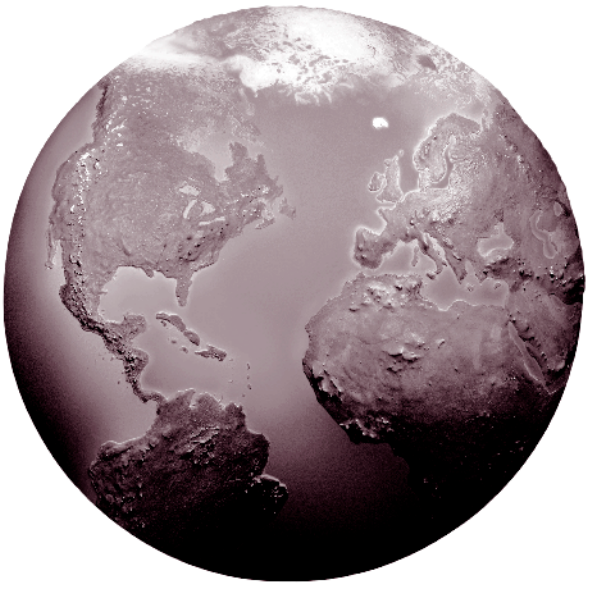
HEI

COMMUNICATION 11

Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research

HEI Accountability Working Group



HEALTH EFFECTS INSTITUTE

September 2003

Contributors

This Communication is the work of a multidisciplinary group, chaired by Dr Jonathan M Samet of the Johns Hopkins Bloomberg School of Public Health, that included participants from the Health Effects Institute's Health Research and Review Committees, HEI staff, and other invited experts. An iterative workshop process was used to develop the Communication's novel concepts and subject matter. For the first workshop in March 2002, HEI convened a multidisciplinary group of experts, several of whom were asked to draft text in advance. At the workshop, participants met in plenary and in small groups, revising draft text on the basis of their discussions. A subset of the workshop participants was then asked by HEI to join the Accountability Working Group responsible for drafting specific chapters. This group met in June 2002 to review and revise the draft chapters and prepare a final draft. The draft was sent for peer review to a broad range of experts and stakeholders, several of whom presented their views at the 2003 HEI Annual Conference. This final document was then prepared by Dr Samet and HEI staff.

HEI Accountability Working Group

Jonathan M Samet, Working Group Chair
Johns Hopkins Bloomberg School of Public Health, HEI Health Research Committee

Aaron J Cohen, *Health Effects Institute*

Kenneth L Demerjian, *State University of New York at Albany, HEI Health Research Committee*

Daniel S Greenbaum, *Health Effects Institute*

Patrick Kinney, *Columbia University*

Thomas A Louis, *Johns Hopkins Bloomberg School of Public Health, HEI Health Review Committee*

Joe L Mauderly, *Lovelace Respiratory Research Institute*

Robert L Maynard, *UK Department of Health*

Richard D Morgenstern, *Resources for the Future*

Robert M O'Keefe, *Health Effects Institute*

C Arden Pope III, *Brigham Young University*

Mark J Utell, *University of Rochester, HEI Health Research Committee (Chair)*

Sverre Vedal, *National Jewish Medical and Research Center, HEI Health Review Committee*

Other Workshop Participants

Ross Anderson, *St George's Hospital Medical School, HEI Health Review Committee*

Steve D Colome, *University of California, Los Angeles*

Mark Goldberg, *McGill University*

Alan J Krupnick, *Resources for the Future*

Michal Krzyzanowski, *WHO European Center for Environment and Health*

M Granger Morgan, *Carnegie Mellon University*

Maria Segui-Gomez* *Johns Hopkins Bloomberg School of Public Health*

Ira Tager, *University of California–Berkeley School of Public Health, HEI Health Research Committee*

Gregory Wagner, *US National Institute of Occupational Safety and Health*

Peer Reviewers

John D Bachmann, *US Environmental Protection Agency*

John C Bailar III, *University of Chicago, HEI Health Review Committee*

Thomas A Burke, *Johns Hopkins Bloomberg School of Public Health*

Richard T Burnett, *Health Canada*

David P Chock, *Ford Motor Company*

Bart E Croes, *California Air Resources Board*

Richard B Stewart, *New York University, HEI Board of Directors*

Jonathan B Wiener, *Duke University*

George T Wolff, *General Motors Corporation*

HEI Project Staff

Jenny Lamont, *Science Editor*

Sally Edwards, *Director of Publications*

Jane Warren, *Director of Science*

Publication Assistance

Melissa R Harke, *HEI Administrative Assistant*

Ruth E Shaw, *Compositor, Cameographics*

* Maria Segui-Gomez deserves special acknowledgment for writing Appendix A, Chapter 1.

CONTENTS

HEI Communication 11

HEALTH EFFECTS INSTITUTE

Continued

CHAPTER 4. HEALTH OUTCOMES FOR MEASURING ACCOUNTABILITY

CHAPTER 5. MODEL DESIGN AND DATA ANALYSIS

CHAPTER 6. DIRECTIONS FOR FUTURE RESEARCH

Publishing History: This document was posted as a preprint on *www.healtheffects.org* and then finalized for print.

Citation for whole report:

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

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

EXECUTIVE SUMMARY

GROWING INTEREST IN ACCOUNTABILITY

Protecting public health from environmental risks involves taking regulatory and other actions on the basis of population statistics or patterns of clinical disease and tracking their consequences so that efforts can be redirected as indicated by subsequent evidence. As some indicators of environmental quality have improved for the United States, specific measures of progress in improving public health have been sought and questions have been raised as to whether public health goals have been met. These questions have emerged with particular force with regard to marked improvements in air quality in the United States in recent decades and ongoing efforts to further improve air quality.

Evaluating the extent to which air quality regulations improve public health is part of a broad effort—termed *accountability*—to assess the performance of all environmental regulatory policies. Air quality has improved substantially in the United States and western Europe in recent decades, with far less visible pollution and dropping concentrations of several major pollutants. In large part, these gains have been achieved through increasingly stringent air quality regulations that often require costly control measures to implement. For example, since 1980, measurements at thousands of monitoring stations across the United States have shown decreasing concentrations for all six criteria pollutants. This progress, of course, has come at a price. The US Environmental Protection Agency (EPA*) estimates that from 1970 to 1990 the annualized cost of air pollution control was about $25 billion per year—more than $500 billion over 20 years. Even as new research findings appear to have strengthened the evidence for health effects, many (including policy makers, legislators, industry, and the public) ask whether past efforts to reduce air pollution have yielded demonstrable improvements in public health and whether future efforts will continue to do so.

Although risk assessments estimate a substantial burden of premature mortality and excess morbidity even at current levels of ambient pollution, direct evidence is lacking about the extent to which control measures have improved health. This dearth has prompted efforts to assess and collect such evidence, including the recent US National Research Council (NRC) report, *Estimating the Public Health Benefits of Proposed Air Pollution Regulations* (NRC 2002); research funded by California Air Resources Board; and recent efforts by the EPA, the US Centers for Disease Control and Prevention (CDC), and other agencies to improve surveillance of the environment and public health. It is also the basis for an initiative by the Health Effects Institute, comprising epidemiologic studies of improvements in air quality and health, and this Communication: *Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research*, written by a multidisciplinary HEI Accountability Working Group. Communication 11 sets out a conceptual framework for accountability research and identifies types of evidence required and methods by which the evidence can be obtained.

CHALLENGE OF MEASURING HEALTH IMPACT OF AIR QUALITY REGULATIONS

Consideration of accountability starts with assessment of the effectiveness of regulations for reducing emissions and whether reductions have affected ambient concentrations as intended. Assessments must then evaluate whether adverse health effects of air pollution have been reduced. Some national governments and public health agencies have attempted to quantify the past health impact of air quality improvements and to estimate future impact. To date, these attempts have used risk estimates from epidemiologic studies to calculate the impact of air pollution on public health in terms of disease burden under hypothetical air quality scenarios. These scenarios reflect either continuation of past patterns of exposure or future patterns of exposure under more stringent controls (EPA 1999). However, these estimates have not been extensively validated against studies of actual regulatory programs and other interventions.

Research on accountability is in its early stages, but even now considerable challenges in assessing the health impact of air quality regulations can be anticipated. Air quality regulations themselves are promulgated to take effect at different times and on multiple governmental levels. Therefore, diverse approaches are needed to evaluate the impact of interventions on human health at national, regional, and local levels and in various time frames. We can also expect the consequences of interventions to extend beyond changes in air quality. Whether or not an intervention improves air quality, it could result in changes in personal activities and behaviors or in economic activities that could in turn affect health. Therefore, the causal pathways from a regulation and its consequences for air quality to a change in its risk to health could be difficult to isolate. Additionally, adverse health

* A list of abbreviations and other terms appears at the end of the Executive Summary.

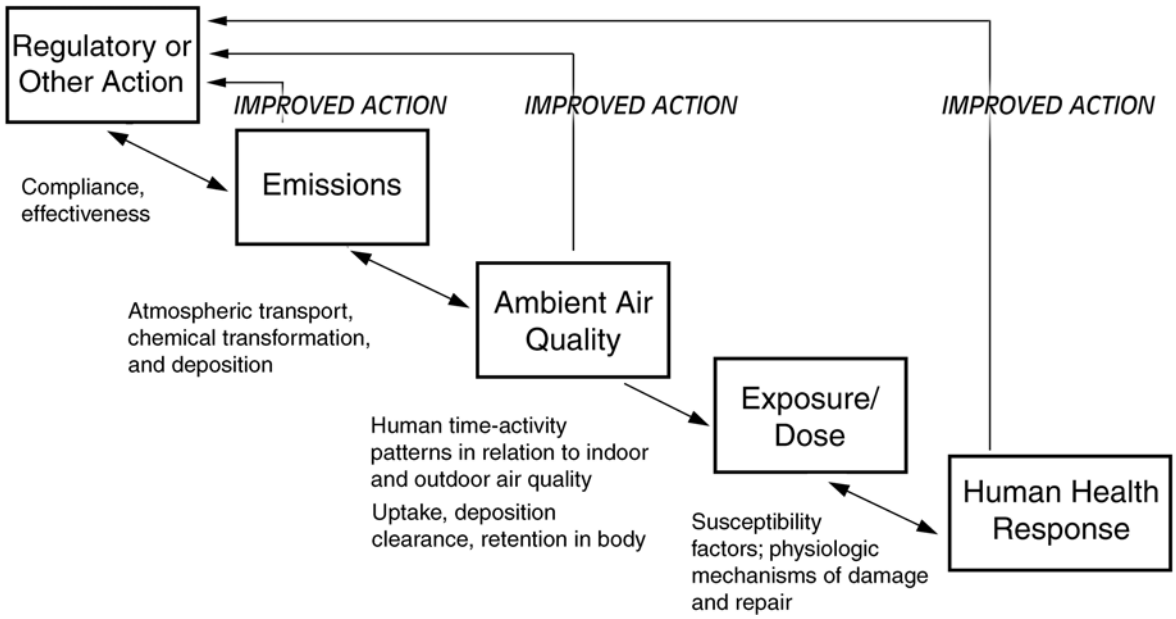
effects that may be caused by exposure to air pollution can also be caused by other factors (some changing over the same time periods as air pollution concentrations).

Regulatory interventions to improve air quality, especially large national programs such as the US Clean Air Act, may not immediately affect either air quality or public health. Once regulations are instated, changes in pollutant emissions, ambient pollutant concentrations, and human exposure to ambient concentrations may not be immediately or uniformly evident, and the biological processes of injury that underlie the health effects of air pollution may not be immediately evident. The longer the time between promulgation of regulations and their effects, the greater the possibility that other events may come into play and interfere with isolating effects of the interventions themselves. The level of enforcement may complicate the analysis by extending the anticipated time between intervention and effect. On the other hand, some interventions may produce relatively rapid changes in air quality, the impact of which may be measurable soon after. Recent studies of the health effects of air quality improvement programs implemented over short time frames in Ireland (Clancy et al 2002) and Hong Kong (Hedley et al 2002) are examples. Rapid changes reduce, but by no means eliminate, the possibility of confounding by other risk factors.

The need to measure the health impact of air quality regulations and other actions, and to improve the evidentiary basis for assessing the effectiveness of those actions, is clear. Whether and to what extent observational study designs can meet these challenges is the task before us. The inherent challenges are well documented in Communication 11, but recent advances in data collection and analytic techniques provide an unprecedented opportunity to improve our assessments of the impact of air quality actions. To this end, this Communication is intended to advance the concept of accountability and to foster development of accountability methods and research throughout the scientific and policy communities.

THE CHAIN OF ACCOUNTABILITY: A FRAMEWORK FOR RESEARCH AND PUBLIC HEALTH ACTIVITY

Future efforts to measure the impact of air quality regulations will need to be based on a conceptual framework that identifies key relations to be estimated and the resources needed to do so. The NRC's Committee on Research Priorities for Airborne Particulate Matter (NRC 1998) set out a framework for linking air pollution sources to adverse health effects. This framework can be used to identify indicators for use in accountability assessment and leads to designation of the chain of accountability, which parallels the links from sources to health effects (Figure 1). The connections between the links correspond to typical points at which quantitative measures of accountability are possible.



Chain of accountability. Each box represents a link between regulatory action and human health response to air pollution. Arrows connecting the links indicate possible directions of influence. Text below the arrows identifies general indices of accountability at that stage. At several stages, knowledge gained from accountability assessment can provide valuable feedback for improving regulatory or other action.

