and little evidence was found to attribute the PM-mortality effect to any of the copollutants studied (nitrogen dioxide [NO_2], carbon monoxide [CO], sulfur dioxide [SO_2], or ozone [O_3]). Subsequent analysis of the NMMAPS data did find associations of O_3 with mortality (Bell et al. 2004a).

The other major multicity project included in APHENA was from Europe, the Air Pollution and Health: A European Approach, or the APHEA project. It had been funded in 1993 by the European Commission 91–94 Environment Programme. The first phase of the project, APHEA1, relied on a database developed from 15 European cities representing 10 different countries with a range of social, climatic, and air quality conditions. The outcome data included daily counts of total and cause-specific deaths and hospital admissions. The second phase of the project APHEA2, expanded the database to include a total of 32 cities, representing 19 countries, and more extensive exposure data than in the earlier study. Various analyses conducted as part of the APHEA1 and APHEA2 projects examined the short-term PM-mortality effects and found that PM concentrations were significantly associated with daily mortality counts (Katsouyanni et al. 1996, 1997, 2001; Zmirou et al. 1998; Samoli et al. 2001; Aga et al. 1993; Samoli et al. 2003; Analitis et al 2006). Results on effects of other pollutants have also been reported (Gryparis et al. 2004; Samoli et al. 2006, 2007). Ultimately, APHEA's contribution to the APHENA project included data from 24 cities representing 15 countries.

Canada provided the third set of single and multicity studies for APHENA. During the 1990s, Canadian investigators also had conducted time-series analyses to examine air pollution related health effects in several of their largest cities (Burnett et al. 1998, 2000; Burnett and Goldberg 2003). Although a project similar to APHEA or NMMAPS had not been established in Canada, multicity studies involving up to 12 cities had been conducted to address some of the same issues as the European and U.S. projects. These multicity studies consequently became part of the multiregion effort developed in the APHENA study.

APHENA's later start allowed it to benefit from methodological refinements that occurred while the component studies were underway. Among them were improved assessments of the impact of measurement error and mortality displacement, as well as revised approaches to regression analyses. The improved regression methods were developed in response to the discovery of problems with the default convergence criteria used in the S-plus generalized additive model (GAM) function. The GAM

function had been used in the Poisson regression methods that had become the method of choice for analyzing time-series data at the end of the 1990s. Concern that the GAM problems could both bias effect estimates upward and underestimate standard errors caused NMMAPS and APHEA2 to undergo extensive reanalysis using different methods (HEI 2003).

In particular, the results of sensitivity analyses included in the HEI reanalysis (2003) suggested that the estimation of the pollution effect was sensitive to the method of smoothing used to control for the confounding effects of time and that the degrees of freedom (*df*) used for smoothing could be very influential. The HEI panel of scientists that reviewed the results from that report concluded that "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 the absence of adequate biological understanding of the time course of PM and weather effects and their interactions, the Panel recommends exploration of the sensitivity of future time-series studies to a wider range of alternative degrees of smoothing and to alternative specifications of weather variables (HEI 2003)." APHENA represents the only major multicountry time-series study to implement those recommendations extensively in its analytic approach.

Substantive questions also remained about the heterogeneity observed in mortality and morbidity effect estimates across individual and multicity studies, beyond that explained by methodological differences among the studies. Specifically, what are the potential roles of different pollutant sources, PM characteristics, copollutants, weather, and socioeconomic and other demographic factors in influencing pollutant effects? Answers to such questions are of particular concern to policy makers interested in more targeted, cost-effective ways to reduce the impacts of air pollution. By involving a broad range of countries and cities in Europe and in North America, APHENA held out the promise both of greater statistical power and of wider variation in possible explanatory factors with which to explore heterogeneity in relative rates of mortality and morbidity.

SUMMARY OF THE STUDY

SPECIFIC AIMS

The APHENA project had two overall goals that followed from this history of air pollution research. The first was the rigorous development of a common analytic protocol with which to characterize relative rates of mortality and hospital admissions associated with airborne PM_{10} and O_3 concentrations across Europe and North America. The second was to explore possible explanations for any remaining heterogeneity in the relative rates that differences in analytic approaches could no longer explain.

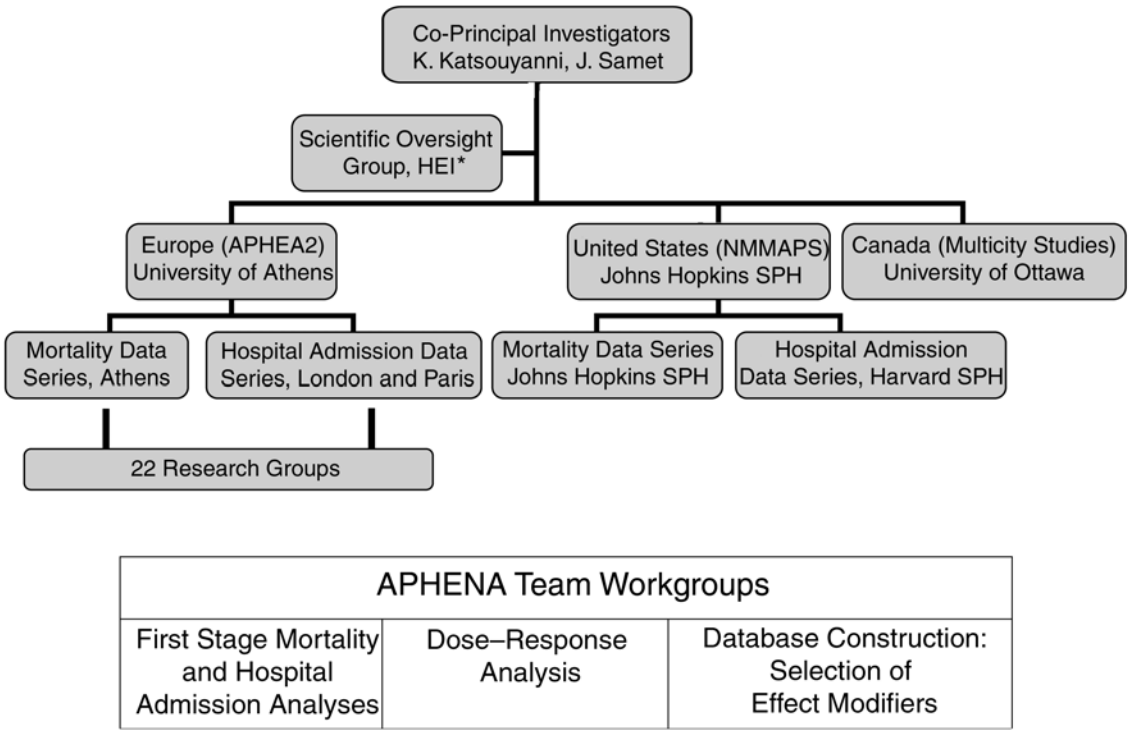
Development of a common analytic protocol required evaluation and selection of approaches to the first-stage analysis of individual city time-series data (i.e., city-level analyses) and to the second stage analysis in which results from individual cities were pooled within each region (i.e., regional-level analyses) and where possible, across regions. Specifically, the investigators' methodological goals were:

1. to develop a common approach or protocol for the first-stage or city-level analysis of mortality and hospital admissions time-series data, making use of assessment of the sensitivity of the findings to key elements of the model, and
2. to perform a comparative evaluation of the different approaches to the second-stage analyses used within the three regional projects.

The exploration of heterogeneity had two major components:

1. development of a database of variables common to the three regions that might act as effect modifiers (and thus sources of heterogeneity) and the exploration of differences in these core variables across the three regions,
2. use of the common protocols agreed upon, under the methodological development goals, to perform analyses of relationships between daily air pollutant concentrations and a) mortality and b) morbidity within and across regions, taking into account the variables identified in the previous step.

As part of their model development process, the APHENA team also conducted sensitivity analyses to explore additional modeling issues: 1) implications of using case-crossover analysis as an alternative approach to calculating effect estimates; 2) the shape of the concentration–response relationships between PM_{10} and O_3 and all-cause mortality at low concentrations — specifically whether or not threshold models were more consistent with the data; and 3) comparison of air pollution effect estimates using *systematically-missing* PM_{10} values with those estimated with daily PM_{10} data. Early plans to compare alternative methods for addressing the question of mortality displacement were dropped from the project.



Commentary Figure 1. Structure of the APHENA collaboration. (* Scientific Oversight Group members are listed at the end of this Commentary.)

STRUCTURE OF THE COLLABORATION

APHENA required a complex multicenter, multicountry effort involving teams of investigators from the United States, Europe, and Canada. The overall structure of the collaboration is shown in Commentary Figure 1. Working groups from the three centers were established and met at intervals to develop methods and to discuss interim results. However, the groups conducted work largely in parallel.

CONSTRUCTION OF THE DATA SETS

From its inception, the APHENA project was designed to build upon the databases originally developed for the APHEA, NMMAPS, and Canadian studies. The European contribution to APHENA was based on data collected for the APHEA2 project and comprised data sets for 24 cities (Katsouyanni et al. 2001). The NMMAPS database included data sets for 90 U.S. cities and is described in the HEI-funded Web site, internet-based Health and Air Pollution Surveillance System (iHAPSS) (*www.ihapss.jhsph.edu*) and in the NMMAPS reports (Samet et al. 2000a,b). The NMMAPS data used in the APHENA project reflect corrections to the database made after publication of the first NMMAPS studies. The Canadian data sets used in APHENA had been developed for a number of different analyses involving multiple cities and mortality or hospitalization outcomes. They were assembled by investigators from Health Canada for 12 cities selected for the availability of appropriate air pollution monitoring data.

The decision to rely on the preexisting times-series databases offered the advantages of lower costs and a more rapid start to the methodological work of the project. Also, the original published analyses of these datasets could serve as a baseline against which to explore the impact of the methodological choices made in APHENA.

However, it had the disadvantage that inherent differences in the ways air pollution or outcome data had been collected by government agencies in the various countries remained a potential source of uncertainty. Furthermore, restrictions placed by various government agencies on the use of the data further limited evaluation of the datasets because the APHENA investigators could not create a central repository for the time-series data.

Because differences did exist in air pollution and in outcome data, the Scientific Oversight Group was concerned about the potential contribution of these differences to heterogeneity in the results and thus to the interpretability of the findings. HEI conducted a quality assurance audit to review and verify the construction of the individual databases. The results of that audit are summarized in Commentary Table 1, which also provides an overview of the

Commentary Table 1. APHENA Dataset Comparison[a]

	Study Region (Research Center)[b]		
	United States (Johns Hopkins University)	Canada (University of Ottawa)	Europe (University of Athens)
Accuracy			
PM_{10} source	U.S. EPA designated reference or equivalent methods	U.S. EPA designated reference or equivalent methods	Different sources, by reporting research groups.
QA/QC	Standardized. U.S. EPA.	Standardized. Comparable to U.S. EPA.	Varied, depending on country submitting data – PM_{10}, PM_{13}, regression of BS or TSP, ratio of TSP
Analyses intercomparison	Contributed 2 cities to 6-city intercomparison	Contributed 2 cities to 6-city intercomparison	Contributed 2 cities to 6-city intercomparison
Data editing	All data visually inspected. Some ozone data at two sites removed – deemed unrepresentative. Edited during NMMAPS	No editing performed. However, period when values at two sites were doubled was identified and corrected.	Documented data validation during APHEA, including some data filling and editing of incorrect submissions
Precision			
Duplicate dataset generation	Dataset independently generated on new computer. Original NMMAPS dataset and APHENA dataset compared	APHENA dataset and datasets previously analyzed were requested separately – compared well	Unable to duplicate
Representativeness			
Data review	All data visually inspected	NR	Summary statistics reviewed
PM_{10} frequency	Every sixth day, though occasionally daily	Every sixth day, though occasionally daily	Daily
City partitioning	By county	By city, per census subdivisions	By city
Excessive mortality	Edited if > 3 sigma	Not edited	Not edited
Calculations	Trimmed means		
Completeness			
Missing data criteria	NR	NR	Fixed monitoring sites with > 25% missing air pollution data for whole study period not included in analyses. Standardized protocol for filling in missing values for selected monitoring stations.
Study period	1987 – 1996	1987 – 1996	Varied by city.
Data management			
Version control files	Yes	No	No

[a] Source: Adapted from Investigators' Report Appendix C: HEI Quality Assurance Statement.

[b] NR indicates not reported.

construction of the PM_{10} and mortality outcome data sets. As will be discussed later in the Methods section, the APHENA investigators also conducted sensitivity analyses in some cases to explore the impact of some of those differences on effect estimates prior to the second-stage or pooled analyses.

An important component of the APHENA team's effort was to establish a database of ecologic covariates with which to explore potential sources of heterogeneity in the air pollution effect estimates observed across cities in the second-stage analysis. The investigators focused in particular on potential effect modifiers based on prior analyses from the APHENA component studies. Those studies had provided indications that estimated effects of air pollution would vary with characteristics of the populations, the air pollutant concentrations, climatic variables, the geographic areas, and the cities themselves. The challenge was in finding measures or indicators of these characteristics that were common to and comparable across all three primary databases. Consequently, the APHENA investigators established a workgroup to identify the existing databases (e.g., World Bank, UN, AIRNET, EUROSTAT, etc.) for these population-level indicators and to assess both the limitations and the comparability of these databases across countries.

Each of the three centers was individually responsible for developing and maintaining the database for its own region. Each database included: 1) air pollution monitoring data (for PM and O_3, NO_2, CO, and SO_2); 2) health outcome data in the form of daily mortality and hospitalization counts; and 3) weather data including daily temperature and relative humidity; day-of-week, holiday, and (for Europe only) influenza indicators. The exception to this center-specific approach was the database of potential effect modifiers that was developed for the project, shared between centers, and used for all the second stage models.

The key differences among the databases are discussed below.

Air Pollution Monitoring Data

The APHENA investigation of the health effects of air pollutants focused on PM_{10} (24-hour average) and O_3 (1-hour maximum) concentrations. The project faced challenges arising from differences in air pollution monitoring protocols across the three regions during the period of the study, particularly for PM_{10}. These differences, based in different regulatory schemes for the various countries, affected the frequency with which data were collected, as well as the measurement methods employed.

For PM_{10}, European and North American studies differed first in the frequency of data collection. In Europe, daily PM measurements were available for all cities, whereas in the United States, measurements were available for PM_{10} only for 1 in 3 or, in most cases, 1 of every 6 days in the majority of cities. In Canada, measurements were available for 1 of every 6 days in all cities. These systematically missing data affected the design of the time-series analyses by limiting the numbers of days following an exposure (i.e., lags) and the numbers of cities for which the health effects could be explored across the three regions.

Although Europe offered more daily PM measurements, the monitoring methods varied from country to country. In contrast to Canada and the United States, no standardization of monitoring methods existed across Europe for the time span of the study. As detailed in the Investigators' Report (Methods — Description of the Databases), among the 24 cities contributing to APHENA from the APHEA project, PM was measured variously as PM_{10}, PM_{13}, total suspended particles, and black smoke. Measurements of PM_{13} and total suspended particles were converted to an estimate of PM_{10} using various approaches based on local studies. Results of sensitivity analyses conducted by the investigators suggested that the effect estimates using the hybrid approaches to PM_{10} measurement were only slightly lower than those estimated using direct measures of PM_{10}. A few differences also existed among the research centers in how the data were edited prior to incorporation into the databases (Commentary Table 1).

The methodologies for O_3 measurements were relatively more consistent across the three regions, although measurements were not consistently available for all cities. Daily, hourly O_3 measurements were available for the full year in Canadian cities, but for only subsets of the European and U.S. cities.

Outcome Data

The APHENA investigators sought to standardize the health outcome data analyzed across the three study centers by using broad disease categories defined by International Classification of Disease (ICD-9) codes.

The protocol specified collection of daily mortality data for the following disease categories: 1) all nonaccidental (all-cause; ICD-9 < 800), 2) cardiovascular (ICD-9 390–459), and 3) respiratory causes of death (ICD-9 460–519). The age groupings differed by disease category: for all-cause and cardiovascular categories, deaths were tallied for people < 75 years and ≥ 75 years; for respiratory categories, deaths were reported for all ages and for people ≥ 75 years.

Morbidity was characterized using hospital admissions data for people ≥ 65 years. The protocol called for collection of data on cardiovascular admissions (ICD-9 390–429, excluding stroke [ICD-9 430–438]) and for respiratory

Commentary Table 2. Number of Cities in the APHENA Data Sets[a]

Study Region	Total Cities	Mortality Database				Hospital Admissions Database
		PM_{10}		O_3		PM_{10} and O_3
		Number of Cities	Measurement Frequency / Type	Number of Cities	Measurement Frequency / Type	Daily[b]
Europe	31	12	Daily	23	Daily (all-year)	8
		10	Daily, PM_{10} estimated from other measures			
USA	90	75	1 of 6 days	54	Daily (all-year)	14
		15	Daily	36	Daily (summer only)	
Canada	12	12	1 of 6 days	12	Daily (all-year)	12

[a] PM_{10} – 24 hour average; O_3 – 1 hour maximum.

[b] Except Canada PM_{10} measurements, which were available only for 1 of every 6 days.

admissions (ICD-9 460–519). Visits to emergency rooms that did not result in hospitalization were to be excluded from admissions counts.

Potential Effect Modifiers

From a starting point of over 80 variables that had been used in one way or another by APHEA, NMMAPS, or Canadian investigators, the APHENA workgroup narrowed the list of common variables to those presented in the Investigators' Report Appendix E Tables E.1–E.3. The final lists differed slightly by pollutant and by age group analyzed, but generally included variables on pollution characteristics (e.g., mean NO_2, mean NO_2/PM_{10} ratio, mean O_3); climatic variables (mean temperature and humidity); and sociodemographic characteristics that included crude mortality rates, age structure (percentage of population ≥ 65 years and percentage of population ≥ 75 years), and percentage unemployed.

Final Data Sets

Finding cities in each region that had both the necessary health outcome and air pollution data proved to be another challenge for the APHENA investigators. Commentary Table 2 gives a breakdown of the numbers of cities that ultimately comprised data sets for different analyses and shows how they differed from the total number of cities originally associated with each research center. Development of the mortality data sets was least affected. For Europe, of the 24 cities provided by APHEA2, 22 cities were available for the PM analyses and 23 for the O_3 analyses. Of the 22 European cities with daily PM data, only 12 cities used direct measurements of PM_{10}; the other estimated PM_{10} from other types of PM measurements (e.g., black smoke or total suspended particles). Daily PM_{10} measurements were available for 15 cities in the United States, but for no cities in Canada. Smaller numbers of cities, particularly in Europe, had both the complete hospital admissions data and the PM and O_3 data required for the analytic protocols.

Time periods covered by the different databases also varied. The U.S. and Canadian data sets included data for the 10-year period from 1987 to 1996 for mortality data. The Canadian hospital admission data were also from that period. However, the United States hospital admissions data spanned 1985 to 1994. The time period covered by the European data sets for mortality and hospital admissions ranged overall from 1990 to 1997, but varied by country; time-series data covered at least three years and usually more than five.

METHODS

The development of methods to characterize the concentration–response relationships between PM_{10} or O_3, and mortality or morbidity followed the two-stage process consistent with other multicity studies. The leading issue was development of the first-stage, or city-level analytic models. These would provide the foundation for design of the second-stage analyses in which the city-level results

were pooled within and, to the extent possible, across the three regions. However, a distinguishing characteristic of APHENA relative to other multicity projects was the extensive analyses conducted to explore the sensitivity of effect estimates to model choices.

First-Stage: City-Level Analyses

In the development of the first-stage protocol, the APHENA work groups undertook extensive analysis of: 1) the choice of the smoothing method and basis functions to control for temporal confounding in the city-level models; 2) the approach for selecting the appropriate number of degrees of freedom for smoothing temporal confounders; and 3) the suite of variables to be included in the model. This detailed analysis reflected the investigators' growing appreciation of the trade-offs between control of temporal confounding to avoid overstating the risks of air pollution and the possibility of over-specifying the statistical model (that is, including too many parameters in the model), thereby potentially underestimating the risks.

Teams from the APHEA2 and NMMAPS projects first undertook simulation studies to explore several issues involved in the specification of models for the APHENA first-stage analysis. The APHEA2 group conducted a simulation study in which they generated Poisson PM_{10} and mortality time-series data from a fully parametric model (Touloumi et al. 2006). The series simulated different patterns for long-term and seasonal trends in mortality, weather conditions, and influenza epidemics. They tested the smoothing functions (locally-weighted scatterplot smoothing [LOESS], natural splines [NS], smoothing splines [SS], and penalized splines [PS]). They evaluated several criteria for choosing the degree of smoothing, including minimization of the absolute value of the sum of the partial autocorrelation coefficient (PACF), the Akaike information criterion (AIC), fixed degrees of freedom, and the generalized cross-validation criterion (results are reported in Appendix A.1 of the Investigators' Report).

The NMMAPS team ran a parallel set of simulations, generating data from nonparametric models under several scenarios reflecting varying levels of confounding. The performance of each modeling method was evaluated using the mean squared error of the estimated air pollution effect as the criterion. The results from this simulation exercise were reported in Peng and colleagues (2006).

On the basis of these and other extensive simulation studies, the investigators decided that neither one model (for all cities and conditions) nor several models (specific to each city) was optimal. They agreed instead upon a final protocol with a suite of models and model assumptions to be used in city-specific analyses of daily mortality and hospital admissions (summarized in the Investigators' Report Table 2 and described in detail in the report's Appendix B). The protocol included both NS and PS as smooth functions for trend, seasonality, and temperature control. The selected degrees of freedom were 3, 8, and 12 per year. In addition, for time series with complete daily data (that is without systematically missing data), the minimization of PACF was used to select the optimal number of degrees of freedom per year. Exposure lags of 1, average of lags 0 and 1, and distributed lags were evaluated depending on the availability of daily data. Other lag terms of 0 and 1 were included for temperature. Dummy variables for day-of-week and holidays were included. Relative humidity and dew point were not included in the model, nor were variables to control for influenza epidemics. Applied systematically to the cities in each of the three study regions, these models provided insight into the sensitivity of effect estimates to an array of model assumptions.

Second-Stage: Regional-Level Estimates

Given agreement on a common protocol for estimating city-level effect estimates, the next step in developing a consistent methodology required comparison of the approaches previously used by the different investigators to pool the estimates across cities and to explore heterogeneity in the effect estimates for specific regions. The NMMAPS and APHEA2 projects represented the two major approaches being used at the time to pool estimates. NMMAPS used Bayesian hierarchical regression models (Gelman et al. 2003), fitted by use of Monte Carlo Markov Chain methods, for a computation called two-level normal independent sampling estimation (TLNISE) (Everson and Morris 2000); APHEA2 used metaregression models fitted by maximizing the likelihood function (MLE-Berkey; Berkey et al. 1998). The HEI Scientific Oversight group had encouraged the investigators to choose one or the other of the two methods to maintain a consistent approach at both stages of the analysis.

The APHENA investigators compared the two statistical approaches using a simulation exercise in which they generated time-series data using a hierarchical model for different hypothetical *true values* of the parameters of the model, representing different levels of heterogeneity. They then re-estimated the model parameters using the two statistical approaches described earlier and compared their relative efficiency at estimating the *true* parameters using a calculation of the mean squared error. The simulation studies did not suggest a clear preference for one method over another, so the investigators continued to "use the two approaches interchangeably" (Katsouyanni and Samet 2008, in a letter to HEI responding to a question from the

HEI Review Committee on this issue). In other words, the investigators in each center continued to use their preferred method for pooling data at the regional level, although it was not always clear which methods underlie the results presented in the Investigators' Report, particularly in the comparisons across regions (for example, Investigators' Report Figures 9 through 18). However, the paper published by Samoli and colleagues (2008), from the APHENA study indicates that the APHEA metaregression method was used for the center-specific and overall estimates of risk.

Exploration of Effect Modification

Using the APHENA protocol for the first-stage analysis but their preferred methods for the second-stage analyses, the investigators in each of the regions evaluated potential effect modification by the few variables identified earlier that were common to all data sets. Modification of the PM_{10} effect on all-cause mortality was assessed for the all-age and ≥ 75 year groups in the cities in Europe and in the United States that had daily PM_{10} data; Canada, lacking daily data, was excluded from the analysis. For O_3, effect modification patterns were evaluated for all-cause mortality in all three regions using the same approach. The results presented in the Investigators Report were only for models using lag 0–1, PS, and 8 *df*/year and for variables that had displayed significant effect modification in four of the eight protocol models applied in at least one center. Thus, the effect modifiers selected for comparison differed somewhat between the PM_{10} and O_3 results.

The investigators also conducted a sensitivity analysis using O_3 data to evaluate the impact of the different regional-level pooling methods on the patterns of effect modification.

Further Analyses

Alternative Analysis Using a Case-Crossover Design

When APHENA began, case-crossover analysis was increasingly being evaluated as an alternative to Poisson time series for estimating the short-term effects of air pollution. Case-crossover analyses are appealing because of their similarity to case-control studies and more straightforward approach to dealing with seasonal confounding. However, the case-crossover design can have several potential disadvantages, including that it does not account for overdispersion and may underestimate the variance of an estimator in each city. It can also induce larger estimates for heterogeneity than do time-series methods. The APHENA investigators were asked to compare the results of this alternative design to those of the second-stage models discussed earlier.

The ultimate analysis conducted was not as comprehensive as originally intended. Schwartz was already using case-crossover methods to analyze the mortality and pollutant data for PM_{10} (Schwartz 2004a,b) and O_3 (Schwartz 2005) from 14 U.S. cities that were a part of the NMMAPS cities. The APHENA project extended this work to the study of PM_{10} in a subset of European cities that were part of the APHEA project (Athens, London, Paris, Madrid, Rome, and Stockholm). No case-crossover analyses were conducted in the Canadian cities as part of APHENA.

Threshold Analysis Most analyses of risk to health from air pollution assume that the form of concentration–response relationship is linear. However, this assumption could result in both biased air pollution effect estimates and greater heterogeneity among estimates if, in fact, threshold or other nonlinear relationships were to exist. Given the implications that thresholds or nonlinear concentration–response relationships could have for the estimated burden of disease from air pollution, and thus for the estimated benefits of air quality standards or pollution-control policies, policy makers have been keenly interested in this issue. Several investigators affiliated with the APHEA and NMMAPS projects had previously used various methods to explore the nature of the concentration–response relationship (Daniels et al. 2000; Schwartz and Zanobetti 2000; Dominici et al. 2002; Samoli et al. 2003, 2005; Daniels et al. 2004).