Along the length of the chain of accountability, a number of indicators can be identified, and accountability questions addressed:

- **Regulatory action**: Have controls on source emissions been put into place?
- **Emissions**: Have the source controls reduced emissions? Have there been unanticipated and untoward consequences? Answers to these questions require evidence about how regulation has changed the practices of emitters and about what changes in emissions have resulted.
- **Ambient air quality**: Have concentrations of air pollutants declined consequent to source control and emissions reductions? Answers to this question require evidence based on the periodic standardized measurement of ambient concentrations of air pollution constituents, such as particles, ozone, and sulfur dioxide.
- **Personal exposure**: Has exposure to air pollution declined? For which groups in the population (particularly as defined with reference to susceptibility to the effects of air pollution)? For instance, have exposures been reduced for disadvantaged people, including racial and ethnic groups, who may experience disproportionately high exposures and may be more susceptible? The relation between concentration and exposure can be modified by time-activity patterns, and such modification may mean that declines in concentration do not lead to proportional declines in exposure. In assessing changes in exposure, two factors should be considered: how ambient concentrations have changed as a result of changes in emissions, and how the intervention may have changed the behavior (and thus exposure) of the population(s) it was intended to protect.
- **Dose to target tissues**: Have reductions in exposure led to reductions in dose? The relative consequences of a dose for a susceptible versus a nonsusceptible person may also be relevant.
- **Human health response**: Have health risks declined? This indicator requires evidence about changes in health endpoints that have resulted from changes in exposure. Research must address which health endpoints and measurement techniques are most directly attributable to air pollution exposure, and thus would be most useful for accountability assessments, as well as how the health endpoints should be defined and characterized for analysis.

At each link in the chain, the opportunity exists to collect evidence to either validate the assumptions that motivated intervention or identify ways in which those assumptions were incorrect. Using such evidence can thus ensure that future interventions are maximally effective.

This framework fits well with the approach taken in the United States for regulation of principal pollutants. Air pollution regulations for criteria pollutants specify National Ambient Air Quality Standards (NAAQS)—maximum concentrations of selected pollutants in air. Each standard specifies a pollutant, its concentration, the averaging time, and the proportion of time that the standard must be met. The concentration limits are met by source control and related reductions of emissions. This approach inherently assumes that source control and emissions reductions to meet target concentrations will reduce human exposure to targeted pollutants and, subsequently, reduce risks of adverse health effects. Assessing the health impact of programs such as the US Clean Air Act calls for a demonstration that implementation of measures to reach the NAAQS has in fact led to a reduced disease burden, which is the final step in the chain of accountability. In the general framework considered here, implementation of a NAAQS would ideally be followed by comprehensive surveillance for each indicator. Although the EPA tracks ambient pollutant concentrations with monitoring networks and requires the development of plans that specify emissions controls, there is currently no way to connect in a comprehensive fashion the sources at one end of the chain to adverse health effects at the other.

Ultimately, the framework for accountability assessment will need to be extended beyond targeted pollutants and associated health risks. The measures that are needed to reduce ambient concentrations of major pollutants may have broad consequences, some unintended and unanticipated, which could reduce or increase risks to public health. Wiener (1998) and others have advanced the concept of a so-called portfolio of effects of a regulation and argue that the full set of effects, not just the intended effects, needs to be evaluated. Although this Communication focuses primarily on measuring the health-related consequences of regulation, each link in the chain of accountability is placed in this broader context.

METHODS FOR ACCOUNTABILITY STUDIES

The success of future accountability assessments will require further development and application of epidemiologic and biostatistical approaches in three areas: assessment of exposure and dose, selection of health outcomes, and study design and data analysis.

ASSESSING POPULATION EXPOSURE AND DOSE

Policies to improve health of the population by controlling emissions will be successful only if the emissions reductions ultimately result in reduced population exposures to and doses of the air pollutant of concern. Several general strategies can be used to measure exposure or dose in accountability studies. One involves the use of large-scale, periodic, random monitoring surveys of the general population to document long-term trends in exposure or dose. Examples of the large-scale survey designs include periodic US National Health and Nutrition Examination Survey (NHANES) (US National Center for Health Statistics 2003) and the US National Human Exposure Assessment Study (NHEXAS) (EPA 2003a). On such a large scale, personal monitoring would be constrained by cost considerations as well as the need to limit the burden on participants. When practical, and if an appropriate biomarker is available, blood or other biospecimens could be collected for analysis. Another strategy involves smaller-scale studies of specific subpopulations to document exposure and dose before and after specific interventions. This strategy could yield richer data on personal exposure and biological dose measurements. Carefully designed and targeted small-scale field studies of personal exposures or doses could play a critical role in assessing interventions directed at reducing pollution and improving health. Such small-scale studies would complement large-scale surveys that aim to track long-term trends broadly across the population.

SELECTING HEALTH OUTCOMES FOR ACCOUNTABILITY STUDIES

Air quality regulations are established with the primary purpose of protecting the public's health. Regulatory action is taken on the basis of evidence of a causal association between exposure to air pollution and health risk. The outcomes considered in assessments of accountability should, therefore, reflect the evidence on which estimates of health benefits and regulation are based.

To estimate the impact of specific regulations, however, certain outcomes may easier or harder to apply. The most serious health outcomes (such as mortality and increased morbidity from cardiovascular and respiratory diseases) are associated with not just one but several pollutants as well as other behavioral and environmental factors. Thus, although researchers planning studies of the health impact of air quality regulations have a variety of possible health outcomes from which to choose, none are associated uniquely with air pollution. In addition, a range of practical considerations will determine the feasibility of using specific endpoints for accountability research. For example, national databases currently exist for some endpoints of interest (eg, mortality via the US National Centers for Health Statistics, hospitalization via the US Health Care Finance Administration). But some data including baseline (preintervention) rates of some endpoints (eg, asthma prevalence in major cities across the United States) may be unavailable or limited.

Choice of health endpoints for assessments of the health impact of air quality regulations will depend critically on temporal relations among changes in pollutant emissions, concentrations, and exposure and on development of a detectable endpoint. Endpoints that might be detectable shortly after exposures change are counts of daily deaths and hospitalizations, certain clinical endpoints such as medication use, and subclinical indices (such as changes in pulmonary function that can be linked to adverse clinical conditions). Biomarkers of health response have the potential to predict the health impact of regulations without waiting for disease outcomes. Using biomarkers is challenging, however. For instance, relations among biomarkers and health endpoints must be demonstrated and biomarkers must be validated under field conditions.

Endpoints that might be appropriate for assessing the long-term impact of air quality regulations include long-term average rates of adult and infant mortality, effects on average population lifespan, incidence of chronic cardiovascular and respiratory disease, and biomarkers such as age-related growth and decline of lung function. Studies of long-term impact will probably need to use information on spatial variation in exposure resulting from regulatory interventions, as has been done in earlier epidemiologic studies. Existing cohorts may offer some opportunities for accountability assessments if the period of observation allows preintervention baselines to be established and the duration of follow up is long enough to allow observation of intervention-related long-term effects.

MODEL DESIGN AND DATA ANALYSIS

Much evidence of health effects of air pollution comes from observational studies that relate changes in health indicators to changes in exposure to air pollution in space and time. The results of such studies are then used to estimate the expected benefits of current and future air pollution regulations. Interventions may be broadly viewed as any intentional (planned, such as a regulatory program) or unintentional (unplanned, such as a labor strike closing an industrial facility) change in air pollutant concentrations or exposures. Direct study of the effect of such interventions may be a more definitive approach to determining whether air pollution regulations actually result in health

benefits. Compared with the usual observational studies, studies of interventions can disrupt links between confounding factors and exposures that may be unavoidable and cause bias in many observational studies of environmental factors and health.

Some might argue that assessing accountability depends only on better science—that scientific methods, in particular those of epidemiology and toxicology, could be enhanced to generate better evidence and provide a platform for decision making. Although improved science is, without question, central to both applications, new scientific methods alone would be insufficient to integrate evidence across the chain of accountability. As Figure 1 indicates, accountability assessment may incorporate evidence over a long, complex chain of relations. Therefore, its evaluation requires integrating information that is directly and indirectly related. Statistical synthesis is central to such integration. It should describe relations among the links of the chain of accountability, generate the statistical relations needed for evaluation, and identify gaps in data or research. Bayesian approaches that explicitly incorporate summaries of prior knowledge may be particularly valuable.

DIRECTIONS FOR FUTURE RESEARCH

On the basis of this Communication, the Health Effects Institute plans to develop methods for accountability assessment and fund studies that address accountability. To that end, Communication 11 makes wide-ranging recommendations for an agenda to advance understanding and assessment of accountability. These recommendations fall into three general categories: (1) developing and implementing new study designs; (2) identifying targets of opportunity for accountability research; and (3) developing surveillance systems to track prospectively the health impact of air quality regulations. As HEI and other organizations move forward on accountability research, priorities among these opportunities will need to be set.

OPPORTUNITIES

The continually changing regulation of air pollution in the United States, Europe, and elsewhere affords an immediate set of opportunities for accountability assessment on national, regional, and local scales. In the United States, possible targets include:

- **$PM_{2.5}$ and ozone NAAQS implementation**. The state implementation plan (SIP) process is now in its initial stages for the recently promulgated $PM_{2.5}$ (particulate matter less than 2.5 μm in aerodynamic diameter) and ozone NAAQS. Extensive data on nationwide $PM_{2.5}$ concentrations are now being collected from a new monitoring network, establishing baseline conditions against which future emissions reductions can be assessed. (Such data already exist for ozone concentrations.) The state implementation plan process for $PM_{2.5}$ and ozone could provide an opportunity for accountability assessments that address changes in emissions, ambient concentrations, and exposures or doses to the population.
- **EPA's Air Toxics Control Plan.** EPA is required to assess risks and, if necessary, control the 188 air pollutants now classified as hazardous. Relevant research for accountability might include longitudinal measurements of pollutant emissions and ambient concentrations and identification of health endpoints that could be tracked in the near term. This approach is most applicable for hazardous air pollutants associated with short-term responses (eg, irritants).
- **Targets at local level.** Relatively rapid changes in ambient concentrations may occur in a local area as a result of a major change in local source emissions due to regulatory action. Numerous opportunities exist for studies of such interventions throughout the United States and elsewhere. For example, the New York City Metropolitan Transit Authority has plans to convert bus fueling and storage depots from diesel to natural gas, thereby possibly reducing neighborhood levels of elemental carbon and other diesel-related particle components. Control programs for major stationary pollution sources might also provide opportunities. Because these interventions occur over relatively short times and small areas, assessment studies aimed at documenting cause-effect relations between emissions changes and changes in exposure or health can be both economically and logistically feasible.

STUDY DESIGNS

Success of future research requires systematic identification of research needs and opportunities and commissioned studies to address them. This research will probably entail both adaptation (or tuning) of existing methods to suit specific needs and development of long-term surveillance of both health outcomes and potential confounders. Both conceptual and methodologic issues need to be addressed, including the fundamental step of assuring a uniform concept of accountability among researchers and regulators. Needed activities include the following:

- Focused research planning and further elaboration of study designs to assess accountability. Planning may include workshops involving the many stakeholders

concerned with accountability and moving toward a shared understanding of the concept.

- Review of available information and development of a mechanism for identifying possibly informative natural experiments and a process for following up on them. Reviewing information will include creating syntheses of studies relevant to accountability assessment in which different approaches are used to combine data across studies in order to properly gauge the weight of the evidence. A formal analysis could help identify the most important gaps in information which, if filled, would exert the most leverage on both scientific knowledge and public policy decisions.
- Analysis of model-based predictions of health impact to compare predicted and observed effects while accounting for model uncertainty. Findings from those analyses that appropriately address uncertainty can provide insight into the information required to make more accurate accountability assessments.
- Development of cooperative research models to anticipate settings that may be affected by changes in air pollution regulations or policies and that could provide opportunities for accountability assessment. Such mechanisms will likely require funding agencies to reach out to regulators and other governmental agencies and affected communities.
- Implementation of specific study designs:
    - Serial cross-sectional studies that could be completed within relatively short time frames and then repeated after an air pollution intervention. The timing of the repeated studies would be determined by the postulated latency period between any changes in exposure and health outcomes.
    - Randomized studies aimed at rigorously providing information about one or several connections in the chain of accountability. Randomizing exposure in real-life settings through randomized manipulation of behavior (eg, provision of air conditioners or indoor versus outdoor exercise regimens) is an example of providing information about the connection between the exposure and health effect links.
    - Cohort studies that may serve as the basis for accountability assessments. Ongoing cohort studies provide limited information for the windows of exposure applicable to study participants (eg, HEI 2001; Pope et al 2002). Insights from cohort studies might be improved by combining evidence from multiple cohorts to widen the exposure windows that could be assessed.

NEED FOR SURVEILLANCE

Taking full advantage of these opportunities will require data collected through ongoing surveillance of major time-varying links in the chain of accountability, or at least those components dealing with exposure and health outcomes. Existing data on air pollution and precedents for using national databases for air pollution analyses (Samet et al 2000) suggest that the issues of air pollution and public health provide an excellent opportunity to pilot the use of emerging surveillance systems for informing decisions about public health interventions. Some elements of such a system already exist, such as the National Death Index of the US National Centers for Health Statistics, NHANES, and the EPA and state air monitoring networks.

A system for the long-term surveillance of the health impact of air quality regulations will initially require evaluation of the adequacy of these existing resources in the context of a proposed study design. This evaluation would also need to consider what kinds of information would be required for long-term evaluation of health impact, including goals for efficacy and effectiveness. Several recent and ongoing efforts have already made important contributions. These include the CDC Environmental Public Health Indicators Project, which is evaluating a range of health indicators that could be used to track changes in health outcomes caused by environmental factors, and two recently released EPA reports, *America's Children and the Environment* (EPA 2003b) and the *Draft Report on the Environment* (EPA 2003c). A Nationwide Health Tracking Act is being considered by the US Congress to develop a comprehensive system for identifying and monitoring chronic diseases and correlating their causes with environmental, behavioral, socioeconomic, and demographic risk factors. New funding is enabling CDC to begin to build such systems with the help of many states and three university-based Centers of Excellence. These systems would be a great asset for accountability research.

CONCLUSIONS

Air quality in the United States and western Europe has improved considerably in recent decades, yet findings of continuing adverse health effects have prompted increasingly stringent air quality regulations. Demonstrating that these regulations are producing the desired health benefits will require creative and rigorous application of epidemiologic research methods and public health surveillance approaches within a conceptual framework for assessing accountability at each stage of the regulatory process. This Communication proposes such a framework and begins to

identify opportunities to implement it. Improvement of our ability to measure the health impact of regulations will require new levels of collaboration among the research community and federal, state, and local agencies charged with protection of both the environment and public health.

REFERENCES

Clancy L, Goodman P, Sinclair H, Dockery DW. 2002. Effect of air-pollution control on death rates in Dublin, Ireland: An intervention study. Lancet 360:1210–1214.

Environmental Protection Agency (US). 1999. The Benefits and Costs of the Clean Air Act 1990 to 2010: EPA Report to Congress. EPA-410-R-99-001. Office of Air and Radiation, Washington DC.

Environmental Protection Agency (US). 2003a. Human exposure measurements: National Human Exposure Assessment Survey (NHEXAS) (last updated 7/21/03). *www.epa.gov/heasd/edrb/ nhexas.htm*. Accessed 8/21/03.

Environmental Protection Agency (US). 2003b. America's Children and the Environment: Measures of Contaminants, Body Burdens, and Illnesses (2nd edition). EPA 240-R-03-001. National Center for Environmental Economics, Washington DC. Available from *www.epa.gov/envirohealth/ children/.*

Environmental Protection Agency (US). 2003c. Draft Report on the Environment. EPA 260-R-02-006. EPA, Washington DC. Available from *www.epa.gov/indicators/ roe/html/roePDF.htm.*

Health Effects Institute. 2001. Airborne Particles and Health: HEI Epidemiologic Evidence. HEI Perspectives. Health Effects Institute, Cambridge MA.

Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. 2002. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: An intervention study. Lancet 360:1646–1652.

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

National Research Council (US). 2002. Estimating the Public Health Benefits of Proposed Air Pollution Regulations. National Academy Press, Washington DC.

Pope CA III, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287:1132–1141.

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

Wiener J. 1998. Managing the iatrogenic risks of risk management. Risk Health Safety Environ 9:39–82.

ABBREVIATIONS AND OTHER TERMS

CDC	Centers for Disease Control and Prevention (US)
EPA	Environmental Protection Agency (US)
NAAQS	National Ambient Air Quality Standard(s) (US)
NHANES	National Health and Nutrition Examination Survey (US)
NRC	National Research Council (US)
$PM_{2.5}$	particulate matter less than 2.5 μm in aerodynamic diameter

Library of Congress Catalog Number for the HEI Report Series: WA 754 R432.

The paper in this publication meets the minimum standard requirements of the ANSI Standard Z39.48-1984 (Permanence of Paper) effective with Report 21 in December 1988; and effective with Report 92 in 1999 the paper is recycled from at least 30% postconsumer waste with Reports 25, 26, 32, 51, 65 Parts IV, VIII, and IX, 91 and 105 excepted. These excepted Reports are printed on acid-free coated paper.

Chapter 1. SEEKING ACCOUNTABILITY: AN OVERVIEW

CONTEXTS AND CHALLENGES

Protecting public health from environmental risks involves taking regulatory and other actions on the basis of population statistics or patterns of clinical disease and tracking their consequences so that efforts can be redirected as indicated by subsequent evidence. As some indicators of environmental quality have improved for the United States, specific measures of progress in improving public health have been sought and questions have been raised as to whether public health goals have been met. These questions have emerged with particular force with regard to marked improvements in air quality in the United States in recent decades and to ongoing efforts to further improve air quality. In response, the US Environmental Protection Agency (EPA*) and other health agencies have sought recently to improve surveillance of the environment and public health.

Evaluating whether air quality regulations improve public health is part of the broad effort—termed *accountability*—to assess the performance of environmental regulatory policy. The emphasis on accountability in air quality regulations can be viewed as a byproduct of improved environmental quality. Air quality has improved substantially in the United States and western Europe in recent decades, with far less visible pollution and dropping concentrations of several major pollutants. In large part, these gains have been achieved through increasingly stringent air quality regulations that often require costly control measures to implement. Even at current low ambient pollution levels, risk assessments estimate a substantial burden of premature mortality and excess morbidity. However, carefully constructed evidence about the extent to which control measures have improved health (ie, *accountability assessments*) is still lacking. This Communication, prepared by the Accountability Working Group of the Health Effects Institute, is intended to promote research to strengthen accountability assessments. In this chapter, we introduce the broad context for the Communication by discussing the concept of accountability that extends throughout the chapters.

The concept of accountability fits within the general framework for problem solving in public health: monitoring and identifying threats to public health; developing prevention and control programs; and tracking the consequences of these programs with redirection when indicated by subsequent evidence. In this framework, various lines of evidence are tracked to identify emerging problems and to monitor the status of those for which control measures have been implemented. For infectious diseases, for example, the experience of individual clinicians and health care institutions is often the means by which problems are first identified. Sudden epidemics, such as the recent emergence of severe acute respiratory syndrome (SARS), are often identified because they present an unusual clinical pattern. For diseases with prevention and control programs in place, various monitoring systems track their occurrence. For example, the United States and other countries have population-based cancer registries that track incidence rates. After decades of vigorous programs to control the use of tobacco, national incidence and mortality data for lung cancer indicate that its occurrence is on the decline, particularly in males. Although not explicitly applied previously to the combustion-related criteria pollutants, this public health framework has been central in limiting lead exposure to children (Cohen et al 1990).

The need for air pollution control is based on extensive scientific evidence for adverse effects of ambient air pollution on a range of health outcomes from eye irritation to premature mortality (see, for example, Bascom et al 1996a,b; American Thoracic Society 2000). Based on this collective evidence, guidelines and standards designed to improve air quality are promulgated by governments and public health agencies from the local to the national levels. The expectation is that past reductions of air pollution levels must have benefited public health and prevented some amount of human disease and death and that future effort will have additional benefits. At the same time, efforts to improve air quality entail large financial costs and possibly substantial

* A list of abbreviations and other terms appears at the end of this chapter.

This chapter is one part of Health Effects Institute Communication 11, *Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research,* which also includes five other chapters and an Executive Summary of the project.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R82811201 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.

unanticipated consequences for society. Therefore, policy makers and other stakeholders seek evidence that interventions to improve air quality do in fact benefit public health. Such evidence would contribute importantly to assessments of both their efficacy for improving public health and their economic costs and benefits.

The current interest in accountability is a response to reasonable questioning of the benefits of further tightening of air pollution regulations. As levels of air pollution have fallen in the United States and other developed nations over the past several decades, the assumption that further health benefits will accrue from regulatory actions to reduce emissions further is being challenged, especially because the incremental costs of additional reductions are potentially high. Since 1980, measurements at thousands of monitoring stations across the country have shown lowered concentrations for the six nationally regulated and tracked principal pollutants (O_3, PM, SO_2, NO_2, lead, CO)—called *criteria pollutants*. According to the EPA's most recent *Trends Report* (EPA 2002), aggregate emissions of all six dropped 25% between 1970 and 2000. Meanwhile, the gross domestic product rose 161%, energy consumption increased 42%, and vehicle miles traveled increased 149%.

This progress has, of course, come at a price. The EPA estimates (EPA 1997) that from 1970 to 1990 the annualized cost of air pollution control was about $25 billion per year—more than $500 billion over 20 years. Further, the EPA estimates (EPA 1999) that when fully implemented the 1990 Amendments to the Clean Air Act will double the annual costs of the Act to an average of more than $50 billion per year. Implementation of the US National Ambient Air Quality Standards (NAAQS) for particulate matter (PM) less than 2.5 μm in aerodynamic diameter ($PM_{2.5}$) and ozone issued in 1997 is anticipated to add billions more dollars. Even as new research findings appear to have strengthened the evidence for health effects, many (including policy makers, legislators, industry, and the public) ask whether we are receiving sufficient value in terms of health benefits for the money spent. Have past efforts to reduce air pollution yielded demonstrable improvements in public health? Will future efforts continue to do so?

If public health is to benefit from reduction of ambient concentrations of targeted pollutants, consideration of accountability needs to start with an assessment of the effectiveness of regulations in reducing emissions followed by assessment of whether emissions reductions do in fact reduce ambient concentrations. Such assessments were proposed by the National Research Council in its 1994 report on ozone control (US National Research Council 1991). This proposal acknowledged the complex and nonlinear path from control strategies directed at reducing emissions to realized changes in ambient ozone concentrations.

SCOPE OF ACCOUNTABILITY

Recent discussions have extended the concept of accountability to health effects, the final link in the chain that begins with air pollution sources. In response to the call for accountability, some national governments and public health agencies have attempted to quantify the past health impact of air quality improvements and to estimate future impact. These attempts have used risk estimates from epidemiologic studies to calculate the adverse impact of air pollution on public health using various measures of disease burden under hypothetical, or counterfactual, air quality scenarios (see, for example, EPA 1997, 1999; Künzli et al 2000).

The term *counterfactual* describes a scenario that does not exist (and is hence counter to fact) that is used to estimate either the extra or avoided burden of disease under the actual conditions of population exposure. In the context of air quality accountability, the counterfactual scenario is that which would have occurred in the absence of regulatory or other action to improve air quality. The counterfactual is designed to be compared with actual patterns of exposure and population health effects that result from the implementation of a particular air quality action or set of actions.

The estimates from counterfactual scenarios have not been extensively validated against studies of the health impact of actual interventions. Direct evidence of effectiveness from population-level data would strengthen the evidentiary basis for regulatory and public health policy and for scientific inference about the health effects of air pollution. This evidence would also reduce the uncertainty associated with these model-based but often-influential calculations. Conversely, evidence of lack of effectiveness could directly inform and improve future regulatory decisions. Recent studies of the health effects of air quality improvement programs implemented over short time periods in Ireland (Clancy et al 2002) and Hong Kong (Hedley et al 2002) provide examples of ways in which the impact of interventions may be directly measurable.

The framework for accountability assessment also needs to be extended beyond targeted pollutants and associated health risks. The measures that are needed to reduce ambient concentrations of major pollutants may have broad consequences, some unintended and unanticipated, which could reduce or increase risks to public health. Wiener (1998) and others have advanced the concept of a so-called portfolio of effects from a regulation; these authors describe the full set of effects that need to be evaluated, not just the intended effects. This full portfolio is clearly important in any overarching accountability assessment. In this Communication we have attempted to note those areas in which these unintended consequences

may occur even as we focus on those areas within accountability that are most within HEI's direct expertise: accountability of regulations or other actions for reducing exposure to targeted pollutants and the intended improvements in public health.

With growing interest in accountability, HEI has been urged by its public-sector and private-sector sponsors to undertake research on the health impact of interventions to improve air quality. The sponsors' interest is shared by others, including the US Congress, various nongovernmental organizations, and the World Health Organization. The proposed research could strengthen the empirical basis for assessment of regulatory interventions to improve air quality. Accountability research is a central component of the *HEI Strategic Plan for the Health Effects of Air Pollution 2000–2005* (HEI 2000). Studies of air quality interventions are challenging to design and conduct, however, which perhaps explains why they have not been more widely attempted. Nonetheless, as the serious health effects of air pollution become better understood and generally accepted (HEI 2001), and governments respond with regulations and other interventions, the need for such studies will become increasingly evident. Indeed, studies to evaluate the consequences of interventions should be part of routine public health surveillance.

This Communication is intended to provide a framework for advancing research on accountability. Chapters 2 through 5 cover relevant concepts and methods, including the data resources for accountability research and the types of studies and methods of analysis that may be informative. The final chapter presents specific recommendations for how such research might proceed. These recommendations are intended to contribute to research planning by HEI and other agencies.

ANTICIPATED CHALLENGES

Research on accountability is in its early stages, but even now considerable challenges in assessing the health impact of air quality regulations can be anticipated. The following challenges need to be considered when data systems are developed and studies are designed for accountability assessment.

Air quality regulations themselves are promulgated to take effect at different times and on multiple governmental levels. Governments pursue a variety of actions to improve air quality from national regulatory programs (such as the US Clean Air Act) to local initiatives (such as conversion of the fuel of municipal bus fleets from diesel to natural gas). Within the context of the US Clean Air Act, research could be designed to measure the health impact of the overall Act, particular titles of the Act (eg, Mobile Sources: Title II), or particular actions taken to implement the provisions of a specific title (eg, reformulated gasoline or natural gas fuel). Therefore, diverse approaches will be needed to evaluate the impact of interventions on human health at national, regional, and local levels and on various time frames.

Measuring the effectiveness of an action to improve air quality in terms of changes in air quality and exposure over time presents considerable challenges, both technical and conceptual. Any study that seeks to provide evidence for the health impact due to regulation-driven air quality improvements will need to establish that the regulations did indeed improve air quality and public health. To determine whether this conclusion is true, both changes in pollutant emissions and the relations among changes in emissions, changes in ambient concentrations, and changes in personal exposure will need to be measured. Estimating baseline or reference pollutant concentrations will require specific attention. To the extent that reduction strategies directed toward one particular pollutant may increase (or decrease) levels of another pollutant, monitoring for other pollutants may be warranted.

Although current US air quality standards are pollutant specific, air pollution is a mixture of pollutants. Researchers can track changes in components of the mixture (eg, ozone or $PM_{2.5}$), but in general they cannot be certain that particular pollutants by themselves are the toxic agents responsible for the effects associated with them. The air pollution mixture complicates interpretation of patterns of changes in health outcomes observed after an intervention aimed at reducing ambient pollution levels. In this Communication we repeatedly visit several examples to illustrate the challenges posed by the mixture as well as possible solutions. These examples include lead and carbon monoxide (which represent single pollutants with specific biologic markers of exposure), and PM and photochemical oxidants (which represent complex mixtures with multiple sources).