The APHENA investigators first conducted a simulation study to investigate whether it was statistically possible to identify a range of hypothetical thresholds in simulated time-series data. Simulated Poisson time-series data were generated using a no-threshold model and a range of assumptions about the underlying strength of the concentration relationships (log relative risk; 0.01, 0.005, 0.001, or 0.0005). Models incorporating a range of hypothetical thresholds (0, 5, 10,….75 $\mu g/m^3$) were then fit to the data (using methods previously applied in both APHEA and NMMAPS, see Daniels et al. 2004) and the fit was assessed using AIC.

They next used a similar approach to investigate the potential for thresholds in the concentration–response relationships between PM_{10} or O_3 and all-age, all-cause mortality in each of the three centers. For PM_{10}, they investigated the existence of thresholds in the 22 European and 15 U.S. cities with daily data. Since the Canadian cities lacked daily PM_{10} data, they used O_3 data for the analysis. Essentially they fit the models with hypothetical thresholds to the city-specific time-series data, and assessed the fit

using AIC. They then computed the mean deviance of the fit, or the AIC value, across all of the cities in each region, theorizing that the existence of a threshold might be indicated by a poor fit of a threshold model to the actual time-series data.

Exploration of the Effect of Systematically Missing Values on Air Pollution Effect Estimates They conducted a sensitivity analysis in four European cities (Athens, London, Milan, and Zurich) for which complete daily time-series data were available for PM_{10} and O_3 for a period of five or more years. For each city, they created parallel synthetic data series with measurements every sixth day by selecting one day in the complete series and excluding the next five, and so on throughout the series. By varying the day with which they started, they created several data sets from the original series. The investigators applied the APHENA analytic protocol to the newly created data series with systematically missing data and compared the resulting effect estimates to those originally estimated with the complete daily time-series data.

SUMMARY OF KEY RESULTS

Methodological Contribution

APHENA began at a time when methodological issues involving time-series analysis of air pollution health effects were receiving considerable analytic attention. The APHENA project investigators effectively built on this work to offer additional insights to the design of time-series analyses for multiple cities. They carefully considered the class of models to be used for the analysis, the approach for selecting the smoothers and the degrees of freedom for confounders, and the other key variables to include in the model. With respect to selecting the number of degrees of freedom, they investigated whether the same number should be used across all locations (cities, countries, regions) and whether this should be fixed a priori or should be estimated from the data.

An important finding was that no single method is adequate or should be preferred for selecting the underlying model and the degrees of freedom for confounder control. The APHENA investigators learned that most of the alternative methods performed similarly, although the sensitivity analyses showed subtle variations in estimates of health effects that were data-dependent. Effect estimates also showed some sensitivity to each element of model specification.

Overall, their extensive simulation results suggested that the type of smoothers selected was less important than the degree of smoothing represented by the degrees of freedom per year. Their sensitivity analyses clarified that effect estimates pooled over several cities tended to stabilize at greater degrees of freedom per year (typically greater than 6–8 *df*/year) whereas such stabilization was not observed for all city-specific results. This finding influenced the investigators' decision to apply a common set of models at the city level using a range of degrees of freedom per year before pooling the resulting effect estimates at the regional level. Results from extended sensitivity analyses using 2–22 *df*/year can be seen in Figures 5 through 8 of the Investigators' Report.

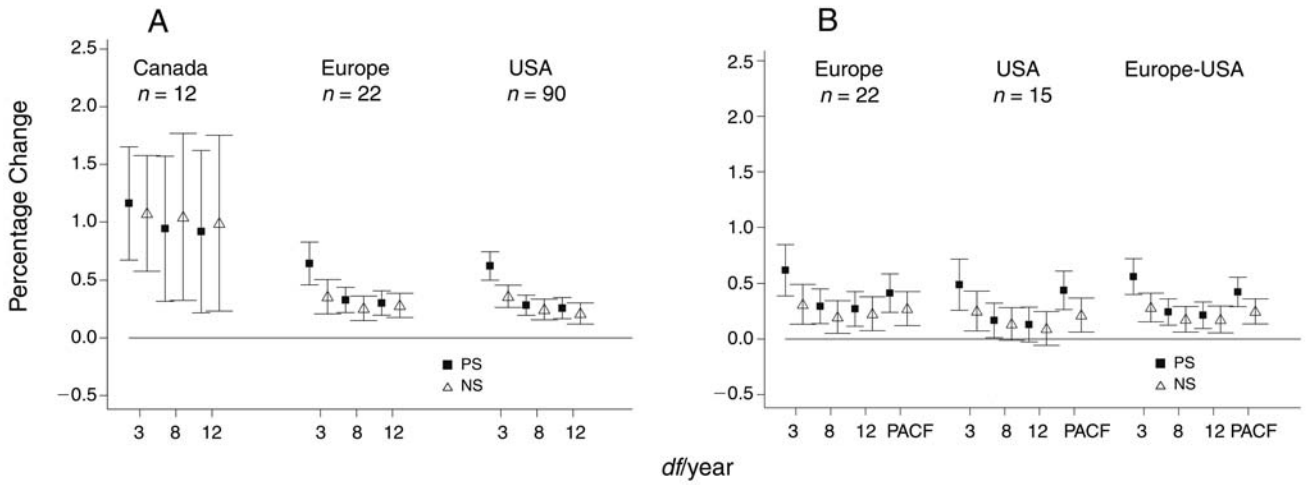
Given differences observed in the sensitivity of results for individual cities to different model specifications, the investigators decided to present results for a suite of models to portray the relationships between PM_{10} and O_3 and daily counts of mortality and morbidity.

Associations between PM_{10}, O_3, and Mortality and Morbidity

PM_{10} and Mortality The investigators reported the PM_{10} mortality results in an extensive series of tables for all-cause, cardiovascular, and respiratory mortality for each of the three regions individually (Investigators' Report Tables 4–6 for Canada, Tables 11–13 for Europe, and Tables 18–20 for the United States). Individual and pooled regional results for PM_{10} are displayed together in Figures 9–15 of the Investigators' Report. To illustrate the key findings, a subset of these results has been compiled in Commentary Figure 2 and Commentary Table 3.

The investigators reported positive and statistically significant associations of all-cause mortality for the all-age group with a 10-μg/m^3 increase in PM_{10} (lag 1) in Canada, Europe, and the United States for an array of model assumptions (Commentary Figure 2A). The pattern of results relative to choice of smoother (NS or PS) and degrees of freedom per year were also similar in the three regions. However, the effect estimates for Canada were two to three times higher than those for Europe and the United States.

The analysis of pooled effects of PM_{10} exposures using the average of lags 0 and 1 (lag 0–1), for cities with daily PM_{10} data, were also positively associated with all-cause mortality in Europe, the United States, and in both centers combined (Commentary Figure 2B). Effect estimates for the lag 0–1 exposures were lower than those for lag 1 in both regions, but especially in U.S. cities. It should be noted, however, that the lag 0–1 effect estimate was based on the subset of cities with daily measurements of PM_{10}, which greatly reduced the number of U.S. cities.



Commentary Figure 2. Percentage change in all-cause mortality, for all ages, associated with a 10-μg/m^3 PM_{10} increase. (Source: Investigators' Report Figure 9) **A:** lag 1, Canada, Europe, and United States; **B:** lag 0–1, Europe, United States, and the two centers combined.

Commentary Table 3. Percentage Increase in the Daily Number of Non-Accidental Deaths (All Ages, ≥75 and < 75 Years of Age) Associated with an Increase of 10 μg/m^3 in PM_{10} (Estimated by Using PS and 8 *df*/Year to Control for Seasonal Patterns) in the Three Centers[a]

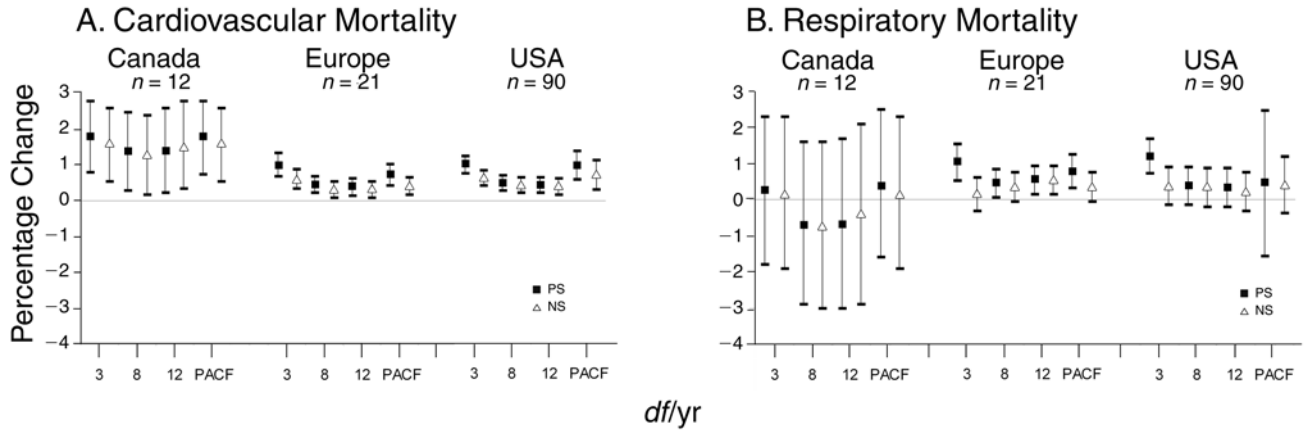
	Percentage Change in All-Cause Mortality[b] Mean (95% CI)			
Age Group (Years)/Center	Lag 1	Controlling for O_3 Lag 1	Average of Lags 0, 1	Distributed Lags 0,1,2
All ages				
Canada	0.86 (0.32 to 1.4)	0.74 (0.19 to 1.3)	NA	NA
Europe	0.33 (0.22 to 0.44)	0.32 (0.21 to 0.42)	0.29 (0.14 to 0.45)	0.20 (−0.01 to 0.42)
United States	0.29 (0.18 to 0.4)	0.24 (0.08 to 0.41)	0.14 (−0.12 to 0.4)	0.26 (−0.08 to 0.61)
≥75 years				
Canada	1.1 (0.35 to 1.9)	1 (0.25 to 1.8)	NA	NA
Europe	0.44 (0.29 to 0.58)	0.41 (0.27 to 0.54)	0.39 (0.19 to 0.59)	0.32 (0.04 to 0.60)
United States	0.47 (0.31 to 0.63)	0.37 (0.16 to 0.59)	0.19 (−0.19 to 0.56)	0.33 (−0.16 to 0.82)
<75 years				
Canada	0.58 (−0.16 to 1.3)	0.41 (−0.35 to 1.2)	NA	NA
Europe	0.25 (0.10 to 0.40)	0.23 (0.07 to 0.39)	0.25 (0.09 to 0.42)	0.11 (−0.20 to 0.43)
United States	0.12 (−0.02 to 0.27)	0.1 (−0.13 to 0.34)	0.09 (−0.2 to 0.38)	0.20 (−0.24 to 0.63)

[a] Compiled from Investigators' Report tables 4, 11, and 18.

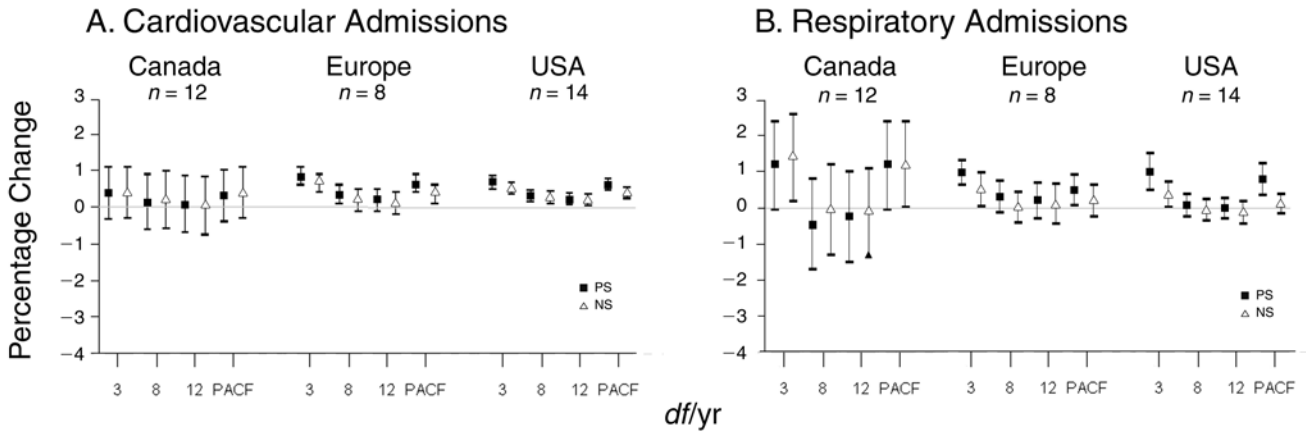
[b] NA indicates not applied because of systematically missing data for PM_{10}.

By presenting only results from models using PS and 8 *df*/year, Commentary Table 3 provides an overview of the influence of age, lag structure, and control for O_3 across the three regions. The percentage increase in mortality was greatest for persons 75 years or older in all three regions. For persons younger than 75 years, the effect estimates were positive, but were not statistically significant except in Europe. Increasing the number of lags, or days over which numbers of deaths were averaged, resulted in lower mortality effect estimates. Inclusion of O_3 in the lag 1 model reduced the PM_{10} effect estimates slightly but not significantly. (Note: This selection of results is similar to that presented in the Samoli and colleagues [2008] paper on APHENA.)

The investigators found positive and significant associations of PM_{10}, lag 1 with cardiovascular mortality, particularly in the oldest age group (Commentary Figure 3A).



Commentary Figure 3. Percentage change in mortality, for age ≥ 75, associated with a 10-μg/m^3 PM_{10} increase (lag 1). A: Cardiovascular (created from data in Investigators' Report Tables 5, 12, and 19); **B:** Respiratory (created from data in Investigators' Report Tables 6, 13, and 20).



Commentary Figure 4. Percentage change in hospital admissions, for age ≥ 65, associated with a 10-μg/m^3 PM_{10} increase (lag 1). A: Cardiovascular (created from data in Investigators' Report Tables 27, 32, and 37); **B:** Respiratory (created from data in Investigators' Report Tables 26, 31, and 36).

Effect estimates from the analysis of European and United States data using lag 0–1 for cities with daily data were also positive and significantly different from zero (Investigators' Report Figure 12B). The effect estimates for persons younger than 75 years, though generally positive, were not statistically significant — except when 3 *df*/year or degrees of freedom per year chosen using the PACF criterion were used in the models (Investigator's Report Figure 13).

The results comparing the effects of PM_{10} exposure (lag 1) on respiratory mortality were less consistent across region and model. The results for the older age group are shown in Commentary Figure 3B. Effect estimates for Canada varied substantially with the number of degrees of freedom but were largely negative and statistically insignificant. Those for Europe and the United States were positive, but statistically significant primarily in Europe. When the two regions were pooled (for lag 0–1 exposures), the effect estimates were statistically significant (Investigators' Report Figure 15B).

The investigators report that their analyses of the effects of PM_{10} exposure on daily all-cause mortality generally replicated the findings reported for previous studies done in the three regions (see Investigators' Report Table 44). The Canadian effect estimates continued to be substantially larger than those for Europe or the United States but the APHENA results were somewhat higher than previously published estimates from pooled analyses from the same group of cities. The Investigators' Report comparisons for the other categories of mortality suggest general consistency with findings from previous studies.

PM_{10} and Hospital Admissions Commentary Figure 4 displays the percentage change in hospital admissions for

cardiovascular disease (A) and respiratory disease (B) for people 65 years or older for a 10-μg/m^3 change in PM_{10}, lag 1. Although not presented together in the Investigators Report, they are juxtaposed to the mortality results here to facilitate comparisons.

The APHENA investigators found that the effect estimates for cardiovascular and respiratory admissions varied across the three centers. In Canada, daily increases in PM_{10} were not associated with significantly increased risks of hospitalization for cardiovascular disease. The risk of respiratory admissions was lower, more variable, and model dependent; it was positive and significant in models using 3 *df*/year (and PACF) but negative and nonsignificant for 8 and 12 *df*/year. In Europe, PM_{10} was associated with increased rates of hospital admissions for both respiratory and cardiovascular disease, although the quantitative estimates were sensitive to both the underlying model (NS or PS) and to the degrees of freedom. In the United States, the estimates for cardiovascular disease admissions were generally positive and statistically significant for all models. PM_{10} effects on respiratory admissions in the United States tended to be lower than those in Europe but showed similar patterns of sensitivity to model assumptions.

O_3 and Mortality As with the PM_{10} results, the investigators reported the O_3 mortality results in an extensive series of tables and figures for all-cause, cardiovascular, and respiratory mortality for each of the three regions. Their key findings have been summarized here in a series of figures and tables analogous to those just presented for PM_{10}.

Commentary Figure 5 summarizes results from the analysis of the effect of O_3 on all-cause mortality in the all-age group, based on annual data with lag 1 or lag 0–1.

Ozone generally had a positive association with all-cause mortality in each of the three regions and, although not shown in this figure, when results from all three regions were combined (see Investigators' Report Figure 16). The estimates for the United States were highly sensitive to the number of degrees of freedom per year in the model; consistent with the sensitivity analyses conducted during methods development, estimates were more stable with the higher degrees of freedom per year. As in the PM_{10} analyses, effects estimates from O_3 exposure were several times higher in Canada than in the other two regions.

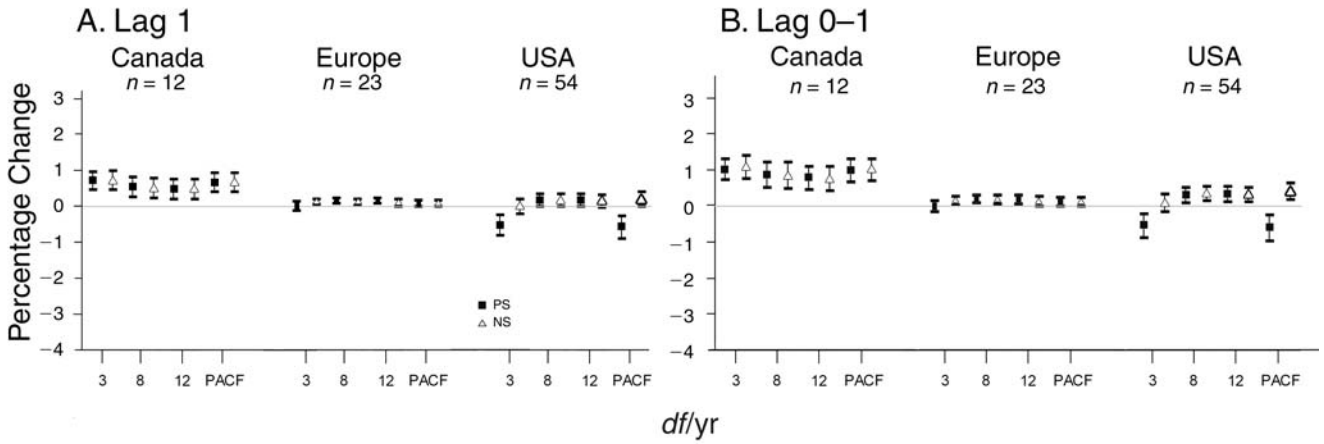
Commentary Table 4 provides insight into the influence of age, lag structure, and control for PM_{10} by comparing results based on a common model (PS, 8 *df*/year). As in the PM_{10} analyses, the effect of O_3 exposure on all-cause mortality was slightly higher in the oldest age group. Effect estimates were generally lower for lag 1 exposures than for lag 0–1 or distributed lags. When PM_{10} was included in the lag 1 models, the O_3 effect estimates generally declined, except in Europe where they were slightly increased.

Commentary Figure 6, summarizes the impact of a 10-μg/m^3, lag 0–1 increase in O_3 on cardiovascular and respiratory mortality in the three regions for people 75 years or older. The estimates of the effect of O_3 on cardiovascular mortality were positive and significant in Canada. In Europe and the United States the effect estimates for cardiovascular mortality were several times lower than in Canada, and significance was more model-dependent. The cardiovascular effect estimates were slightly higher than those for all-cause mortality in their respective regions, but not consistently so. The associations between O_3 and respiratory mortality, though still slightly higher in magnitude in Canada relative to those estimated for Europe and the United States, were generally close to zero and were not significant in any region or in the combined estimate from all three regions (Investigators' Report Figure 18).

For the summer-only analyses, the overall patterns describing the relative influence of O_3 on all-cause, cardiovascular, and respiratory mortality were similar, but the effect estimates were generally higher and more significant in all three regions than in the analyses of full-year data. Results in Canada differed from both the annual results and the other regions because the effect of O_3 on all-cause mortality was higher than for cardiovascular mortality alone, and the effect on respiratory mortality was twice that of all-cause mortality.

The investigators concluded that the O_3 mortality effect estimates from APHENA were generally consistent with those reported in earlier studies from the same data sets (Investigators' Report Table 44 for all-cause mortality, one pollutant models).

O_3 and Hospital Admissions Estimates of the effect of O_3 (lag 0–1) on hospital admissions for cardiovascular disease and respiratory disease are displayed in Commentary Figure 7. In contrast to the cardiovascular mortality results (Commentary Figure 6A), the estimates of the effect of O_3 on cardiovascular hospital admissions were closer to zero and not statistically significant — particularly for models with 8 and 12 *df*/year. Effect estimates were higher for respiratory than for cardiovascular disease admissions and both higher and more uncertain in Canada than in the other two regions. Model sensitivity varied across the data sets; the same set of assumptions (choice of NS or PS, *df*/year, and lags) did not result in the same pattern of results in each data set.



Commentary Figure 5. Percentage change in all-cause mortality, for all ages, associated with a 10-μg/m^3 O_3 increase (created from data in Investigators' Report Tables 7, 14, and 21). A: lag 1; **B:** lag 0–1.

Commentary Table 4. Percentage Increase in All-Cause Mortality (All Ages, ≥ 75 and < 75 Years of Age) Associated with an Increase of 10 μg/m^3 in O_3 (Estimated by Using 8 *df*/year to Control for Seasonal Patterns and PS) in the Three Centers[a]

	Percentage Change in Total Mortality Mean (95% CI)			
Age Group (Years)/ Center	Lag 1	Controlling for PM_{10} Lag 1	Average of Lags 0,1	Distributed Lags 0,1,2
All ages				
Canada	0.56 (0.48 to 0.83)	0.48 (−0.18 to 1.2)	0.85 (0.51 to 1.2)	0.75 (0.25 to 1.3)
Europe	0.17 (0.09 to 0.25)	0.19 (0.10 to 0.28)	0.18 (0.07 to 0.30)	0.25 (0.10 to 0.40)
United States	0.18 (0.00 to 0.35)	0.13 (−0.18 to 0.44)	0.31 (0.09 to 0.52)	0.43 (0.11 to 0.75)
≥75 years				
Canada	0.61 (0.22 to 1)	0.07 (−0.99 to 0.87)	0.98 (0.5 to 1.5)	0.9 (0.19 to 1.6)
Europe	0.14 (0.04 to 0.24)	0.16 (0.05 to 0.28)	0.12 (−0.02 to 0.26)	0.17 (−0.03 to 0.37)
United States	0.21 (−0.04 to 0.46)	−0.12 (−0.39 to 0.63)	0.33 (0.02 to 0.64)	0.4 (−0.05 to 0.85)
<75 years				
Canada	0.52 (0.14 to 0.89)	1 (0.11 to 2)	0.74 (0.28 to 1.2)	0.62 (−0.06 to 1.3)
Europe	0.18 (0.07 to 0.29)	0.24 (0.12 to 0.37)	0.25 (0.10 to 0.40)	0.37 (0.14 to 0.59)
United States	0.15 (−0.07 to 0.37)	0.12 (−0.33 to 0.56)	0.3 (0.04 to 0.56)	0.5 (0.09 to 0.92)

[a] Compiled from Investigators' Report tables 7, 14, and 21.

Effect Modification

The APHENA investigators reported that PM_{10} effect-modification patterns were generally consistent across statistical model used (i.e., choice of smoother and *df*/year), but that the evidence for effect modification was weaker when more degrees of freedom per year were used for control of seasonality. The results presented in the Investigators' Report were only for models using lag 0–1, PS, and 8 *df*/year.

Modification of the PM_{10} effect by copollutant and climatic variables was not entirely consistent across Europe and the United States (see Investigators' Report Table 41). For example, higher annual mean NO_2 concentrations and larger NO_2/PM_{10} ratios were associated with higher PM_{10} mortality effects in Europe and in the United States, but to a lesser extent. Higher temperatures and drier locations were associated with increased PM_{10} effects on mortality in Europe, but not in the United States.

The most consistent evidence of effect modification was found for age and unemployment rate in Europe and in the United States. A higher percentage of older people and a higher percentage of unemployment were each associated

with a greater effect of PM_{10} on mortality in both regions. However, in cities with higher O_3 concentrations, a smaller PM_{10} effect on mortality in the older age group was found, particularly in the United States.

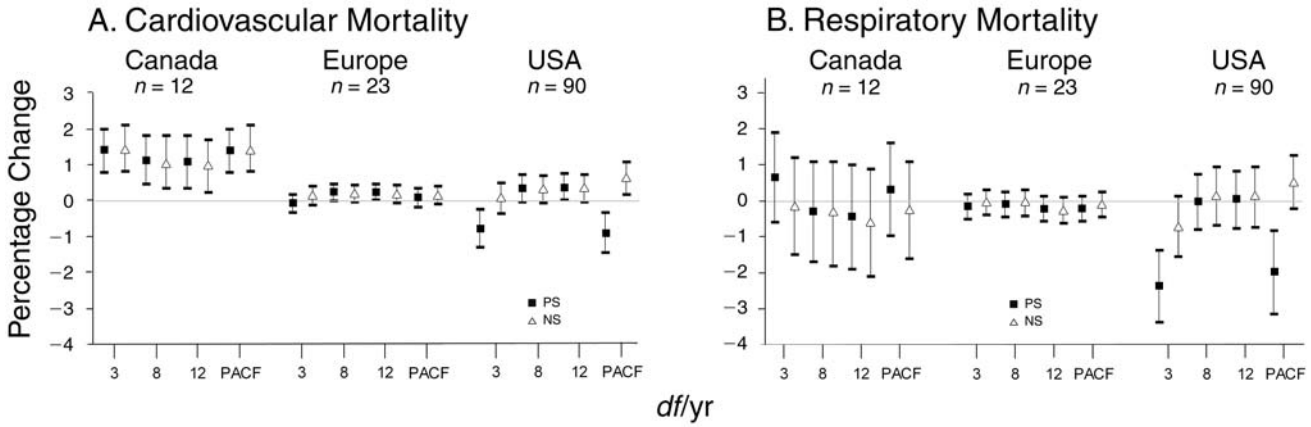
The investigators found no consistent patterns of effect modification for O_3 across the three regions, but suggested there was evidence of effect modification by some variables in the United States and Canada (Investigators' Report Table 42).