The consequences of interventions may extend beyond changes in air quality. Interventions could result in changes in personal activities and behaviors (such as more time engaged in physical activity outdoors) or in economic activities (such as decisions to close facilities and reduce employment) that could in turn affect health. Therefore, the causal pathways from a regulation and its consequences for air quality to a change in its risk to health could be difficult to isolate.

Adverse health effects that may be caused by exposure to air pollution can also be caused by other factors (some changing over the same time periods as air pollution concentrations). For example, prolonged exposure to air

pollution has been associated with increased long-term mortality from cardiovascular diseases in recent epidemiologic studies (Dockery et al 1993; Krewski et al 2000; Hoek et al 2002; Pope et al 2002). But cardiovascular diseases are also related to diet, cigarette smoking, and other factors. In order to estimate the effects of air pollution accurately, epidemiologic studies need to account for these other risk factors. Efforts to track the long-term impact of air quality regulations will also need to consider these possibly confounding factors.

Regulatory interventions to improve air quality, especially large national programs such as the US Clean Air Act, may not immediately affect air quality or public health. Once regulations are instated, changes in pollutant emissions, ambient pollutant concentrations, and human exposure to ambient concentrations may not be immediately or uniformly evident over space and time and may be offset in whole or in part by changes in land use patterns and rates of activity (eg, miles driven). Also, the biological processes of injury that underlie the adverse health effects of air pollution may not directly follow changes in exposure due to regulatory action. The longer the time between promulgation of regulations and their effects, the greater the possibility that other causes of adverse health outcomes may come into play and interfere with measuring the effects of the interventions themselves. The level of enforcement may complicate the analysis by extending the anticipated time between intervention and effect. On the other hand, some interventions may produce relatively rapid changes in air quality and we might expect to be able to observe their effects on some health indicators quite quickly. Rapid changes reduce, but by no means eliminate, the possibility of confounding by other risk factors. Examples of rapid changes, namely the rapid changes in particle levels associated with operation of a steel mill in the Utah Valley and the regulation-mandated changes in fuel use in Ireland and Hong Kong during the 1990s (Clancy et al 2002; Hedley et al 2002), are presented throughout this Communication.

Stakeholder expectations for regulatory accountability are high and may exceed the realistic possibilities of accountability assessment. All stakeholders hope, and some may expect, that health effects research can provide more definitive evidence concerning the efficacy of interventions to improve air quality. HEI intends this Communication to help set forward-looking yet realistic expectations for accountability research in two ways: by identifying current and probable future accountability methods and their limitations, and by identifying opportunities for accountability research.

CONCEPTUAL FRAMEWORKS FOR ASSESSING ACCOUNTABILITY

PUBLIC HEALTH PARADIGM AND ROLE OF SURVEILLANCE

Regulatory programs enacted to improve air quality, such as the US Clean Air Act of 1970 and its subsequent revisions, seek to protect and improve public health. Protection of the public's health is based on a pragmatic paradigm that is implemented sequentially and iteratively, from evidence on the state of the public's health and on factors that adversely affect it to program planning and then to interventions directed at these factors. The process is ongoing and is redirected as surveillance data show whether interventions have proved effective and whether new problems have emerged. This pragmatic approach assumes that public health data can identify problems to be addressed and that the efficacy of interventions can be tracked through surveillance and other evaluation data.

This paradigm has often been applied to help control a wide range of human maladies, including serious injuries and infectious and chronic diseases. The paradigm has long been in use for control of infectious diseases. It helped identify acquired immune deficiency syndrome (AIDS) and its causative agent and has been exemplary in subsequent efforts to control the infection. Appendix 1.A describes a case study using this paradigm to evaluate automobile safety regulations. This case study shows how data can be used to track the consequences of an intervention intended to save lives and to redirect interventions if needed. Other examples of the paradigm's effectiveness can be found among the chronic diseases that became epidemic over the last century, such as coronary heart disease, lung cancer, and diabetes. Massive intervention programs, directed at risk factors such as cigarette smoking, physical inactivity, and untreated hypertension, have been implemented. These interventions address multiple levels: individual lifestyle, community norms, medical practices, and state and federal policies and laws. Program evaluation has been integral to these interventions, although often challenging, because multiple programs may be at work simultaneously with actions directed at multiple levels.

CAUSATION AND PREDICTION OF HEALTH EFFECTS

The concept of causality—specifically, that reduction of one or more risk factors will reduce the risk of disease occurrence—is inherent in the public health paradigms. Designating a risk factor as a cause of disease implies a prediction that reducing the factor will in turn reduce disease

occurrence. Criteria have been developed for concluding that an association is causal for public health purposes. Based in the Henle-Koch postulates elaborated in the 19th century, these criteria are tailored to observational data, which are often paramount for human diseases. The criteria still in use are often referred to as the *Bradford Hill criteria*, in acknowledgment of the British medical statistician (Hill 1965), or as the *Surgeon General's criteria*, in recognition of their elaboration and application in the 1964 Report of the Surgeon General on Smoking and Health (US Public Health Service 1964). The principal criteria include temporality (exposure precedes disease), consistency (the association has been replicated by different investigations and investigators for different populations), strength of association (stronger associations are more likely to be causal), and plausibility (the full scope of evidence is supportive). Unlike the Henle-Koch postulates, these criteria do not require experimental evidence.

For the health effects of air pollution, epidemiologic evidence, based on observation of human populations, has been the foundation of the conclusion that air pollution can cause excess morbidity and mortality. The epidemiologic findings are supported by and interpreted in relation to multiple other lines of evidence: information on physical and chemical characteristics of the pollutants, knowledge of the mechanisms by which the pollutants cause injury, and toxicologic information from in vitro and in vivo assays. The epidemiologic data come from observational studies because randomized controlled trials cannot readily be used to assess the impact of reducing exposure to air pollution. Observational data need to be evaluated carefully to ensure that health effects attributed to air pollution are not actually due to other factors, bias in participant selection, or error in data collection.

For those factors identified as causing disease, epidemiologic methods can be used to estimate the burdens of disease associated with a factor. The attributable burden of disease depends on the proportion of the population exposed and the levels of exposure as well as on the risks associated with the exposure. Estimates of attributable burden compare the observed disease burden to the burden that would have been estimated in a counterfactual scenario. This counterfactual scenario could represent an observed background concentration or a regulatory target value.

FRAMEWORK FOR AIR POLLUTION CONTROL AND ACCOUNTABILITY

CHAIN OF ACCOUNTABILITY

The National Research Council's Committee on Research Priorities for Airborne Particulate Matter report (US National Research Council 1998) sets out a framework for linking air pollution sources to adverse health effects (Figure 1.1). This framework encompasses sources of air pollution (extending from their emissions to the resulting concentrations of pollutants in air), exposure of human populations and the doses individuals receive, and finally the human health response. It can be used to identify indicators for use in accountability assessment and leads to designation of the chain of accountability (Figure 1.2), which parallels Figure 1.1 in its links from sources to health effects. The connections between the links in Figure 1.2 correspond to typical points at which quantitative measures of accountability are possible. This Communication uses the chain of accountability as a framework for each chapter.

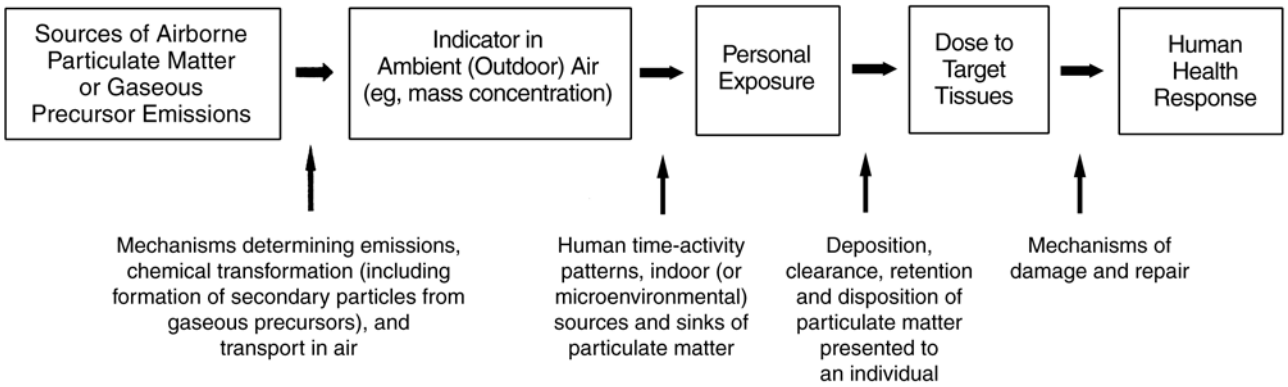


Figure 1.1. A framework for linking air pollution sources to adverse health effects. Adapted with permission from NRC 1998.

This framework fits well with the approach taken in the United States for regulation of principal pollutants. US air pollution regulations for criteria pollutants specify air quality standards (the NAAQS for maximum concentrations of selected air pollutants). Each standard specifies a pollutant, its concentration, the averaging time, and the proportion of time that the standard must be met. The concentration limits are met by source control and related reductions of emissions. This approach inherently assumes that source control and emissions reductions to meet target concentrations will reduce human exposure to targeted pollutants and, subsequently, reduce risks of adverse health effects. Assessing the health impact of programs such as the US Clean Air Act calls for a demonstration that implementation of measures to reach the NAAQS have in fact led to a reduced disease burden, which is the final step in the chain of accountability.

INDICATORS FOR ACCOUNTABILITY

Along the length of the chain of accountability (Figure 1.2) are a number of indicators that can be used to assess the effectiveness of an air quality regulation.

- **Regulatory action**: Have controls on source emissions been put into place?
- **Emissions**: Have the source controls reduced emissions? Have there been unanticipated and untoward consequences? Answers to these questions require evidence about how regulation has changed the practices of emitters relative to what they would have been without regulation and about what changes in emissions have resulted.
- **Ambient air quality**: Have concentrations of air pollutants declined due to source control and emissions reductions? Answers to this question require evidence based on periodic standardized measurement of ambient concentrations of air pollution constituents, such as PM, ozone, and sulfur dioxide.
- **Personal exposure:** Has exposure to air pollution declined? For which groups in the population (particularly as defined with reference to susceptibility to the effects of air pollution)? For instance, have exposures been reduced for disadvantaged people, including racial and ethnic groups, who may experience disproportionately high exposures? The relation between concentration and exposure can be modified by time-activity patterns, and such modification may mean that declines in concentration do not lead to proportional declines in exposure. In assessing changes in exposure, two factors should be

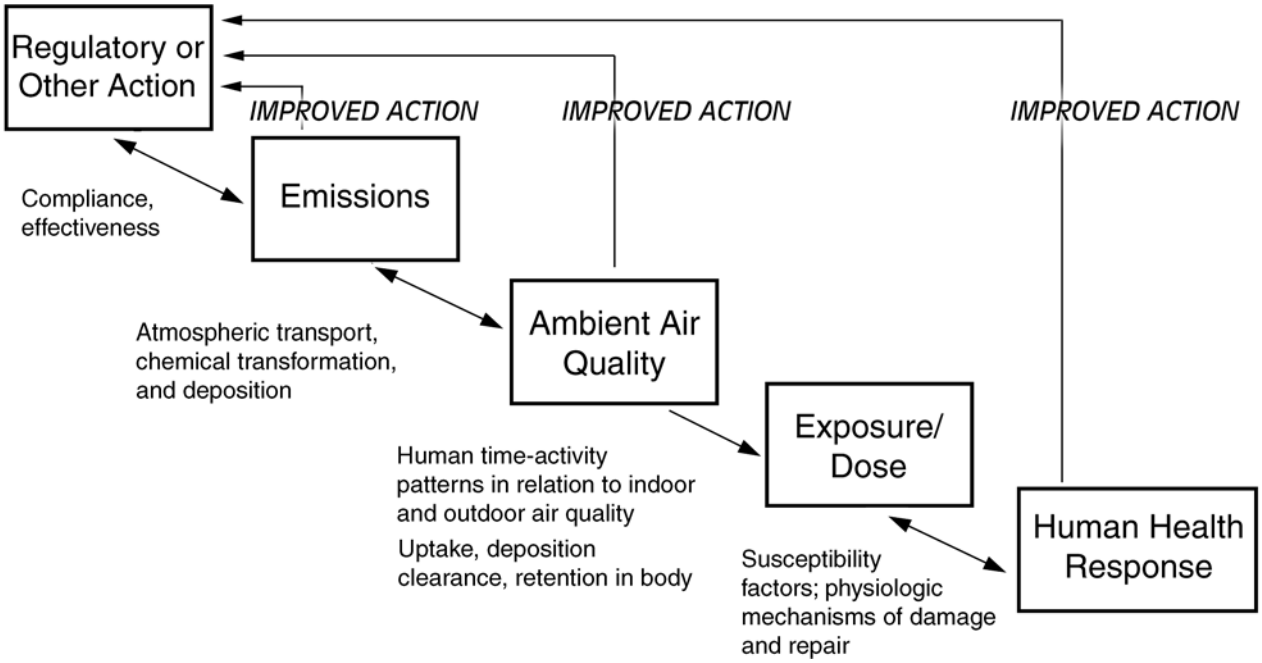


Figure 1.2. Chain of accountability. Each box represents a link between regulatory action and human health response to air pollution. Arrows connecting the links indicate possible directions of influence. Text below the arrows identifies general indices of accountability at that stage. At several stages, knowledge gained from accountability assessment can provide valuable feedback for improving regulatory or other action.

considered: how concentrations have changed as a result of changes in emissions, and how the intervention may have changed the behavior (and thus exposure) of the population(s) it was intended to protect.

- **Dose to target tissues**: Have reductions in exposure led to reductions in dose? The relative consequences of a dose for a susceptible versus a nonsusceptible person may also be relevant.
- **Human health response**: Have health risks declined? This indicator requires evidence about changes in health endpoints that have resulted from changes in exposure. Research must address which health endpoints and measurement techniques are most directly attributable to air pollution exposure and thus would be most useful for accountability assessments as well as how the health endpoints should be defined and characterized for analysis.

In the general framework considered here, implementation of a standard would ideally be followed by comprehensive surveillance for each potential indicator. Although the EPA tracks ambient pollutant concentrations with monitoring networks and requires the development of plans that specify emissions controls, no effort currently connects in a comprehensive fashion the sources at one end of the chain to their adverse health effects at the other.

HEALTH OUTCOMES FOR ACCOUNTABILITY

Linking Air Quality Regulation and Health Outcomes

Regulations designed to improve air quality and protect the public health most frequently target changes in the quantity and nature of emissions (the left side of the chain of accountability). The challenge of demonstrating that the intended consequences have occurred becomes progressively more difficult in the right side of the chain where the possibility increases that other factors are also at play.

If health outcomes were caused only by air pollution, our task would be much simpler. In the real world, however, health outcomes associated with air pollution, such as mortality and morbidity from cardiovascular and respiratory disease, can be caused by factors besides air pollution. Some of these factors have far stronger effects than air pollution. Nonetheless, because most regulatory actions are based on evidence for adverse effects on these health outcomes that are not uniquely associated with air pollution, these same outcomes are the focus for assessing regulatory effectiveness. Assessments may focus on specific outcomes on the basis of other factors, including practicality, public-policy objectives, and availability of research methods and data.

Considering Other Factors Related to Health Outcomes

Any assessment of accountability must also consider factors other than changes in air pollution levels that may be affecting the same health indicators as the intervention of interest. Adverse health effects of air pollution are affected by multiple factors that may determine either incidence or severity of the disease or condition. Exposure to these factors may change contemporaneously with exposure to air pollution, and as a result, isolating consequences of air pollution control from those of other factors may be complicated. For example, morbidity and mortality rates of some major chronic diseases linked to air pollution have changed across recent decades. Asthma prevalence has risen substantially, cardiovascular mortality has dropped sharply, and mortality from chronic obstructive pulmonary disease (COPD) has risen. The prevalence of cigarette smoking has declined. The US population itself has changed, with far larger numbers of older people than previously.

The multifactorial etiology of health-outcome measures poses a challenge to assessment of accountability and will require researchers to develop and implement analytic strategies to separate the consequences of changes in air pollution from those of other factors.

MODELS AND RISK ASSESSMENT

Unable to directly track the consequences of regulation through an accountability chain, the EPA and other organizations often turn to models to predict the possible consequences of regulation. The chain of accountability (Figure 1.2) is itself a model of how changes in emissions may affect public health. Connections within the chain are represented by specific submodels reflecting the relations of emissions from pollution sources with pollutant concentrations in the air, of concentrations with personal exposures, and of exposures with risks to health. Specific source-receptor models, some quite extensive and elaborate, have been developed to estimate atmospheric pollutant concentrations generated by pollution sources. Other models have been developed to estimate personal exposures to air pollutants.

As an alternative to comprehensive direct assessment of progress in meeting the public health objectives of the Clean Air Act, the EPA and other interested parties use risk assessment approaches based on estimating risk attributable to exposure to air pollution. The Agency uses this approach, for example, in estimating the benefits from reduction in levels of criteria pollutants undertaken under the mandate of Section 812 of the Clean Air Act. Modeling can be used to predict the consequences of various source-control strategies

for reduction of emissions and concentrations. For estimating health risks, pollutant concentration is often assumed as a surrogate for exposure and dose, and the disease burden to be prevented by an intervention is estimated, using an attributable risk approach.

Estimating the burden of disease attributable to air pollution requires a model for the relation between pollutant concentration (or exposure or dose) and risk to health. For this purpose, the most frequently used approaches, albeit simplistic, use dose-response relations from observational epidemiologic studies. For example, the EPA has used the dose-response estimates observed in long-term cohort studies to estimate the burden of premature mortality associated with particulate air pollution. Use of such estimates implicitly assumes that exposure-risk relations derived from observational studies will accurately estimate the effects of an intentional intervention. Uncertainty arising from this assumption is viewed as a limitation of the epidemiologic approach to estimating attributable burden of disease and thereby to documenting the public health impact of interventions.

The dose-response relation (more accurately, the exposure-response or concentration-response relation) is most often estimated through epidemiologic study designs based on exploring gradients of health risk across gradients of naturally occurring exposure. For example, mortality and morbidity might be compared across regions with differing air quality or on days with differing concentrations of the pollutants of interest.

On occasion, planned or unplanned interventions may allow observations of health risks under rapidly changing concentrations or exposures. One widely cited example is closing of the Provo, Utah, steel mill for one year in the 1980s (see Pope 1989 and Chapter 4); another is the reduction in traffic and ozone levels during the 1996 Summer Olympic Games in Atlanta (Friedman et al 2001). In the Utah Valley, Pope and colleagues based a series of mortality and morbidity studies on pollution from the steel mill. They examined particle concentrations and characteristics before, during and after the year in which the plant was not operating. Friedman and colleagues analyzed asthma morbidity rates in relation to changes in air pollution levels. The changing air quality in the former Soviet bloc countries has also been tracked and changes in respiratory health of children have been monitored (Heinrich et al 2002).

The Health Effects Institute and other funding agencies are actively soliciting proposals for research that might address other interventions that have led to changes in pollutant concentrations. Effect estimates derived from identified interventions complement evidence from observations of more typical variation in source emissions. Additionally, the consequences of a rapid intervention may be more readily detected along the chain of accountability than more gradual interventions.

GENERATING SCIENTIFIC EVIDENCE FOR ACCOUNTABILITY ASSESSMENT

Research aimed at measuring the health impact of actions to improve air quality should consider each link in the accountability chain (Figure 1.2). Consideration of the whole chain requires involvement of experts from a broad range of disciplines, including public policy, engineering, atmospheric science, exposure assessment, epidemiology, biostatistics, toxicology, and economics. Such collaboration, although frequently advantageous even within the narrower scope of epidemiologic research designed to identify risk factors for disease, is seldom realized in practice.

DATA RESOURCES

Intervention Design and Objectives

In order to assess the impact of a specific regulatory intervention, the design and objectives of the intervention need to be understood. The timing of the intervention and how it was implemented have implications for assessment of its effectiveness. Investigators may require the cooperation of regulatory agencies, and possibly the regulated industries, to obtain this information. Chapters 2 and 3 provide detailed discussions of the conceptual and practical aspects of air quality regulations to be considered in the design of studies to measure the intended health benefits of those regulations.

Air Quality Monitoring

Air quality monitoring networks established to monitor compliance with air quality standards and guidelines provide critical data used to assess the health impact of air quality regulation over time in the United States, Europe, and other locales. In the United States, measurements taken by these networks are modified in response to changes in NAAQS; for example, measurement of $PM_{2.5}$ began in 1999 in support of the new fine particle standards. In addition, new monitoring networks have been established at the state and national levels to obtain more detailed information on physical and chemical characteristics of PM. Information from these new networks will be invaluable for scientific research but may also be quite useful for policy-related research. The use of air quality

monitoring data in accountability assessments is reviewed in Chapter 3.

Assessing Exposure or Dose

Changes in air quality affect health risks of pollutants by altering exposures to, and ultimately doses of, air pollution. Although ambient air pollution levels provide useful surrogates for exposure in epidemiologic research, they do not necessarily reflect the levels of pollution that individuals are actually breathing in the different environments where they spend time. Differences among ambient concentrations, often measured at one or more central locations, and personal exposure reflect variations in pollution levels in both time and space as determined by the time-activity patterns. The relation between ambient concentration and personal exposure varies among the criteria pollutants, such as ozone and PM. In some cases, understanding required to link changes in ambient concentrations to changes in exposure may be derived from ambient measurements and assumptions about their relation to exposure. But more formal modeling of their relation, using additional information about temporal and spatial variation, may be needed. Large-scale national exposure surveys (which can involve individual exposure monitoring) and collection of biological samples may also contribute to research on the health impact of air quality regulations, particularly if repeated over time. Approaches to linking air quality and individual exposure in accountability research are discussed in Chapter 3.

Public Health Data

Data on births, deaths, and hospitalizations have been invaluable resources for epidemiologic research on the health effects of air pollution. These data are collected routinely over time and in a relatively standardized fashion in many countries for administrative and public health purposes. Assessments of these data have made possible the replication of important observations on the impact of air pollution on daily rates of mortality and morbidity in many locations around the world (Steib et al 2002; Cohen et al 2003). Recent evaluations of the health impact of air quality regulations in Ireland and Hong Kong (Clancy et al 2002 and Hedley et al 2002; discussed in Chapter 4) have also used such data.

Public health data will continue to play a key role in strategies for measuring the health impact of air quality regulations, but for such strategies to be effective, the quality of some types of data will need to improve. For example, inaccuracies in cause-specific mortality data are well known to researchers and public health experts. Comprehensive data on the incidence and prevalence of certain health effects to which air pollution contributes, notably asthma, are not generally available, even in developed countries. These issues are discussed further in Chapters 4 and 5.

Study Designs

Actions taken to improve air quality can be viewed as intentional public health interventions in which improved air quality is a treatment applied to an entire population. Viewing actions to improve air quality as population-level interventions makes clear that the design, conduct, and interpretation of studies to measure the health impact of air quality regulations need to be based on well-formulated theories of causal effects and methods for their estimation in epidemiologic research. In practice, however, interventions to improve air quality are not applied at random as in a laboratory experiment. Rather, interventions such as the US Clean Air Act have been implemented over broad areas over extended periods of time. As a result, the same types of observational epidemiologic designs used to measure the adverse effects of air pollution (eg, time-series and cohort studies) have begun to be used to estimate the impact of air quality regulations (Clancy et al 2002; Burnett et al 2003; Hedley et al 2002). Therefore, the relative strengths and limitations of these designs should be considered when used for this purpose. These issues are explored in Chapter 5.

Surveillance Systems

After promulgation of an air quality regulation, events along the entire chain of accountability unfold sequentially over time. For major national programs, the time course of these events may be on the order of years, although it may be considerably shorter for less complex, local interventions. Regardless of the time frame, surveillance is needed to anticipate each event in order to measure the impact of regulations at each link. Surveillance requires collection, maintenance, and coordinated analysis of data: pollutant emissions, ambient concentrations and other determinants or indicators of exposure, and health outcomes presumably caused by exposure to air pollution. In the United States, some of these elements, such as a national system of air quality measurements, already exist in some form. Others do not: particularly, the health outcome data. The effort required to incorporate all of the elements into a surveillance system (including the development of statistical methods to analyze the data) is considerable, even though some work has begun in the United States and Europe (Medina et al 2001; US Centers for Disease Control and Prevention 2003). Chapters 3, 4

and 5 address issues relevant to the design and implementation of surveillance systems.

Data Analysis

Considerable challenges to estimating the health impact of air quality regulations may arise in data analysis. Some of these challenges include data aggregation, confounding, and inadequate statistical power. For example, recent epidemiologic studies on air pollution (including the Dublin and Hong Kong studies of air quality improvements [Clancy et al 2002; Hedley et al 2002]) have used data aggregated over time (eg, yearly mortality rates) and space (city-wide or regional data). Analyses of such aggregated data need to be interpreted in terms of effects at the level of the individual. Such interpretation depends on the careful specification of the elements of statistical models, especially those dealing with risk factors other than air pollution (Morgenstern 1998).

Interventions to improve air quality may also change behavior within the target human population (see also Chapter 3). Because the interventions are not assigned at random, any behavioral changes not addressed in data analysis could obscure the true relation between improved air quality and health. Accounting for such changes requires the use of statistical models for such behaviors to adjust study results, using techniques such as causal modeling and propensity scores. Issues of data analysis are discussed in Chapter 5.

INTERPRETING ACCOUNTABILITY ASSESSMENTS

Need for a Counterfactual Scenario

Accountability assessment inherently involves a comparison between the links in the chain as they actually occur and their corresponding counterfactual scenarios: how they would have been, without the changes introduced by a regulation or other control measure. The anchoring point is, appropriately, the time at which the control measure is implemented. If relatively immediate consequences are anticipated, comparison with the state of the chain of accountability at that moment will be informative. If the consequences are to be delayed, however, scenarios other than this baseline condition may be appropriate when sources and control measures change over time. For example, comparison of accountability indicators measured in 2017 with baseline conditions in 1997 may be uninformative if the sources have changed substantially for reasons other than the control strategy (eg, some polluting processes may no longer be operative for primarily economic reasons).

Model Assumptions and Related Uncertainties

Models have been used to estimate the impact of air quality regulations in the past and will continue to be necessary in the future. The estimated benefits derive in part from the assumptions made in the models that also introduce uncertainties. Uncertainties are unavoidable, but their impact on interpretation of study data can be mitigated when attention is paid to describing and quantifying their magnitude.

Analysts will need to describe and, if possible, quantify at least two types of uncertainty to which their estimates will be subject. Stochastic, or sampling, uncertainty is the result of imprecision due to inadequate study sample size. The much more substantial, and difficult to quantify, uncertainty results from not knowing the correct statistical models to describe the data. Statisticians have formal methods for quantifying stochastic uncertainty once a model is chosen for the data. But sensitivity analyses, based on carefully chosen scenarios, undoubtedly will be needed to quantify uncertainty about the correct model for the data.