The comparison of the hierarchical and metaregression modeling approaches to the second-stage analysis of effect modification found the results to be consistent with one another (Investigators' Report Table 43).

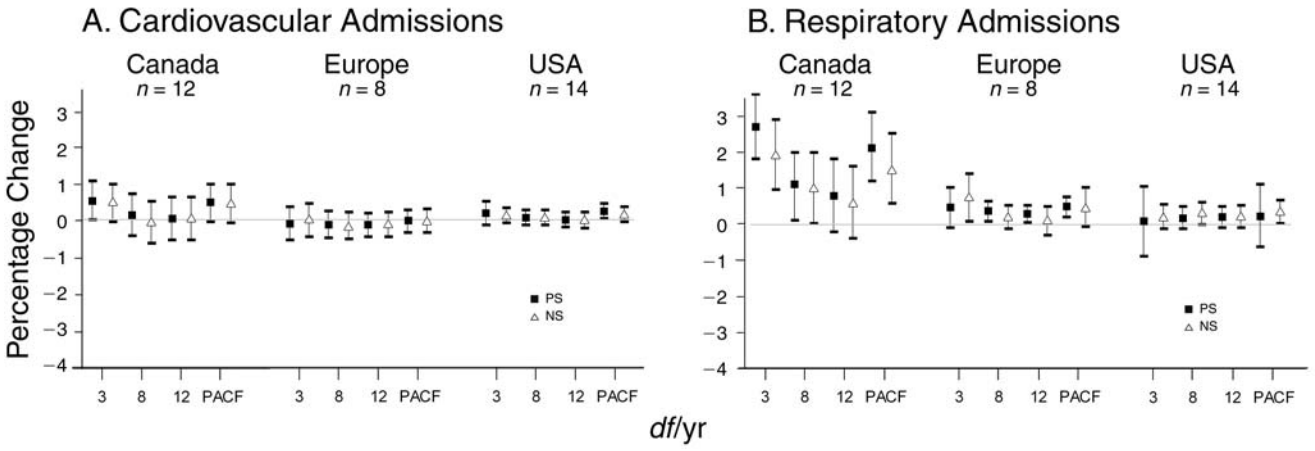
Results from Further Analyses

Case-Crossover Analyses The APHENA application of case-crossover analysis to the six selected cities from the APHEA project (Athens, London, Paris, Madrid, Rome, and Stockholm) found a 0.6% increase in all-cause mortality with a 10-μg/m^3 increase in PM_{10} in the pooled analysis. This effect estimate was higher than pooled estimates based on all the APHEA cities, but the authors point out that the original effect estimates in the individual were generally higher than in the APHEA cities overall.

Threshold Analysis The multicity results of the threshold analyses conducted by each the three centers also did not provide convincing evidence for a threshold. The Canadian investigators, who examined only O_3 effects, reported that the results of their pooled analysis of 12 Canadian cities appeared to suggest a threshold around 30 or 35 ppb, but that the differences in the fitting measure they used (the AIC) were not statistically significant. The Europeans conducted the threshold analysis for PM_{10} and mortality and for O_3 and mortality in all of their cities. They also reported that their results did not support the



Commentary Figure 6. Percentage change in mortality, for age ≥ 75, associated with a 10-μg/m^3 O_3 increase (lag 0–1). A: Cardiovascular (created from data in Investigators' Report Tables 8, 15, and 22); **B:** Respiratory (created from data in Investigators' Report Tables 9, 16, and 23).



Commentary Figure 7. Percentage change in hospital admissions, for age ≥ 65, associated with a 10-μg/m^3 O_3 increase (lag 0–1). A: Cardiovascular (created from data in Investigators' Report Tables 29, 34, and 39); **B:** Respiratory (created from data in Investigators' Report Tables 28, 33, and 38).

hypothesis of a threshold for either pollutant. The investigation of the 15 NMMAPS cities with daily PM_{10} data also did not find any evidence of a threshold in the United States.

These findings were not surprising given the investigators' threshold analyses using simulated time-series data; they had concluded that the power to detect a threshold would be limited, particularly if the true association between PM_{10} and mortality were relatively small, as is the case for many time-series studies. For the single-city simulations, there appeared to be little potential for discriminating among possible thresholds if the thresholds were to exist at relatively low PM_{10} concentrations.

Analysis of Systematically Missing Data The investigators compared the impact of systematically missing data on the mortality effect estimates for PM_{10} and O_3, using the same APHENA analytic protocol but only in four European cities (Athens, London, Milan, and Zurich).

For PM_{10}, the effects estimated for the series with missing data were smaller than those estimated with the full series, with the exception of Zurich, regardless of choice of model. The difference was more substantial for two cities (Athens and London). The pooled effect across cities was consistently smaller for all models (Investigators' Report Figures 26 and 28).

The results for O_3 showed a pattern similar to those for PM_{10}, although the effects estimated when using the series with missing data were consistently smaller for all cities, under all model assumptions. The difference was again most pronounced in Athens (Investigators' Report Figures 27 and 29).

While random variation was observed in each of the effect estimates, it was greater in the estimates based on the series with missing data, as would be expected given the more limited data.

The sensitivity analysis conducted by the investigators in Athens, in which they varied the starting day from which they began eliminating data from the complete PM_{10} series, suggested that the sequence of monitored days may be another source of variation in the estimates (Investigators' Report Figure 30). Three of the four series generated using alternative starting points resulted in similar degrees of reduction in the PM_{10} mortality effect estimates. One series resulted in substantially less reduction, and depending on the smoother used, even increased effect estimates.

HEI EVALUATION OF THE STUDY

The APHENA project was an ambitious collective effort designed to provide greater insight into methodological and scientific issues surrounding the estimation of the acute effects of air pollution. Given the active debates underway at the time over modeling choices and other explanations for observed variations in effect estimates across geographic areas, the objectives of the project were reasonable and laudable.

The project team was highly qualified to undertake this challenge. APHENA brought together a highly competent and experienced group of investigators from Europe, the United States, and Canada representing three major research efforts: APHEA, NMMAPS, and the Canadian multicity studies. The air pollution monitoring networks available to these investigators comprised the vast majority of the air pollution monitoring sites existing worldwide at the time.

Ultimately, the question is whether this multicity, multicountry analysis adds meaningfully to the findings of the three influential multicity analyses for these regions that were performed and published earlier. The HEI Review Committee, as discussed below, believes that the APHENA project has made a meaningful contribution, albeit inevitably constrained by limitations in data and other factors.

METHODOLOGY

The analytic problem the APHENA investigators addressed in their substantial methodological work was essentially one of model selection. The wide availability of large modern databases, and the ability to link them, enables scientists to assemble large collections of observations related to a problem of interest. Such large data sets then permit the estimation of quite large numbers of parameters, and widely available software can fit a wide variety of models. From a statistical standpoint, these models are meant to summarize the data, not to identify a mechanism by which pollution acts on mortality; that is, there is no way to tell, absent other supporting scientific data, whether any particular model is correct (for example, has correctly characterized the underlying biologic relationship).

First-Stage: City-Level Analyses

Currently, the most widely used models assume that the mortality or morbidity counts in a given city (or other small area) can be modeled using the Poisson distribution. The mean of the Poisson is modeled as a function of pollution, and also of potential confounding variables that may

both affect pollutant concentrations and be associated with mortality or morbidity. Variables related to weather (e.g., humidity or temperature) or time (e.g., season, day-of-week, holidays) are examples of confounding variables. This basic model is described by Equation 3 of the Investigators' Report.

Having made this model assumption, there are many details still to be specified in order to separate the effect of pollution on mortality or morbidity from those of potential confounding variables. The effects of time, temperature, and humidity are modeled with smooth functions, but there are several classes of smooth functions from which to choose. Each has a tuning parameter, usually described as degrees of freedom, that is used to describe how many parameters are estimated by the smooth function, or equivalently, how smooth the fitted function is. In one class, the smooth function is a sum of basis functions, which might, for example, be sine and cosine functions (but are now usually taken to be spline functions). In this class, the number of degrees of freedom is the number of such basis functions, as each one has a single parameter to be estimated. In the second class, the smooth function is specified with a very large number of basis functions, but the coefficients are *shrunk*, or regularized, in recognition of the fact that the smooth function is over-specified. In this class, the number of degrees of freedom is less than the total number of basis functions. Natural splines, with a fixed number of basis functions specified, are in the first class; smoothing splines are in the second class.

Experience with these time-series models has shown that the estimate of the pollution effect, which is of most interest, is affected by choices in both the type of smoothing method and the degrees of freedom tuning parameter, essentially because pollution and weather are correlated, and both have an impact on mortality. A common approach is to use some criterion of prediction accuracy, or more generally a measure of goodness-of-fit of the model. However, most of these methods are relatively ad hoc and their relevance for time-series studies of pollution and health are not guaranteed (i.e., model fit is not necessarily a guarantee that adequate control for confounding or other factors has been achieved).

An important related problem is that the estimation of the standard error of the estimated pollution effect is also more or less difficult depending on the smoothing method used; roughly, simpler smoothing methods lead to simpler methods for estimation of standard errors. Reliable estimates of the standard error are important for understanding the degree of uncertainty in the estimated effects of air pollution on health.

With that background, the APHENA investigators had twin goals of choosing both the type of smoothing, and the amount of smoothing (i.e., associated degrees of freedom) that seemed to provide the most suitable properties over a wide range of data types. The related question was whether to use a city-specific or uniformly applied degree of smoothing. They evaluated these options with extensive simulations, and in rather broad terms the amount of smoothing was assessed to be more important than the choice of smoothing method.

The APHENA investigators' finding that the amount of smoothing is more critical than the type of smoothing is important. It suggests that in research of this type a relatively simple choice of method would usually be appropriate, but that investigators should assess the sensitivity of their results by investigating several choices of the amount of smoothing. This is the approach that the APHENA investigators took in their common approach to the analysis of the three data sets.

Furthermore, the investigators' sensitivity analyses of the degree of smoothing revealed that results tended to stabilize at higher degrees of freedom in regional effect estimates (that is, estimates pooled over several cities) whereas such stabilization was not observed for all city-specific results. This finding is of interest because in many multicity studies these models have been separately fit to data from individual cities; in principle the method and amount of smoothing might be different for each city. This approach then makes it more difficult to compare and combine estimates across cities in the second stage of the analysis. APHENA's analysis suggests that it may also be a source of heterogeneity among the estimates.

However, these helpful conclusions must be tempered by the limitations of the data itself. Problems arise in any study relying on large administrative databases developed for purposes other than epidemiologic research (for example, data collected by governments or other agencies for administration of various regulatory programs, social policies, or services). These problems can include inconsistencies in data collection, differences in definitions of variables from one location to another, and the possible presence of large numbers of missing observations — among others. Potentially, they can have as much or more of an impact on effect estimates than do technical details of model selection. A reasonable guideline is to choose the simplest model that seems to capture the main variability in the data, and to explore in detail the sensitivity of the most scientifically relevant conclusions.

Second-Stage: Regional Analyses and Exploration of Effect Modification

The second major component of methods development in APHENA was to evaluate approaches to combining the city-specific results into overall summary results, while trying to find any underlying factors that influence differences (or heterogeneity) among cities or regions in the estimates of the health effects of air pollution. A simple way to assess heterogeneity among cities in a given country is to consider the ensemble of estimates of the effect of a given pollutant on a specific health outcome, along with their estimated standard errors (e.g., the effect of PM_{10} on mortality). These individual estimates are modeled as deviating from some overall (country-wide) average value, with some part of this deviation possibly explained by city-specific characteristics, and another part of the deviation ascribed to random (i.e., as yet unexplained) differences between cities. This concept is behind both the hierarchical models and the metaregression analysis that have been preferred by the NMMAPS and APHEA investigators, respectively, and that were compared in the APHENA project.

The HEI Review Committee found it reassuring that the results of the simulation studies indicated little reason to choose between the two methods. Both of these methods assume a multivariate normal distribution for the effect estimates, but differ slightly in the assumptions about the covariance matrix and in the implementation of the estimation.

Unfortunately, the Committee found that the information provided in the report was insufficient to independently assess whether one method should be preferred over the other. Details of the modeling can be found in some of the papers published from the APHENA work (Samoli et al. 2001; Peng et al. 2006; Touloumi et al. 2006); the equations presented in the report are not adequate for this purpose. The aggregation of the United States estimates into regional and country-wide estimates is described in more complete detail in NMMAPS II (Samet et al. 2000b, HEI 2003).