Biomedical Considerations

Understanding the natural history of the disease process is necessary to any assessment of an implemented regulation. Coronary artery disease and COPD are infrequently manifested before 50 years of age although the biological processes started decades earlier. This timing of onset contrasts with that for other chronic respiratory diseases, such as asthma, for which onset is actually highest in the first 10 years of life. Thus, study of diseases with long incubation phases would benefit from developing and identifying sensitive and stable biomarkers that reliably predict disease occurrence, whereas for diseases that manifest in childhood, measurement of disease incidence itself might be feasible.

Some biomedical factors could limit detection of health benefits due to a reduction in air pollution:

- Long induction time for exposure and disease. The long time period between exposure to air pollution and manifestation of disease required to detect a reduction in the incidence of COPD or cardiac disease may prohibit direct study. Benefits that are actually observed might, therefore, appear less or more than those predicted from epidemiologic observations.
- Other risk factors for cardiovascular and respiratory disease. Other factors contribute to the development of cardiovascular and respiratory disease (eg, cigarette smoke, workplace exposures, indoor air pollution, and diet) via mechanisms that may be similar to those associated with the effects of some outdoor air pollutants. Exposure to these factors, which may well be

more potent determinants of disease than outdoor air pollution, may complicate our ability to identify the impact of changes in outdoor air pollution.

- Changes in other environmental exposures concurrent with air pollution regulations. Reductions in tobacco use or increases in workplace dust exposures may occur at the same time as concentrations of regulated pollutants change. The interactions between changes due to regulations and changes due to other factors may be complex and not simply additive.
- Changes in biomedical technology and medical treatment. These changes, such as the introduction of new medications, may alter the natural history of a disease and complicate evaluation of the health impact over time. For example, the natural history of coronary heart disease has been continually changing over recent decades as surgical approaches and nonsurgical revascularization procedures improve.

Meeting the Challenge

Observational studies have been described as a blunt instrument with inherent limitations that preclude finding definitive answers about the effects of exposure to low levels of environmental pollution. Recently, however, epidemiology's evident success in identifying the hazards of exposure to air pollution even at the relatively low levels now prevalent in developed countries has raised expectations that the health impact of air quality regulations can be measured with some accuracy.

Consideration of factors outlined above that might determine the limits of detecting the health impact of air quality regulations is important. Actual health impact may be less than or greater than those anticipated from the existing epidemiologic evidence. The models used to generate the epidemiologic evidence, although reasonable summaries of data, may not reflect accurately the effects of other factors that could be involved with regulatory action.

Whether and to what extent observational study designs can meet these challenges and current expectations is the task before us. The need is clear, however, to measure of the health impact of air quality regulations and other actions and to improve the evidentiary basis for assessing the effectiveness of those actions. Although the inherent challenges are well documented in the chapters that follow, recent advances in data collection and analytic techniques provide an unprecedented opportunity to test and advance our ability to better support efforts to assess the impact of air quality actions in the future. To this end, this Communication is intended to advance the concept of accountability and to foster development of accountability methods and research throughout the scientific and policy communities.

REFERENCES

American Thoracic Society. 2000. What constitutes an adverse health effect of air pollution? Am J Respir Crit Care Med 161:665–673.

Bascom R, Bromberg PA, Costa DA, Devlin R, Dockery DW, Frampton MW, Lambert W, Samet JM, Speizer FE, Utell M (Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society). 1996a. Health effects of outdoor air pollution, Part 1. Am J Respir Crit Care Med 153:3–50.

Bascom R, Bromberg PA, Costa DL, Devlin R, Dockery DW, Frampton MW, Lambert W, Samet JM, Speizer FE, Utell M (Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society). 1996b. Health effects of outdoor air pollution, Part 2. Am J Respir Crit Care Med 153:477–498.

Burnett RT, Cakmak S, Bartlett S, Stieb D, Jessiman B, Raizenne M, Blagden P, Brook JR, Samson PR, Dann T. 2003. Measuring progress in the management of ambient air quality: The case for population health. J Toxicol Environ Health. In press.

Centers for Disease Control and Prevention (US). 2003. Environmental public health tracking program (last updated 6/24/03). *www.cdc.gov/nceh/tracking/default.htm*. Accessed 7/21/03.

Clancy L, Goodman P, Sinclair H, Dockery DW. 2002. Effect of air-pollution control on death rates in Dublin, Ireland: An intervention study. Lancet 360:1210–1214.

Cohen AJ, Anderson HR, Ostro B, Pandey KD, Krzyzanowski M, Kuenzli N, Gutschmidt K, Pope CA, Romieu I, Samet JM, Smith K. 2003. Mortality impacts of urban air pollution. In: Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors (Ezzati M, Lopez AD, Rodgers A, Murray CJL, eds). World Health Organization, Geneva, Switzerland. In press.

Cohen J, Brion G, Haines J. 1990. Review of the National Ambient Air Quality Standards for Lead: Assessment of Scientific and Technical Information. EPA 450/2-89/022. NTIS PB91-206185. US Environmental Protection Agency, Research Triangle Park NC. Available from *www.ntis.gov*.

Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG Jr, Speizer FE. 1993. An association between air pollution and mortality in six US cities. N Engl J Med 329:1753–1759.

Environmental Protection Agency (US). 1997. The Benefits and Costs of the Clean Air Act, 1970 to 1990. EPA-410-R-97-002. Office of Air and Radiation, Washington DC.

Environmental Protection Agency (US). 1999. The Benefits and Costs of the Clean Air Act 1990 to 2010: EPA Report to Congress. EPA-410-R-99-001. Office of Air and Radiation, Washington DC.

Environmental Protection Agency (US). 2002. Latest Findings on National Air Quality: 2001 Status and Trends. EPA 454/K-02-001. Office of Air Quality Planning and Standards, Research Triangle Park NC. Available from *www.epa.gov/air/aqtrnd01/*.

Friedman MS, Powell KE, Hutwagner L, Graham LM, Teague WG. 2001. Impact of changes in transportation and commuting behaviors during the 1996 Summer Olympic Games in Atlanta on air quality and childhood asthma. JAMA 285:897–905.

Health Effects Institute. 2000. HEI Strategic Plan for the Health Effects of Air Pollution 2000–2005. Health Effects Institute, Cambridge MA.

Health Effects Institute. 2001. Airborne Particles and Health: HEI Epidemiologic Evidence. HEI Perspectives. Health Effects Institute, Cambridge MA.

Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. 2002. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: An intervention study. Lancet 360:1646–1652.

Heinrich J, Hoelscher B, Frye C, Meyer I, Pitz M, Cyrys J, Wjst M, Neas L, Wichmann H-E. 2002. Improved air quality in reunified Germany and decreases in respiratory symptoms. Epidemiology 13:394–401.

Hill AB. 1965. The environment and disease: Association or causation. Proc R Soc Med 58:295–300.

Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA. 2002. Association between mortality and indicators of traffic-related air pollution in the Netherlands: A cohort study. Lancet 360:1203–1209.

Krewski D, Burnett RT, Goldberg MS, Hoover K, Siemiatycki J, Jerrett M, Abrahamowicz M, White WH. 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.

Künzli N, Kaiser R, Medina S, Studnicka M, Chanel O, Filliger P, Herry M, Horak F Jr, Puybonnieux-Texier V, Quenel P, Schneider J, Seethaler R, Vergnaud JC, Sommer H. 2000. Public-health impact of outdoor and traffic-related air pollution: A European assessment. Lancet 356:795–801.

Medina S, Plasència A, Artazcoz L, Quénel P, Katsouyanni K, Mücke HG, De Saeger E, Krzyzanowsky M, Schwartz J, and the contributing members of the APHEIS group. 2001. APHEIS: Monitoring the Effects of Air Pollution on Public Health in Europe. Scientific Report 1999–2000. Institut de Veille Sanitaire, Saint-Maurice, France.

Morgenstern H. 1998. Ecologic studies. In: Modern Epidemiology, 2nd edition (Rothman KJ, Greenland S, eds). Lippincott-Raven Publishers, Philadelphia PA.

National Research Council (US). 1991. Rethinking the Ozone Problem in Urban and Regional Air Pollution. National Academy Press, Washington DC.

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

Pope CA III. 1989. Respiratory disease associated with community air pollution and a steel mill, Utah Valley. Am J Public Health 79:623–628.

Pope CA III, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287:1132–1141.

Public Health Service (US). 1964. Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service. US Department of Health, Education, and Welfare, Washington DC.

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 Manage Assoc 52:470–484.

Wiener J. 1998. Managing the iatrogenic risks of risk management. Risk Health Safety Environ 9:39–82.

APPENDIX 1.A. Case Study: Airbags and Front-Seat Motor Vehicle Occupant Safety in the United States

By Maria Segui-Gomez, MD, ScD
Johns Hopkins Bloomberg School of Public Health

The introduction of motor vehicles to the United States has been both beneficial and detrimental. Motor vehicles led to increased mobility and improved the economy, but they also increased the risks of death or injury in crashes. As the numbers of vehicles on the road and miles traveled increased, so did the number of individuals suffering temporary or permanent health losses. Toward the end of the 1960s, some 52,000 US residents died each year because of cars; many more sustained temporal or permanent physical impairments (Bonnie et al 1999). Motor vehicle crashes became the leading cause of death between the ages of 1 and 44 years and the leading cause of years of potential life lost.

By the late 1960s, a group of physicians, public health professionals, and engineers had introduced the concept of crash and injury preventability and established the field of research called *injury control*. As part of their efforts to reduce the burden of motor-vehicle crashes, they advocated for a federal agency to be responsible for motor-vehicle safety, a regulatory body to increase the safety of motor vehicles, and legislation to establish surveillance systems.

SURVEILLANCE SYSTEMS AND DATABASES

As a result, a number of surveillance systems were developed. One of the first, the Special Crash Investigation, was established in 1972 to provide in-depth engineering data on specific crashes. Population-based surveillance systems were the next development, starting with the Fatality Analysis Sampling System in 1975, which generated a census of all police-reported crashes in the United States for which one or more people died within 30 days. This data system comprises some 50 data elements: mainly data from accident reports completed by police at the scenes of crashes, with supplementary vehicle registration and drivers' license data.

The National Automotive Sampling System was established in 1978 as a probability sample of police-reported crashes after which at least one vehicle was towed. A subset system, the General Estimates System, involves annual sampling of some 50,000 crashes, for which data from police accident records only are included (for a total of 90 data elements). A second subset, the Crashworthiness Data System, focuses on crashes involving passenger vehicles. An in-depth investigation is conducted for each of 5000 crashes sampled annually. Detailed crash, vehicle, and medical information are collected in the approximately 400-element database.

All these databases are publicly available (US National Highway Traffic Safety Administration 2003). They have enabled researchers to track the magnitude of the motor vehicle crash problem, investigate possible risk factors, and evaluate preventive measures.

AIRBAGS VERSUS SAFETY BELTS

Research conducted with the help of these national databases and others confirmed that head injuries due to contact with the windshield and frontal panel, including the steering wheel, during frontal crashes are the most dangerous types of crash injury in terms of their lethal and severe nonlethal consequences (Graham 1989). Safety belts and airbags were the most helpful devices available to minimize these injuries. Safety belts had been introduced into motor vehicles in the early 1960s, but their efficacy was hampered in the real world by their low use rates: at the time, fewer than 10% of motor vehicle front seat occupants used their safety belts (US Department of Transportation 1997).

Airbags had the potential to be more protective than safety belts because they do not require any action on the part of occupants. Designed in the late 1940s, airbags are inflatable cushions that automatically deploy during collision. Whereas safety belts protect by restraining the occupant and reducing their likelihood of hitting the dashboard, steering wheel, or windshield during the deceleration associated with a crash, frontal airbags protect by providing a soft cushion against which the occupant comes to rest. However, the necessarily quick deployment of the bags releases large amounts of energy, which could injure vehicle occupants.

Whether vehicles in the United States should have airbags (instead of, or in addition to, safety belts) became one of the longest and most intense debates among professionals concerned with motor vehicles. Consumer protection organizations and injury control specialists, including agents of the National Highway Traffic Safety Administration, favored the passive protection of airbags because they were skeptical of the potential to considerably increase safety belt use. In contrast, manufacturers were concerned about the added costs associated with airbags. In addition, they and some safety experts expressed concerns about possible airbag-related injuries (Graham 1989).

The 20-year debate culminated in 1991 with legislation requiring all cars sold in the United States to be equipped with frontal airbags for drivers and front-seat passengers, effective model year 1997 (US Department of Transportation 2001). (To date, the United States is the only country

to have passed such legislation.) This legislation was passed in spite of some experimental evidence presented by manufacturers that confirmed possible damage to out-of-position or smaller occupants. After enactment, manufacturers introduced frontal airbags very rapidly into the fleet. The percentage of passenger cars with driver-side airbag systems increased from 23% in model year 1990 to 100% in model year 1995.

AIRBAG DESIGN AND PERFORMANCE

Airbags had to be designed to protect potential occupants from sustaining forces greater than a prespecified level when exposed to a frontal crash of specified severity. Performance standards were developed early in the debate on the basis of experimental and real-world crash data. By 1991, frontal airbag systems were expected to reduce the probability of death for a driver by 22% (US Department of Transportation 1996). The effects on nonfatal injuries and front-seat passengers were assumed to be similar.

Soon after airbag systems were implemented, however, the need for performance monitoring became clear. Numerous researchers began monitoring performance primarily by using the databases created in the 1970s. In mid 1996, several children died from airbag deployment during otherwise minor crashes. Even before these highly publicized accidents occurred, several reports of airbag-induced injuries had already been published in the medical literature, showing that certain specific injuries (ie, blunt or hyperextension-related) occur when an occupant interacts with an inflating bag. By the end of 1996, the Special Crash Investigations system confirmed that 39 child passengers, 2 adult passengers, and 31 drivers (mostly female, several more than 60 years old) died from these injuries (US National Center for Statistics and Analysis 2003).