The original concept in APHENA, that the three regional estimates might then be combined into an overall estimate of effect — in essence a "third-stage" analysis — and could then be used to evaluate sources of heterogeneity across regions, ultimately made little sense. One obvious problem was that the estimates from the Canadian studies were an order of magnitude higher and were therefore difficult to combine to a common estimate. The HEI Review Committee suspects that there are too many differences in the databases for such an analysis to measure any quantity of scientific, and therefore of political, interest.

Regional Effect Estimates The HEI Review Committee agrees that the APHENA study has basically corroborated the estimates of the effects of PM_{10} and O_3 on mortality that were published in earlier analyses of the same data sets from Canada, Europe, and the United States (see Investigators' Report Table 44). However, the report would have benefited from a more concerted effort to digest the extensive series of results and to put them more carefully into the broader context of the time-series findings for these two pollutants. Investigator's Report Table 44 and related discussion in the report is quite limited in that regard.

Nonetheless, the APHENA investigators' findings of a small but significant effect of PM_{10} and ozone on daily mortality — based on a common analytic approach to city-level analyses between air pollution and health effects — represents an important improvement over those estimated using meta-analytic approaches that rely on published results. They showed that the effects of air pollution could not be attributed to model choice alone, and demonstrated the importance of a more transparent and complete presentation of model assumptions and results.

Coherence Between Mortality and Hospital Findings

The APHENA report would have benefited from a careful discussion of coherence between the mortality and hospital admissions results in the three regions and the implications of their findings.

Coherence in the setting of air pollution health effects research has come to have a variety of meanings (Bates 1992), including, in the context of observational studies, the extent to which hospitalization findings are consistent with findings on mortality. This aspect of coherence is motivated by the not unreasonable notion that if people are dying from exposure to air pollution, those that are less adversely affected by the same exposure might instead be hospitalized.

A general impediment to the assessment of coherence using mortality and hospitalization data is that the classifications of cause-of-death and reason-for-hospitalization are qualitatively different. Cause-of-death in time-series mortality studies is typically defined as the underlying cause of death that is abstracted from a death certificate. The underlying cause of death is often not the same as the immediate cause of death. Similarly, cause-of-hospitalization, derived from a discharge diagnosis in a hospital record, may not correspond to the underlying cause of the death, but rather to the immediate cause of the death. Therefore, lack of correspondence between the cause of death and the reason for hospitalization may not indicate a lack of coherence in findings from mortality and hospitalization time-series studies, although that is one possible explanation.

In APHENA this general problem was compounded by the lack of comparable time series of mortality and hospitalizations, with respect to both the length of the respective time series and the specific cities included. This in part arose because the existing city–hospitalization time series with which the investigators had to work were relatively limited in comparison with the city–mortality time series. In NMMAPS, the cities originally selected for the hospitalization time-series analyses had daily PM_{10} data, but several NMMAPS cities included in the mortality analyses that had daily PM_{10} data were not included in the hospitalization analyses; some cities included in the hospitalization analyses were not even part of the NMMAPS mortality analyses. These limitations preclude a more thorough and quantitative evaluation of the coherence of pollutant effects on hospitalization and mortality.

In spite of these issues, there seems to be reasonable coherence to the PM_{10} mortality and hospitalization effects, except in Canada. On the other hand, it is remarkable how little coherence there is for the O_3 effects. In all three study regions, O_3 was associated with increased cardiovascular mortality, but not with increased respiratory mortality. However, also in all three regions, O_3 was associated with increases in respiratory hospitalizations, but not with increases in cardiovascular hospitalizations. These findings do not "complement one another" as the investigators suggest, but on the face of it appear somewhat paradoxical. It is not known to what extent such apparent lack of coherence is due to differential levels of hospitalization for cardiovascular versus respiratory events, to expected differences in mortality and hospitalization effects resulting from the different interpretations of cause-of-hospitalization and cause-of-death, as discussed earlier, or to the markedly different sources of the hospitalization and mortality data in APHENA. To the extent that such findings support a lack of coherence, plausibility of the mortality or hospitalization findings might be questioned.

Effect Modification

One of primary goals of the APHENA project was to take advantage of the broader range of geographic areas, with their potential differences in pollution and population characteristics, to explore possible sources of heterogeneity in health effect estimates. Unfortunately, limitations in the underlying databases for APHENA restricted the extent to which interesting new hypotheses could be tested. The investigators do identify some potential factors, including unemployment rate, mean pollutant concentrations, and age structure of the population, that would quite plausibly influence the estimated effect of pollution on health outcomes.

The differences identified for PM mortality effect modifiers between the United States and Europe are interesting. For the United States, there was increased PM-related mortality with higher proportion of older persons in the population, higher mortality rate, higher unemployment, and higher NO_2 and humidity. For Europe, increased PM mortality was seen with higher unemployment, higher proportion of older persons, higher NO_2, and higher temperature, but not with higher humidity. While several interpretations could be suggested to explain these different findings, one interpretation is that the crude ecologic variables used in the study identified something different in Europe than in the United States. Unemployment was the one consistent effect modifier, but no other ecologic or sociodemographic variables were measured with enough consistency across the European cities to allow them to be used in the second stage of the analysis (the effect modification stage). As described in the methodology section, modification of the PM_{10} effect was only assessed using cities with daily PM_{10} data, limiting the strength of the analysis. The authors have explained that using the daily data (lag 0–1) results provided more heterogeneous results from which they observed more stable effect modification patterns, but the comparison data were not shown.

No consistent patterns of effect modification were found for O_3 across regions, although there was a suggestion of larger O_3 effects with higher unemployment in the United States and Canada.

While methods for pooling hospitalization data for assessment of effect modification were presented, no findings on effect modification of the PM or O_3 effects on hospitalizations was presented. It is possible that the investigators elected not to proceed with this aspect of the analysis because of the relatively small number of cities that contributed hospitalization time-series data.

Ultimately, these results are another reflection of the fact that large databases collected for administrative purposes will not usually be well suited for detailed scientific study of specific issues. They could not, in particular, explain the persistent and puzzling large differences between the air pollution effect estimates in Canada and those in the United States and Europe. They do, however, give us a broad picture, in this case identifying some potential factors, including unemployment rate, mean pollutant concentrations, and age structure of the population, that would quite plausibly influence the estimated effect of pollution on health outcomes.

Further Analyses

Case-Crossover Analyses The HEI Review Committee felt the relatively restricted analyses, ultimately conducted for only four European Cities, provided little further insight to the use of case-crossover analysis for multicity analyses.

Thresholds The APHENA analysis of the issue of thresholds did not substantially add to the earlier work conducted in NMMAPS III by Daniels and colleagues (2004). The APHENA investigators essentially applied the earlier methodology to data from Canada, APHEA, and NMMAPS; consistent with these earlier analyses, they found no evidence of a threshold.

The simulation study carried out by the APHENA investigators was a useful exercise, in which threshold models were fit to Poisson distributed time-series data that were generated under different assumptions about the strength of the underlying association. The study illustrated that it is difficult to detect a threshold in data when the strength of the association is small, as it is in many time series.

However, it is important to be careful when interpreting evidence from such analyses. The simulation data are also generated from a linear model that assumes the absence of a threshold. Thus, failure to find the threshold effect in simulations is of some comfort, as it indicates that when no such threshold exists, the methods used are not likely to mistakenly find one. However, it does not give any information on whether the methods are powerful enough to find any threshold that does exist in the data.

Ultimately, detection of thresholds, or change points, in a linear regression remains extremely difficult because it depends on a small subset of the data near the threshold. While quite sophisticated methods have been developed in the theoretical literature, their application to the available air pollution time-series databases would be very difficult. (For a more in-depth discussion of these issues, see the report and HEI Commentary on the NMMAPS III analyses [Daniels et al. 2004]).

Systematically Missing Data The APHENA investigators assessed the effect of using every-6th-day exposure data by comparing effect estimates from a daily time series to those in which 5 out of every 6 days of data were removed, artificially producing a data set with only every-6th-day data. The interesting, and potentially very important, finding from this exercise was that mortality effects estimated from the series with systematically missing data were generally lower and more uncertain than those estimated from daily data. Although this finding is based on a small subset of cities, the HEI Review Committee felt it could have important implications for interpreting effect estimates from NMMAPS and Canadian time series that have systematically missing pollutant data. It is possible that this factor could contribute to the heterogeneity in effect estimates across the study regions in APHENA. It is somewhat surprising that other investigators have not examined the impact of systematically missing pollution data in time-series studies more fully, although the number of cities with daily PM data necessary to carry out such analyses is still relatively small.

LIMITATIONS

The primary motivation for carrying out an integrated and standardized analysis of the Canadian, United States, and European multicity databases was to attempt to gain insight into between-region differences in study findings by eliminating or reducing between-region differences in study methods. The investigators were successful in reducing many of these methodological differences, particularly in the statistical methods, and thereby either completely or largely removed them from the list of possible explanations for variation in study findings.

However, there were many challenges to carrying out a fully integrated and standardized analysis of the three regions beyond statistical methodology. Those challenges that were effectively beyond the control of the investigators stemmed largely from a decision HEI and the investigators made at the outset; they decided to conduct a relatively cost-effective analysis — one that would rely on existing databases. Though reasonable, this decision limited the degree to which time-series data for the different cities and regions could be better aligned and impacts the ability to fully interpret the study findings.

There was an unavoidable lack of standardization in measuring or estimating PM_{10} concentrations in Europe. PM_{10} concentration data, either measured or estimated, was available in only 24 of the 31 cities, and in only 12 of these cities was mass actually measured (PM_{10} or PM_{13} [in two European cities]). In 10 European cities, measurement of black smoke or of total suspended particles was used as an estimate of PM_{10}. Methods used to estimate PM_{10} concentrations from surrogate data also differed across the European cities. It was somewhat reassuring that, in spite of these sources of variability, the investigators observed that the health findings were not meaningfully different for analyses using only the 12 cities with PM_{10} or PM_{13} measurements than for analyses based on the full suite of cities.

A related complication is that some time series did not include daily 24-hour concentration averages, most notably in Canada and in the U.S. cities, where most cities

contributed PM_{10} data only on every 6th day. This systematically missing data effectively shortened the time series and prevented a thorough investigation of combined-day lags. In APHENA, when analysis was restricted to the combined effects of day lags 0 and 1, only data from some of the APHEA cities and those few NMMAPS cities with daily data could be included. Also, the investigators chose to restrict their assessment of effect modification to cities with daily air pollution data, effectively excluding the Canadian cities and all but 15 of the 90 NMMAPS cities from this analysis. The implications of restricting the analysis to this more limited set of cities are unclear, but remain a concern.

The investigators did fully acknowledge one of the most unavoidable and obvious aspects of poor standardization — the varied sources of information used to characterize sociodemographic features of the populations of individual cities. This problem was most notable across the many European countries in APHEA2 where definitions of data terms and content, and likely, data quality varied. The immediate effect of these disparities on the APHENA investigation was that a large number of population characteristics of interest (i.e., potential effect modifiers) that might have helped account for heterogeneity in effect estimates across the three main study regions, could not be examined. The investigators were left with a very restricted set of potential effect modifiers that could reasonably be assumed to have common meanings across all of the cities. Specifically, in addition to readily measured features, such as age, only city unemployment rate (as a crude indicator of socioeconomic status) could be utilized with some confidence. (Although, even an unemployment figure could be influenced in different countries and regions by varying social welfare policies. For example, Canada and many European countries had much longer and more generous unemployment benefits than did the United States.)

For some European cities, it was not absolutely clear whether or not emergency room visits that did not lead to a hospital admission were included in the hospitalization time series. Allusion to emergency visits being "excluded in Barcelona" suggests that these might have been included elsewhere. If so, interpretation of hospitalization effects would become more complicated. While this may not be a significant issue, lack of certainty in this regard reduces confidence in the meaningfulness of cross-region comparisons of the hospitalization effects.

An additional hurdle to better integration and standardization was that a centralized data management center for the study could not be established. Again, this limitation was unavoidable given restrictions placed by various government agencies on the use of their data. The implications for APHENA were that quality assurance and quality control procedures for study data and procedures were therefore not standardized and may have resulted in differences in quality across the study sites.

One of the challenges that might have been addressed, but was not, was the lack of harmonization of many of the mortality and hospitalization time series with respect to cities, time period, and duration. Specifically, the time series from European cities ranged from 3 to 7 years versus 10 years in both Canada and NMMAPS. These disparities are in large part a function of the decision to use existing data. They may also reflect investigator preference for statistical power over standardization. Statistical power in time-series studies such as these is a direct function of the number of counts per day (deaths or hospitalizations) and the length of the time series. Effect estimates are also sensitive to the length of the time series. They are more variable in shorter time series, especially those as short as three years, and when the number of counts per day is not large, as in the smaller cities.

In addition, investigators from each region carried out the city-level analyses separately, albeit with guidelines as to the analytic methods to be used and some interaction between the investigating teams in the planning and execution of the analyses. Nevertheless, separately carrying out the analyses introduced a largely avoidable element of uncertainty. It was reassuring to see that results from a round-robin approach to some of the analyses were essentially indistinguishable.

From the information provided in the APHENA study, it is not possible to determine the extent to which these inconsistencies in the many data sets contributed individually or collectively to the heterogeneity in the health effect estimates. This uncertainty to some degree undermines the attempt to attribute heterogeneity to those factors that were well measured, and generally complicates interpretation of the findings on effect modification.