Concurrent statistical evaluations of airbag performance reported mixed results. Fatalities were reduced by 34% among drivers without safety belts in frontal crashes, by 11% for drivers and front seat passengers in all crashes, regardless of safety belt use, and by 16% in cases of serious head injuries among drivers in all types of crashes (US Department of Transportation 2001). Less favorable were reports of a net 63% increase in child fatalities (Graham et al 1998) and a 14% increase in serious upper-extremity injuries among drivers (US Department of Transportation 2001). Passenger-side airbags were less cost effective than driver-side airbags, even when adult occupants only were included in the analyses (Graham et al 1997; Larkin et al 1998).

These unintended negative consequences of airbags reignited the debate about their use, which involved motor vehicle manufacturers, airbag system manufacturers, consumer representatives, government safety officials, lawyers, victims' associations, and others. Because airbags were inducing some injuries, had they been introduced carelessly? Had the airbags been designed to perform too aggressively or to deploy too frequently or in crash conditions that were not sufficiently severe? On the other hand, because overall airbags were protective, had they not been introduced soon enough?

AIRBAG IMPROVEMENTS

The renewed debate led to remedial measures. Manufacturers were requested to send letters to owners of airbag-equipped vehicles, warning them of the dangers of airbags, and to add three warning stickers in the interior of new vehicles. Permission was granted to install a switch to turn off the airbag systems, which was an illegal operation prior to this time. Legislation requiring proper restraint of child passengers was strengthened throughout the country; some states started to require children below certain ages to occupy rear seats only. In 1997, a final rule on depowering, or allowing less powerful frontal airbags to be used, was introduced. Lastly, performance standards required for approval of the systems were revised in 2000.

Effective model year 1998, many US manufacturers introduced changes in their frontal airbag systems (US Department of Transportation 2001). Investigation of Special Crash Investigation data on the performance of these new systems has revealed no new airbag-induced deaths since 2000. Preliminary evaluations using the larger, population-based datasets suggest reductions in airbag-induced injuries and no increases in the types of injuries the airbags were intended to protect against. A recently formed coalition between automotive manufacturers and government has led to an expansion of the National Automotive Sampling System Crashworthiness Data System so that more cases of vehicles with newer airbag systems can be collected and analyses can be done sooner. Even then, it will take a few more years before a formal evaluation can be made.

The history of airbag regulation in the United States illustrates the interaction between political, public health, and private industry forces that occurs during adoption of public health measures. It also illustrates several needs for accountability of such measures: specific monitoring of the measure's performance, appropriate sources of data for monitoring, and vigilance for unexpected outcomes.

REFERENCES FOR APPENDIX 1.A

Bonnie RJ, Fulco CE, Liverman CT, eds. 1999. Reducing the Burden of Injury: Advancing Prevention and Treatment. Committee on Injury Prevention and Control, Division of

Health Promotion and Disease Prevention, Institute of Medicine. National Academy Press, Washington DC.

Department of Transportation (US). 1996. Third Report to Congress: Effectiveness of Occupant Protection Systems and Their Use. National Highway Traffic Safety Administration, Washington DC.

Department of Transportation (US). 1997. America's experience with seat belt and child seat use. In: Presidential Initiative for Increasing Seat Belt Use Nationwide; Recommendations from the Secretary of Transportation. National Highway Traffic Safety Administration, Washington DC. Available from *www.nhtsa.dot.gov/people/injury/airbags/presbelt/america_seatbelt.html*.

Department of Transportation (US). 2001. Fifth/Sixth Report to Congress: Effectiveness of Occupant Protection Systems and Their Use. DOT HS 809 442. National Highway Traffic Safety Administration, Washington DC.

Graham JD. 1989. Auto Safety: Assessing America's Performance. Auburn House Publishing Co, Dover MA.

Graham JD, Goldie SJ, Segui-Gomez M, Thompson KM, Nelson T, Glass R, Simpson A, Woerner LG. 1998. Reducing risks to children in vehicles with passenger airbags. Pediatrics 102:e3. Available from *www.pediatrics.org/cgi/content/full/102/1/e3*.

Graham JD, Thompson KM, Goldie SJ, Segui-Gomez M, Weinstein MC. 1997. The cost-effectiveness of air bags by seating position. JAMA 278:1418–1425.

Larkin GL, Weber JE, Graham JD, Thomson KM, Goldie SJ, Segui-Gomez M, Weinstein MC. 1998. Cost-effectiveness of air bags in motor vehicles. JAMA 279:506–507.

National Center for Statistics and Analysis. Special Crash Investigations (US). 2003. Children fatally injured by PAB, normalized for a 12-month period. *www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/SCI/2Q_2003/HTML/PABChild/NTCC.htm*. Accessed 7/01/03.

National Highway Traffic Safety Administration (US). 2003. National Center for Statistics and Analysis: Available information. *www-nrd.nhtsa.dot.gov/departments/nrd-30/ncsa/TextVer/index.html*. Accessed 7/01/03.

ABBREVIATIONS AND OTHER TERMS

CO	carbon monoxide
COPD	chronic obstructive pulmonary disease
EPA	Environmental Protection Agency (US)
NAAQS	National Ambient Air Quality Standard(s) (US)
NO_2	nitrogen dioxide
O_3	ozone
PM	particulate matter
$PM_{2.5}$	PM less than 2.5 μm in aerodynamic diameter
SO_2	sulfur dioxide

Chapter 2. STATE OF THE ART

INTRODUCTION

Levels of air pollution fell in many industrialized nations over the last decades of the 20th century. The likelihood that further regulatory action to reduce air pollution will produce health gains is being challenged, however, especially as the incremental costs of additional reductions may be high. Even as new research suggests that continued regulatory actions to protect public health may be necessary, some government leaders and regulatory critics are asking whether the additional money spent will sufficiently benefit the public. This chapter presents the historical context in which this debate has developed, drawing on experiences in the United States and the United Kingdom, two countries with the longest histories of clean-air regulation.

In the early days of air-pollution control, poor air quality in major cities, such as Los Angeles and London, was frequently visible to the naked eye. Because the health effects of air pollution were so overwhelming, questions about regulatory accountability were not a major concern. In fact, this air pollution was considered a public health emergency.

By the 1970s, however, concerns about high and possibly unjustified costs of air pollution control began to be raised, and calls for regulatory accountability are now an important part of the political landscape. Over the past decade, for example, the US Congress enacted several statutes designed to develop better information about regulatory outcomes and to compare these outcomes to the effort expended to reach them.

This chapter describes the social, political and regulatory contexts for the current interest in accountability of air pollution control regulation using the United States and United Kingdom as case studies. The next three sections outline the history of major laws and regulations adopted to address air pollution and describes a growing interest in regulatory accountability. Both nations have long histories of air quality regulation and decades-long trends of improving air quality. Also in those sections are brief reviews of the most recent analyses of the health benefits and costs of the US and UK Clean Air Acts. Although they are not a focus of this Communication, cost-benefit and cost-effectiveness analyses are relevant to decision making and are reviewed briefly in this chapter. These considerations extend accountability assessments by overlaying cost considerations on them. Emerging European and Canadian Accountability describes recent efforts in Europe and North America to use epidemiologic time-series studies to measure and track the health effects of air quality regulations. Finally, in the Conclusions, the strengths and limitations of current approaches to measuring the health impact of actions to improve air quality are discussed.

EARLY AIR POLLUTION PROGRAMS AND GROWING INTEREST IN ACCOUNTABILITY

Early concerns about air pollution—mostly focused on dense smoke and soot from coal combustion—date back for centuries. In London, coal burning was a focus of regulation as early as the 13th century (Brimblecombe 1987; Brimblecombe 1999). In 1881, Chicago adopted the first local air pollution control regulation in the United States, declaring emissions of dense smoke from any smokestack (including on boats and locomotives) to be a public nuisance. In the early 20th century, similar regulations were adopted in industrial areas in the United States and elsewhere.

By the mid 20th century, several major public health problems raised public awareness and concern about the connection between air pollution and public health. During the 1940s, photochemical smog in Los Angeles was first recognized. In December 1952, the so-called great London smog caused at least 3000 excess deaths in a week and as many as 12,000 over the following three months (UK Ministry of Health 1954; Brimblecombe 1987; Bell and Davis 2001). Finally, a less well-known air pollution incident occurred in October 1948 in Donora, a small mill town in western Pennsylvania where a higher proportion

This chapter is one part of Health Effects Institute Communication 11, *Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research,* which also includes five other chapters and an Executive Summary of the project.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R82811201 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.

of the population perished in the first week than in London smog (Schrenk et al 1949; Ciocco and Thompson 1961).

These air pollution problems prompted strong local responses in a number of areas. In 1947, California adopted the first statewide Air Pollution Control Act in the United States, which became a model for later federal legislation. After the report of the Beaver Committee in the United Kingdom, which established coal smoke as the cause of air pollution in London, the UK Clean Air Act of 1956 was adopted. This legislation was followed in the 1950s and 1960s by initial national air quality legislation in the United States.

Since these early actions, major efforts to reduce air pollution have been taken in many countries. Progress toward cleaner air has been accompanied by a trend toward greater scrutiny of the benefits and costs of such efforts. This chapter reviews this trend in the United States, United Kingdom, and other countries and regions.

THE US EXPERIENCE[1]

HISTORY OF AIR QUALITY ISSUES AND CONTROL

The push for federal clean air legislation in the United States began in the early 1950s. The first federal statute, the Air Pollution Control Act of 1955, focused exclusively on financial and technical support for emerging state programs and on research. The Clean Air Act of 1963 authorized establishment of air quality criteria. Because these early efforts lacked federal enforcement authority, however, they were little more than voluntary programs and had limited influence.

The Clean Air Act Amendments of 1970 strengthened the federal role significantly and established the fundamental structure and policies that remain in effect today, thirty years later after several rounds of Congressional revisions. The US Environmental Protection Agency (EPA*) was set up with the authority to establish air quality standards, the *National Ambient Air Quality Standards* (NAAQS), and the states were allowed to meet the standards by adopting measures under so-called state implementation plans (SIPs). For specific pollutants, the Clean Air Act directs the EPA administrator to set primary ambient air quality standards "allowing an adequate margin of safety… [to] protect the public health," regardless of economic or technical feasibility (American Thoracic Society 2000).

The challenge of meeting the initial NAAQS was seriously underestimated, and many areas of the country missed the deadlines established by Congress. The 1977 Clean Air Act Amendments extended the compliance schedule to 1982, with extensions until 1987 in nonattainment areas, and added new control requirements for both new and existing sources. Notwithstanding these efforts, however, air pollution in many areas (eg, Los Angeles) remained a serious problem.

The 1990 Amendments to the Clean Air Act mandated the next and, so far, the latest round of federally mandated control requirements. Based largely on concerns about acid rain, these Amendments contain provisions to reduce emissions of sulfur dioxide (SO_2) as well as oxides of nitrogen (NO_x). A major new emissions permit system (which allowed emissions to be traded) established a nationwide limit of 8.9 million tons of sulfur (SO_2) per year, representing a reduction of approximately 50%. Overall, the 1990 amendments increased stringency of the provisions and of enforcement of the Clean Air Act.

A key feature of the US Clean Air Act is its emphasis on using scientific evidence in setting standards. The EPA is required to review the scientific basis for the NAAQS every five years to assure that the standards provide adequate protection of public health. US law also requires EPA to consult with the Clean Air Scientific Advisory Committee, part of the EPA's Science Advisory Board, a Congressionally mandated group of independent scientific and technical experts drawn from academia, industry, and the states. Since 1980, EPA has completed reviews of six NAAQS. Currently, EPA is increasingly focusing its efforts on tracking and controlling two of the criteria pollutants that are key components of smog and haze: ground-level ozone and fine particles (ie, particulate matter [PM] less than 2.5 μm in aerodynamic diameter [$PM_{2.5}$]).

EVOLUTION TOWARD COST-BENEFIT ANALYSIS

The roots of the current regulatory accountability debate probably lie in the earliest days of the Clean Air Act. At the time when Congress was considering the 1970 Amendments, strong sentiments were expressed about the potentially high costs of air pollution control. In May 1971, a year after passage of the Act, President Nixon instituted the so-called quality of life reviews: All proposed federal rule makings involving health, safety, and environmental issues (including those issued by the EPA) would be reviewed by other relevant agencies in a process coordinated by the President's Office of Management and Budget (OMB). The stated purpose of these reviews was to ensure that economic development and fiscal concerns were accorded sufficient consideration in the process of writing regulations.

1. Some of the material in this section is drawn from Krupnick and Morgenstern (2002).

* A list of abbreviations and other terms appears at the end of this chapter.

Environmentalists claimed that this review process violated the Clean Air Act because it enabled the OMB director, rather than the EPA administrator, to make the final decision on the content of air quality regulations. EPA administrator William Ruckelshaus maintained, however, that while he paid close attention to the recommendations of other agencies, the final decision in all cases was his alone.

In 1978, the Carter administration inaugurated its own version of Executive Branch review by creating the Regulatory Analysis Review Group (RARG) by Executive Order (EO) 12044 (43 Fed Regist 12661 [3/24/78]). This EO required all federal agencies to conduct "economic impact analyses" of any regulation they were proposing that would have "substantial economic impact (ie, a cost greater than 100 million dollars a year)."

Despite the Executive Branch reviews instituted by Presidents Nixon and Carter, by the late 1970s concern was growing that the nation's pollution control efforts were fueling an accelerating inflation rate while yielding only modest health and welfare benefits. Within a month of taking office, President Ronald Reagan issued an EO that substantially increased the requirements for economic analysis and executive branch review of regulations. Specifically, EO 12291 (46 Fed Regist 13193 [2/19/81]) required EPA and other regulatory agencies to prepare regulatory impact analyses (RIAs) on all major regulations and to submit them to OMB before taking regulatory action. Importantly, this EO required that "the potential benefits outweigh the costs," and that "of all of the alternative approaches to the given regulatory objective, the proposed regulation will maximize the net benefits to society."