SUMMARY AND CONCLUSIONS

APHENA was an ambitious project undertaken by a highly qualified team of investigators from Europe, the United States, and Canada. With access to data from three geographic areas, it offered the opportunity of a much larger dataset with which to address methodological as well as scientific issues about the relationships between PM_{10}, O_3, and mortality and morbidity that were the subject of lively debates at the time the project was first conceived in 1999.

By the time the project was fully underway in 2003, some of the specific methodological issues, for example the implications of the convergence problems in the GAM S-plus software, were essentially resolved. However, the APHENA project still made important methodological contributions that continue to resonate through ongoing reliance on papers that were published from the APHENA work (Peng et al. 2006; Touloumi et al. 2006; Samoli et al. 2008).

The investigators conducted extensive sensitivity analyses at the city and regional levels to compare different modeling approaches favored by the three research centers and to evaluate the sensitivity of PM_{10} and O_3 effect estimates to a wide range of model specifications. They found that effect estimates remained fairly stable across a broad spectrum of model assumptions, but illustrated the importance of presenting a more complete array of results. Their approach allows a more transparent examination of the impact of model assumptions than does the selection and presentation of the results from only one model.

APHENA's use of a common analytic approach to city-level analyses of the relationships between air pollution and health effects was an important advance over other meta-analytic approaches that rely on published results. The investigators were successful in reducing the impact of methodological differences as an explanation for variations in study findings. Their corroboration of earlier findings of small, but significant effects of PM_{10} and O_3 on daily mortality, and to a lesser extent on hospital admissions, showed that the effects of air pollution could not be attributed to the vagaries of model choice.

The hope that the common analytic strategy might help reveal some of the other potential contributors to variations observed within and between the three regions was largely unfulfilled. Few new insights were possible given the limited number of potential effect modifiers that were common to the databases for the three regions, and also given the restriction of the analyses to the smaller number of cities with daily PM_{10} and O_3 data within regions. As a result, some of the more puzzling differences between regions therefore remained unexplained — in particular the much higher effect estimates for PM_{10} and O_3 in Canada relative to Europe and the United States.

The APHENA study demonstrated the substantial challenges that face efforts to standardize and integrate data from different countries. Overcoming government agency reluctance or other impediments to establishing centralized databases might help, but some challenges can remain beyond the control of investigators. Basic underlying disparities in the existing databases with respect to air pollutant monitoring methods and frequency, mortality and hospitalization records, and sociodemographic data are very difficult to fix retrospectively.

The authors suggest that periodic pooling of data, as in APHENA, should be considered both to explore methodological questions and to assess the progress of air pollution controls in reducing health impacts. The HEI Review Committee believes this recommendation should be evaluated cautiously. Studies like APHENA that use a well-reasoned, common analytic strategy, may offer the best approach for comparing and combining data across regions or countries. However, APHENA also illustrated how the limitations of using existing data sets can impact the ability to make clear comparisons and to explain the health effects of air pollution — sometimes just as much the technical details of model selection. For these limitations to be overcome in any future collaboration across international boundaries, thought needs to be given to substantially greater coordination and harmonization of the air pollution monitoring, health outcome, and covariate data collected in different countries. The costs of undertaking such exercises would need to be weighed against the expected advances in our understanding of air pollution and health effects.

ACKNOWLEDGMENTS

The Health Review Committee thanks the ad hoc reviewers for their help in evaluating the scientific merit of the Investigators' Report. The Committee is also grateful to Aaron Cohen for his oversight of the study, to Katherine Walker and Kate Adams for reviewing the edited report and preparing its Commentary, to Carol A. Moyer for science editing and managing the publication process, to Ruth Shaw for its composition, and to Fred Howe, Bernard Jacobson, Flannery C. McDermott, Melissa Norcross, and Quentin Sullivan for their roles in preparing this Research Report for publication.

REFERENCES

Aga E, Samoli E, Touloumi G, Anderson HR, Cadum E, Forsberg B, Goodman P, Goren A, Kotesovec F, Kriz B, Macarok-Hiti M, Medina S, Paldy A, Schindler C, Sunyer J, Tittanen P, Wojtyniak B, Zmirou D, Schwartz J, Katsouyanni K. 2003. Short-term effects of ambient particles on mortality in the elderly: Results from 28 cities in the APHEA2 project. Eur Respir J Suppl 40:28s–33s.

Analitis A, Katsouyanni K, Dimakopoulou K, Samoli E, Nikoloulopoulos AK, Petasakis Y, Touloumi G, Schwartz J, Anderson HR, Cambra K, Forastiere F, Zmirou D, Vonk JM,

Clancy L, Kriz B, Bobvos J, Pekkanen J. 2006. Short-term effects of ambient particles on cardiovascular and respiratory mortality. Epidemiology 17:230–233.

Anderson HR. 2009. Air pollution and mortality: A history. Atmos Environ 43:142-152.

Bates DV. 1992. Health indices of the adverse effects of air pollution: the question of coherence. Environ Res 59:336–349.

Bell ML, McDermott A, Zeger SL, Samet JM, Dominici F. 2004a. Ozone and short-term mortality in 95 US urban communities, 1987–2000. JAMA 292:2372–2378.

Bell ML, Samet JM, Dominici F. 2004b. Time-series studies of particulate matter. Annu Rev Public Health 25:247–280.

Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. 1998. Meta-analysis of multiple outcomes by regression with random effects. Stat Med 17:2537–2550.

Braga AL, Zanobetti A, Schwartz J. 2001. The lag structure between particulate air pollution and respiratory and cardiovascular deaths in 10 US cities. J Occup Environ Med 43:927–933.

Burnett RT, Brook J, Dann T, Delocla C, Philips O, Cakmak S, Vincent R, Goldberg MS, Krewski D. 2000. Association between particulate- and gas-phase components of urban air pollution and daily mortality in eight Canadian cities. Inhal Toxicol 4:15–39.

Burnett RT, Cakmak S, Brook JR. 1998. The effect of the urban ambient air pollution mix on daily mortality rates in 11 Canadian cities. Can J Public Health 89:152–156.

Burnett RT, Goldberg MS. 2003. Size-fractionated particulate mass and daily mortality in eight Canadian cities. Revised analyses of time-series studies of air pollution and health. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report. pp 85–89. Health Effects Institute, Boston, MA.

Daniels MJ, Dominici F, Samet JM, Zeger SL. 2000. Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. Am J Epidemiol 152:397–406.

Daniels MJ, Dominici F, Zeger SL, Samet JM. 2004. Part III. PM_{10} concentration–response curves and thresholds for the 20 largest US cities. In: The National Morbidity, Mortality, and Air Pollution Study. Research Report 94. Health Effects Institute, Boston, MA.

Derrienic F, Richardson S, Mollie A, Lellouch J. 1989. Short-term effects of sulphur dioxide pollution on mortality in two French cities. Int J Epidemiol 18:186–197.

Dockery DW, Schwartz J, Spengler JD. 1992. Air pollution and daily mortality: Associations with particulates and acid aerosols. Environ Res 59:362–373.

Dominici F, Daniels M, Zeger SL, Samet JM. 2000a. National models for estimating the effect of particulate matter on mortality in US cities. (Technical Report). Johns Hopkins University, Baltimore, MD.

Dominici F, Daniels M, Zeger SL, Samet JM. 2002. Air pollution and mortality: Estimating regional and national dose-response relationships. J Am Stat Assoc 97:100–111.

Dominici F, McDermott A, Daniels M, Zeger SL, Samet JM. 2003. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report. pp 9–24. Health Effects Institute, Boston, MA.

Dominici F, Samet J, Zeger SL. 2000b. Combining evidence on air pollution and daily mortality from the largest 20 U.S. cities: A hierarchical modeling strategy (with Discussion). J R Stat Soc [Ser A] 163:263–302.

Everson P, Morris C. 2000. Inference for multivariate normal hierarchical models. J Royal Stat Soc (B) 62:399–412.

Fairley D. 1990. The relationship of daily mortality to suspended particulates in Santa Clara County, 1980-1986. Environ Health Perspect 89:159–168.

Gamble JF, Lewis RJ. 1996. Health and respirable particulate (PM_{10}) air pollution: A causal or statistical association? Environ Health Perspect 104:838–850.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2003. Bayesian Data Analysis. Chapman and Hall Inc., New York, NY.

Gryparis A, Forsberg B, Katsouyanni K, Analitis A, Touloumi G, Schwartz J, Samoli E, Medina S, Anderson HR, Niciu EM, Wichmann HE, Kriz B, Kosnik M, Skorkovsky J, Vonk JM, Dörtbudak Z. 2004. Acute effects of ozone on mortality from the "air pollution and health: A European approach" project. Am J Respir Crit Care Med 170:1080–1087.

Hatzakis A, Katsouyanni K, Kalandidi A, Day N, Trichopoulos D. 1986. Short-term effects of air pollution on mortality in Athens. Int J Epidemiol 15:73–81.

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

Katsouyanni K. 2003. Ambient air pollution and health. Br Med Bull 68:143–156.

Katsouyanni K, Karakatsani A, Messari I, Touloumi G, Hatzakis A, Kalandidi A, Trichopoulos D. 1990. Air pollution and cause-specific mortality in Athens. J Epidemiol Community Health 44:321–324.

Katsouyanni K, Schwartz J, Spix C, Touloumi G, Zmirou D, Zanobetti A, Wojtyniak B, Vonk JM, Tobias A, Pönkä A, Medina S, Bachárová L, Anderson HR. 1996. Short term effects of air pollution on health: A European approach using epidemiologic time series data: The APHEA protocol. J Epidemiol Community Health 50 Suppl 1:S12–S18.

Katsouyanni K, Touloumi G, Samoli E, Gryparis A, Le Tertre A, Monopolis Y, Rossi G, Zmirou D, Ballester F, Boumghar A, Anderson HR, Wojtyniak B, Paldy A, Braunstein R, Pekkanen J, Schindler C, Schwartz J. 2001. Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 European cities within the APHEA2 project. Epidemiology 12:521–531.

Katsouyanni K, Touloumi G, Spix C, Schwartz J, Balducci F, Medina S, Rossi G, Wojtyniak B, Sunyer J, Bachárová L, Schouten JP, Pönkä A, Anderson HR. 1997. Short-term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: Results from time series data from the APHEA project. Air Pollution and Health: A European Approach. BMJ 314:1658–1663.

Lipfert FW, Wyzga RE. 1995. Air pollution and mortality: Issues and uncertainties. J Air Waste Manag Assoc 45:949–966.

Lipfert FW, Wyzga RE. 1997. Air pollution and mortality: The implications of uncertainties in regression modeling and exposure measurement. J Air Waste Manag Assoc 47:517–523.

McMichael AJ, Anderson HR, Brunekreef B, Cohen AJ. 1998. Inappropriate use of daily mortality analyses to estimate longer-term mortality effects of air pollution. Int J Epidemiol 27:450–453.

Moolgavkar SH, Luebeck EG, Hall TA, Anderson EL. 1995. Air pollution and daily mortality in Philadelphia. Epidemiology 6:476–484.

Peng RD, Dominici F, Louis T. 2006. Model choice in multi-site time series studies of air pollution and mortality. J R Stat Soc Ser A 169:179–203.

Peng RD, Dominici F, Pastor-Barriuso R, Zeger SL, Samet JM. 2005. Seasonal analyses of air pollution and mortality in 100 U.S. cities. Am J Epidemiol 161: 585–594.

Pope CA III. 1999. Mortality and air pollution: Associations persist with continued advances in research methodology. Environ Health Perspect 107:613–614.

Pope CA III, Dockery DW. 2006. Health effects of fine particulate air pollution: Lines that connect. J Air Waste Manag Assoc 56:709–742.

Pope CA III, Kalkstein LS. 1996. Synoptic weather modeling and estimates of the exposure–response relationship between daily mortality and particulate air pollution. Environ Health Perspect 104(4):414–420.

Pope CA III, Schwartz J, Ransom MR. 1992. Daily mortality and PM_{10} pollution in Utah Valley. Arch Environ Health 47:211–217.

Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000a. The National Morbidity, Mortality, and Air Pollution Study. Part I. Methods and Methodologic Issues. Research Report 94. Health Effects Institute, Cambridge, MA.

Samet JM, Zeger SL, Berhane K. 1995. Particulate Air Pollution and Daily Mortality: Replication and Validation of Selected Studies. The Phase I Report of the Particle Epidemiology Evaluation Project. Special Report. Health Effects Institute, Cambridge, MA.

Samet JM, Zeger SL, Dominici F, Curriero F, Coursac I, Dockery DW, Schwartz J, Zanobetti A. 2000b. 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.

Samet JM, Zeger S, Kelsall J, Xu J, Kalkstein L. 1998. Does weather confound or modify the association of particulate air pollution with mortality? An analysis of the Philadelphia data, 1972–1980. Environ Res 77(1):9–19.

Samoli E, Aga E, Touloumi G, Nisiotis K, Forsberg B, Lefranc A, Pekkanen J, Wojtyniak B, Schindler C, Niciu E, Brunstein R, Dodic Fikfak M, Schwartz J, Katsouyanni K. 2006. Short-term effects of nitrogen dioxide on mortality: An analysis within the APHEA project. Eur Respir J 27:1129–1138.

Samoli E, Analitis A, Touloumi G, Schwartz J, Anderson HR, Sunyer J, Bisanti L, Zmirou D, Vonk JM, Pekkanen J, Goodman P, Paldy A, Schindler C, Katsouyanni K. 2005. Estimating the exposure-response relationships between particulate matter and mortality within the APHEA multicity project. Environ Health Perspect 113:88–95.

Samoli E, Peng R, Ramsay T, Pipikou M, Touloumi G, Dominici F, Burnett R, Cohen A, Krewski D, Samet J, Katsouyanni K. 2008. Acute effects of ambient particulate matter on mortality in Europe and North America: Results from the APHENA Study. Environ Health Perspect 116:1480–1486.

Samoli E, Schwartz J, Wojtyniak B, Touloumi G, Spix C, Balducci F, Medina S, Rossi G, Sunyer J, Bacharova L, Anderson HR, Katsouyanni K. 2001. Investigating regional differences in short-term effects of air pollution on daily mortality in the APHEA project: A sensitivity analysis for controlling long-term trends and seasonality. Environ Health Perspect 109:349–353.

Samoli E, Touloumi G, Zanobetti A, Le Tertre A, Schindler C, Atkinson R, Vonk J, Rossi G, Saez M, Rabczenko D, Schwartz J, Katsouyanni K. 2003. Investigating the dose-response relation between air pollution and total mortality in the APHEA-2 multicity project. Occup Environ Med 60:977–982.

Schwartz J. 1991. Particulate air pollution and daily mortality in Detroit. Environ Res 56(2):204–213.

Schwartz J. 1993. Air pollution and daily mortality in Birmingham, Alabama. Am J Epidemiol 137: 1136–1147.

Schwartz J. 2004a. Is the association of airborne particles with daily deaths confounded by gaseous air pollutants? An approach to control by matching. Environ Health Perspect 112:557–561.

Schwartz J. 2004b. The effects of particulate air pollution on daily deaths: A multi-city case crossover analysis. Occup Environ Med 61:956–961.

Schwartz J. 2005. How sensitive is the association between ozone and daily deaths to control for temperature? Am J Respir Crit Care Med 171:627–631.

Schwartz J, Dockery DW. 1992a. Increased mortality in Philadelphia associated with daily air pollution concentrations. Am Rev Respir Dis 145:600–604.

Schwartz J, Dockery DW. 1992b. Particulate air pollution and daily mortality in Steubenville, Ohio. Am J Epidemiol 135:12–19.

Schwartz J, Marcus A. 1990. Mortality and air pollution in London: A time series analysis. Am J Epidemiol 131:185–194.

Schwartz J, Zanobetti A. 2000. Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. Epidemiology 11:666–672.

Sunyer J, Sáez M, Murillo C, Castellsague J, Martínez F, Antó JM.1993. Air pollution and emergency room admissions for chronic obstructive pulmonary disease: A 5-year study. Am J Epidemiol 137:701–705.

Thurston DG, Kinney PL. 1995. Air pollution epidemiology: Considerations in time-series modeling. Inhalation Toxicol 7:71–83.

Touloumi G, Pocock SJ, Katsouyanni K, Trichopoulos D. 1994. Short-term effects of air pollution on daily mortality in Athens: A time-series analysis. Int J Epidemiol 23:957–967.

Touloumi G, Samoli E, Pipikou M, Le Tertre A, Atkinson R, Katsouyanni K. 2006. Seasonal confounding in air pollution and health time-series studies: Effect on air pollution effect estimates. Stat Med 25:4164–4178.

Zmirou D, Schwartz J, Saez M, Zanobetti A, Wojtyniak B, Touloumi G, Spix C, Ponce de León A, Le Moullec Y, Bacharova L, Schouten J, Pönkä A, Katsouyanni K. 1998. Time-series analysis of air pollution and cause-specific mortality. Epidemiology 9(5):495–503.

SCIENTIFIC OVERSIGHT GROUP FOR THE APHENA PROJECT

RELATED HEI PUBLICATIONS: PARTICULATE MATTER AND OZONE

Number	Title	Principal Investigator	Date*
Research Reports			
140	Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality	D. Krewski	2009
139	Effects of Long-Term Exposure to Traffic-Related Air Pollution on Respiratory and Cardiovascular Mortality in the Netherlands: the NLCS-AIR Study	B. Brunekreef	2009
138	Health Effects of Real-World Exposure to Diesel Exhaust in Persons with Asthma	J. Zhang	2009
137	The Influence of Improved Air Quality on Mortality Risks in Erfurt, Germany	A. Peters	2009
131	Characterization of Particulate and Gas Exposures of Sensitive Subpopulations Living in Baltimore and Boston	P. Koutrakis	2005
130	Relationships of Indoor, Outdoor, and Personal Air (RIOPA)		
	Part I. Collection Methods and Descriptive Analyses	C.P. Weisel	2005
	Part II. Analyses of Concentrations of Particulate Matter Species	B.J. Turpin	2007
127	Personal, Indoor, and Outdoor Exposures to $PM_{2.5}$ and Its Components for Groups of Cardiovascular Patients in Amsterdam and Helsinki	B. Brunekreef	2005
125	Uptake Distribution of Ozone in Human Lungs: Intersubject Variability in Physiologic Response	J.S. Ultman	2004
124	Particulate Air Pollution and Nonfatal Cardiac Events		2005
	Part I. Air Pollution, Personal Activities, and Onset of Myocardial Infarction in a Case–Crossover Study	A. Peters	
	Part II. Association of Air Pollution with Confirmed Arrhythmias Recorded by Implanted Defibrillators	D. Dockery	
123	Time-Series Analysis of Air Pollution and Mortality: A Statistical Review	F. Dominici	2004
99	A Case–Crossover Analysis of Fine Particulate Matter Air Pollution and Out-of-Hospital Sudden Cardiac Arrest	H. Checkoway	2000
98	Daily Mortality and Fine and Ultrafine Particles in Erfurt, Germany *Part I.* Role of Particle Number and Particle Mass	H-E. Wichmann	2000
97	Identifying Subgroups of the General Population That May Be Susceptible to Short-Term Increases in Particulate Air Pollution: A Time-Series Study in Montreal, Quebec	M.S. Goldberg	2000
95	Association of Particulate Matter Components with Daily Mortality and Morbidity in Urban Populations	M. Lippmann	2000
94	National Morbidity, Mortality, and Air Pollution Study		
	Part I. Methods and Methodologic Issues	J.M. Samet	2000
	Part II: Morbidity and Mortality from Air Pollution in the United States	J.M. Samet	2000
	Part III: Concentration–Response Curves and Thresholds for the 20 Largest US Cities	M.J. Daniels	2004
	Part IV: Hierarchical Bivariate Time-Series Models—A Combined Analysis of PM_{10} Effects on Hospitalization and Mortality	F. Dominici	2005

Continued

** Reports published since 1998.*
Copies of these reports can be obtained from the Health Effects Institute and many are available at www.healtheffects.org.

RELATED HEI PUBLICATIONS: PARTICULATE MATTER AND OZONE

Number	Title	Principal Investigator	Date*
91	Mechanisms of Morbidity and Mortality from Exposure to Ambient Air Particles	J.J. Godleski	2000
90	Aldehydes (Nonanal and Hexanal) in Rat and Human Bronchoalveolar Lavage Fluid After Ozone Exposure	M.W. Frampton	1999
82	Acute Effects of Ambient Ozone on Asthmatic, Wheezy, and Healthy Children	E.L. Avol	1998
81	Methods Development for Epidemiologic Investigations of the Health Effects of Prolonged Ozone Exposure		
	Part I: Variability of Pulmonary Function Measures	I.B. Tager	1998
	Part II: An Approach to Retrospective Estimation of Lifetime Ozone Exposure Using a Questionnaire and Ambient Monitoring Data (California Sites)	I.B. Tager	1998
	Part III: An Approach to Retrospective Estimation of Lifetime Ozone Exposure Using a Questionnaire and Ambient Monitoring Data (U.S. Sites)	P.L. Kinney	1998
Special Reports			
	Traffic-Related Air Pollution: A Critical Review of the Literature		In Press
	Revised Analyses of Time-Series Studies of Air Pollution and Health		2003
HEI Communications			
14	HEI's Research Program on the Impact of Actions to Improve Air Quality: Interim Evaluation and Future Directions		2009
11	Assessing the Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research		2003
8	The Health Effects of Fine Particles: Key Questions and the 2003 Review (Report of the Joint Meeting of the EC and HEI)		1999
HEI Research Program Summaries			
	Research on Diesel Exhaust and Other Particles		2003
	Research on Particulate Matter		1999
HEI Perspectives			
	Understanding the Health Effects of Components of the Particulate Matter Mix: Progress and Next Steps		2002
	Airborne Particles and Health: HEI Epidemiologic Evidence		2001

** Reports published since 1998.*
Copies of these reports can be obtained from the Health Effects Institute and many are available at www.healtheffects.org.

HEI BOARD, COMMITTEES, and STAFF

Board of Directors

Richard F. Celeste, Chair *President, Colorado College*

Enriqueta Bond *President Emeritus, Burroughs Wellcome Fund*

Purnell W. Choppin *President Emeritus, Howard Hughes Medical Institute*

Michael T. Clegg *Professor of Biological Sciences, University of California–Irvine*

Jared L. Cohon *President, Carnegie Mellon University*

Stephen Corman *President, Corman Enterprises*

Gowher Rizvi *Vice Provost of International Programs, University of Virginia*

Linda Rosenstock *Dean, School of Public Health, University of California–Los Angeles*

Warren M. Washington *Head, Climate Change Research, National Center for Atmospheric Research*

Archibald Cox, Founding Chair *1980–2001*

Donald Kennedy, Vice Chair Emeritus *Editor-in-Chief Emeritus,* Science*; President Emeritus and Bing Professor of Biological Sciences, Stanford University*

Health Research Committee

Mark J. Utell, Chair *Professor of Medicine and Environmental Medicine, University of Rochester Medical Center*

Kenneth L. Demerjian *Ray Falconer Endowed Chair, Director and Professor, Atmospheric Sciences Research Center and Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York*

Joe G.N. Garcia *Lowell T. Coggeshall Professor of Medicine, University of Chicago*

Uwe Heinrich *Professor, Medical School Hannover, Executive Director, Fraunhofer Institute for Toxicology and Experimental Medicine, Hanover, Germany*

Grace LeMasters *Professor of Epidemiology and Environmental Health, University of Cincinnati College of Medicine*

Sylvia Richardson *Professor of Biostatistics, Department of Epidemiology and Public Health, Imperial College School of Medicine, London, United Kingdom*

Howard E. Rockette *Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh*

James A. Swenberg *Kenan Distinguished Professor of Environmental Sciences, Department of Environmental Sciences and Engineering, University of North Carolina–Chapel Hill*

Ira B. Tager *Professor of Epidemiology, School of Public Health, University of California–Berkeley*

HEI BOARD, COMMITTEES, and STAFF

Health Review Committee

Homer A. Boushey, Chair *Professor of Medicine, Department of Medicine, University of California–San Francisco*

Ben Armstrong *Reader in Epidemiological Statistics, Public and Environmental Health Research Unit, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, United Kingdom*

Michael Brauer *Professor, School of Environmental Health, University of British Columbia, Canada*

Bert Brunekreef *Professor of Environmental Epidemiology, Institute of Risk Assessment Sciences, University of Utrecht, the Netherlands*

Alan Buckpitt *Professor of Toxicology, Department of Molecular Biosciences, School of Veterinary Medicine, University of California–Davis*

John R. Hoidal *Professor of Medicine, the Clarence M. and Ruth N. Birrer Presidential Endowed Chair, Department of Medicine, University of Utah Health Sciences*

Stephanie London *Senior Investigator, Epidemiology Branch, National Institute of Environmental Health Sciences*

William N. Rom *Sol and Judith Bergstein Professor of Medicine and Environmental Medicine and Director of Pulmonary and Critical Care Medicine, New York University Medical Center*

Armistead Russell *Georgia Power Distinguished Professor of Environmental Engineering, School of Civil and Environmental Engineering, Georgia Institute of Technology*

Lianne Sheppard *Professor of Biostatistics, School of Public Health, University of Washington–Seattle*

Officers and Staff

Daniel S. Greenbaum *President*
Robert M. O'Keefe *Vice President*
Rashid Shaikh *Director of Science*
Barbara Gale *Director of Publications*
Jacqueline C. Rutledge *Director of Finance and Administration*
Helen I. Dooley *Corporate Secretary*

Kate Adams *Staff Scientist*
Aaron J. Cohen *Principal Scientist*
Maria G. Costantini *Principal Scientist*
Philip J. DeMarco *Compliance Manager*
Terésa Fasulo *Science Administration Manager*
Hope Green *Editorial Assistant (part time)*
L. Virgi Hepner *Senior Science Editor*
Sarah Katz *Research Assistant*
Anny Luu *Administrative Assistant*
Francine Marmenout *Senior Executive Assistant*
Flannery Carey McDermott *Editorial Assistant*
Sumi Mehta *Senior Scientist*
Nicholas Moustakas *Policy Associate*
Tiffany North *Research Associate*
Hilary Selby Polk *Science Editor*
Robert A. Shavers *Operations Manager*
Geoffrey H. Sunshine *Senior Scientist*
Annemoon M.M. van Erp *Senior Scientist*
Katherine Walker *Senior Scientist*

HEALTH
EFFECTS
INSTITUTE

101 Federal Street, Suite 500
Boston, MA 02110, USA
+1-617-488-2300
www.healtheffects.org

RESEARCH
REPORT

Number 142
October 2009