In September 1993, President Clinton issued EO 12866 (58 Fed Regist 51735 [10/4/93]), replacing the stipulation that benefits "outweigh" costs with "a reasoned determination that the benefits of the intended regulation justify its costs." Further, agencies are to "include both quantifiable measures (to the fullest extent that these can be usefully estimated) and qualitative measures of costs and benefits that are difficult to quantify" and to "select those approaches that maximize net benefits (including potential economic, environmental, public health, and safety, and other advantages; distributive impact; and equity) unless a statute requires another regulatory approach." This formulation endorses benefit-cost analysis as a tool for helping to choose among alternative regulatory (and nonregulatory) options while not requiring that benefits quantitatively exceed costs. The notion of maximizing equity was and is, however, difficult to define. Under President Clinton's EO, agencies are required to make explicit the implications of policy choices but are not forced to adhere to any rigid decision-making formula.

Throughout the history of the Clean Air Act, RIAs (and other economic analyses) have been prepared for dozens of major, air regulations. However, a comprehensive approach for assessing the overall health, welfare, ecologic and economic benefits of EPA's Clean Air Act programs or a cost-benefit comparison was not specifically mandated. Consequently, Congress added Section 812 to the Clean Air Act in 1990, a section that required EPA to conduct periodic, scientifically reviewed studies of how the Act affected the "public health, economy, and the environment of the United States." Section 812 further stipulated that the assessments reflect central tendency or best estimate assumptions rather than the conservative assumptions sometimes used in setting standards.

To date, EPA has published two studies as required under Section 812, referred to as the *Section 812 studies*. The so-called retrospective study, *The Benefits and Costs of the Clean Air Act, 1970 to 1990* (EPA 1997a), covers the first 20 years of the strengthened Clean Air Act. The prospective study, *The Benefits and Costs of the Clean Air Act, 1990 to 2010* (EPA 1999), the so-called prospective study, analyzes the changes embodied in the 1990 Amendments. The study results are summarized and their key assumptions are discussed in detail in the next section.

The US Congress showed more interest in improving the efficiency and efficacy of federal programs with the Government Performance and Results Act of 1993 (GPRA). The GPRA covers all major governmental programs, including those stemming from the Clean Air Act. The premise underlying GPRA is that policymaking at all levels of government is handicapped by imprecise program goals and inadequate information on program performance. GPRA requires agencies to conduct regular and systematic measurement and reporting of program performance compared with preestablished goals. GPRA is designed to provide the information necessary to strengthen program management, to objectively evaluate program performance, and to set realistic, measurable goals for future performance.

ESTIMATING CLEAN AIR ACT BENEFITS AND COSTS

Design

EPA's retrospective and prospective studies were designed to examine a specific set of policies enacted by Congress in 1970 and implemented by the EPA. Both studies were conducted at a highly aggregated, economy-wide level. The retrospective study did not estimate benefits or costs of individual regulations, pollutants, or subcategories (eg, stationary vs mobile sources) of the federal air pollution control program. The prospective study estimated costs but not benefits by title of the Clean Air Act Amendments.

Both studies reflected the state of the art when they were prepared, and they are probably the most intensive and expensive public health and economic analyses ever done by EPA. Under the auspices of the Agency's Science Advisory Board, both studies were scrutinized throughout their decade-long preparation by at least three expert committees of outside economists, air quality modelers, epidemiologists, and other health experts.

Key Assumptions and Approaches

From a policy perspective, an analysis of total benefits and total costs necessarily simplifies a complex issue. This type of aggregate analysis would not be used to decide whether to continue existing programs. More relevant to policy are the benefits and costs of specific regulations and marginal changes to each regulation. In carrying out its highly aggregated analyses, the EPA offered the principal rationale that while costs can be attributed to individual regulations or programs, benefits are analyzed by broad-scale methods that preclude estimating benefits by regulation or program (especially since some pollutants, like NO_x, are addressed in multiple portions of the Act).

Notwithstanding, analyses in the Section 812 studies are presented at too gross a level to be relevant to most policy decisions (such as whether to continue or to expand existing programs or whether to initiate new ones). So far the EPA has not fully embraced the recommendations made by its own Clean Air Science Advisory Committee and others regarding the need for less aggregated analysis.

In both Section 812 studies, the EPA assessed the consequences of air pollution programs by comparing specific scenarios. The retrospective study compared a scenario reflecting economic and environmental conditions observed with the Act in place with a counterfactual scenario of economic and environmental conditions assuming air pollution control technologies as they were in 1970. The prospective study compared a scenario in which all rules promulgated or expected to be promulgated pursuant to the 1990 Amendments with a counterfactual scenario assuming federal, state, and local air pollution controls as they existed in 1990. Both studies held constant the geographic distribution of populations and economic activities across the scenarios.[2]

The counterfactual assumption that technologies remain static—an obvious simplification—is central to the overall results of these analyses. Arguably, in the absence of new federal regulation, one would expect some abatement of air pollution due to state or local regulation or, possibly, voluntary measures. Local actions to reduce air pollution in the United States date back to the 19th century. More recently, some states (eg, California) have imposed particularly stringent controls that exceed the requirements of the Clean Air Act. If state and local regulations were equivalent to federal regulations, a benefit-cost analysis of the federal Clean Air Act would be a meaningless exercise: both benefits and costs would equal zero. For both Section 812 studies, EPA and outside experts wrestled with the possibility of developing more realistic counterfactual scenarios. In the end, they concluded that they could not predict how state and local regulations or voluntary efforts would have differed from those of the Clean Air Act.

In each Section 812 study, the aggregate actual and counterfactual scenarios were evaluated by a sequence of models (economic, emissions, air quality, physical effect, economic valuation, and uncertainty models) to estimate the differences in economic, human health, and environmental consequences of air pollution. Both studies examined the benefits and costs of reducing volatile organic compounds (VOCs), NO_x, SO_2, carbon monoxide (CO), inhalable PM (specifically, PM less than 10 μm in aerodynamic diameter [PM_{10}]), and fine PM ($PM_{2.5}$).[3]

The retrospective analysis also assessed the effect of Clean Air Act provisions governing lead in the environment. Because the 1990 Amendments do not include new controls of lead levels, the prospective analyses did not consider lead. Both studies covered the federal programs broadly, but some were omitted, largely because of data or modeling limitations. Emissions of hazardous air pollutants were not extensively considered in either study.[4] The most recent revisions to NAAQS for PM and ozone were omitted from the prospective study although EPA analysis indicates that, because of similarities in the baseline assumptions, the benefits and costs reported in the RIA for PM and ozone can be considered incremental to the results of the prospective analysis (EPA 1997a). Estimates for Title VI of the Clean Air Act Amendments, regarding stratospheric ozone depletion, were developed in the prospective study but were not fully integrated into the main analysis.

2. Although the scenarios do reflect the basic trends in population and economic growth across the United States over the relevant time periods, they do not allow for the possibility that people would respond to pollution by moving away from the dirtiest areas.

3. The incremental effects of the Clean Air Act Amendments on directly emitted primary particles such as PM_{10} and $PM_{2.5}$, called *primary PM*, are relatively small. PM in the atmosphere comprises both primary PM and secondary PM (formed in the atmosphere by conversion of gaseous emissions of SO_2, NO_x, and organic compounds). Because the US Clean Air Act, especially the 1990 Amendments, require substantial reductions in secondary PM, they probably have a much larger effect on PM_{10} and $PM_{2.5}$ than might be apparent if analyses only consider the changes in primary PM.

4. Some pilot analyses of hazardous air pollutants were conducted, but the poor quality of the available information precluded comprehensive quantification of the effects.

Emissions estimates reflect expected growth in population, transportation, electric power generation, and other economic activity. Different estimation procedures are used for stationary, mobile, and area sources although the benefit and cost estimates are not disaggregated in that manner. Costs are estimated as increases in expenditures to meet the additional control requirements of the Clean Air Act Amendments, including operation and maintenance expenditures and amortized capital costs (ie, depreciation plus interest costs associated with the existing capital stock).[5]

After an extensive review of the epidemiologic literature, including both short-term and long-term studies, EPA relied primarily on the long-term American Cancer Society (ACS) cohort study by Pope and colleagues (1995) to estimate adult mortality associated with elevated PM concentrations for the Section 812 studies.

Relevant work continues to be published. Interestingly, a reanalysis of the ACS study by Krewski and coworkers (2000) (not yet published at the time of the Section 812 studies) estimated slightly larger effects of $PM_{2.5}$ on mortality than found by Pope and colleagues in 1995, but under some model specifications, the effects were markedly lower and in some cases not even statistically significant. Krewski and coworkers (2000) found that the mortality effects of fine particles ($PM_{2.5}$) varied with education level: the estimated coefficients were higher for individuals without a high school education than for those with higher levels of education. Recently, both Abbey and coworkers (1999, in a California cohort) and Lipfert and associates (2000, in a study of veterans' mortality and pollution) reported mixed findings for PM and mortality. Studies from around the world have also strengthened the case for infant mortality caused by exposure to fine particles (Woodruff et al 1997; Bobak and Leon 1999; Lipfert et al 2000). Such effects were not considered in either of the 812 studies, which could only have reflected the evidence available when they were carried out. Krewski and coworkers also estimated the effects of other pollutants on mortality and found relatively large and statistically significant effects of SO_2.

Although efforts to model air quality in the Section 812 studies focused on the full range of pollutants, both studies attributed the majority of total benefits to changes in PM concentrations. However, the studies considered neither specific sources nor chemical composition of particles. Thus, secondary PM formed from SO_2, NO_x, and VOCs were all treated as fine particles.

The retrospective study found considerable benefits associated with reductions in lead—principally lead in gasoline. The health benefits of air quality improvement that were expressed in monetary terms include reduced incidence of human health effects—principally premature mortality—as well as improvements in visibility and damage to agricultural crops avoided. Despite efforts to characterize the impact of air pollution on ecosystems, the inability to quantify in monetary terms (or monetize) the damages precluded the development of benefit estimates for this purpose.[6] The analysis was similarly limited for possible cancer risk and certain other health effects associated with criteria pollutants.

The monetized benefits reflect interpretations of the science and economic literature made by the EPA in consultation with its outside experts. As a form of sensitivity analysis, a number of alternative interpretations of the literature were also examined. The quantitatively most important interpretations concern the valuation of premature mortality: $5 million per life (1990 dollars). In both their retrospective and prospective analyses, the EPA developed an alternative scenario based on the loss of life-years to reflect the greater susceptibility of older individuals to mortality induced by air pollution. In both studies this scenario yielded notably lower benefits. The prospective study also examined alternative assumptions about the incidence of mortality, the incidence and valuation of chronic bronchitis, and certain other effects. For Title VI, sensitivity analysis addressed the possibility of risk-averting behaviors, such as remaining indoors or using sunscreen or hats more frequently.

The techniques applied by EPA to estimate health benefits in the Section 812 studies and in several RIAs (including those for the 1997 PM and ozone NAAQS), recently were reviewed independently by a panel of the National Research Council (2002). The panel concluded that "despite many inherent uncertainties … regulatory benefits analysis can be a useful tool for generating information valuable to policy makers and the public." The panel supported the EPA's general approach, including the "[application] of concentration- or exposure-response functions (derived from the health literature) to estimated changes in population exposures" but suggested several areas in which the EPA's implementation of the general approach could have been improved. Among other recommendations, the panel suggested that future analyses

5. Costs for meeting Title IV requirements through the SO_2 trading program instituted under the Clinton administration were estimated by a model that effectively allocates the cost of emissions reductions, within the context of responding to signals in the electric power and tradable allowance markets.

6. Although the prospective study estimated benefits and costs for Title VI (stratospheric ozone), these calculations were not included in the retrospective analysis. Accordingly, for comparative purposes, the results for Title VI are excluded from Table 1.

should consider a range of regulatory options and compare the benefits of each and should also consider secondary or unintended effects of regulations. The panel noted that future analyses needed to include more complete information about emissions, including the contribution of specific emission sources to human exposure, in order to estimate counterfactual or baseline conditions and their respective uncertainties. In addition, the panel recommended that information about morbidity be improved, in part by categorizing the severity of specific disease outcomes. Finally, the panel recommended that EPA provide more comprehensive quantitative assessments of the uncertainty of benefits estimates and make those assessments primary, rather than ancillary, components of the benefits analysis. Interestingly, the panel noted the need to "develop methods of evaluating causal uncertainty relating to key outcomes" and to include this uncertainty in benefits estimates.

Table 2.1. Central Estimates of Monetized Present-Value Benefits and Costs of the US Clean Air Act (1970–1990) and Amendments (1990–2010)

	1970–1990[a]	1990–2010[a]
Benefits[b]		
Avoided mortality	$17,971	$610
Avoided morbidity	$3,595	$49
Ecologic and welfare effects	$607	$29
Total monetized benefits	$22,171	$690
Costs[b]		
Total costs	$523	$357

[a] Estimates in billions of 1990 dollars.

[b] Total benefits and total costs are only for Titles I–IV of the Clean Air Act, not Titles V and VI (see text for explanation) (EPA 1997a, 1999).

Results

Table 2.1 summarizes the central present-value estimates for both benefits and costs developed in the two Section 812 studies for Titles I through IV of the Clean Air Act.[7] These results demonstrate that the aggregate monetized benefits of air pollution control are estimated to exceed costs by more than an order of magnitude for 1970 to 1990 (the subject of the retrospective study). Note that more than four fifths of these estimated benefits derive from mortality reductions (avoided mortality) at $4.8 million per life.[8] The majority of mortality benefits are associated with the drop in PM concentrations during the first 20 years of the Clean Air Act. The phase-down of lead in gasoline also contributed to reduced mortality during this period. Reductions in morbidity, including chronic bronchitis and other effects, account for about 16% of the total benefits. Ecologic and welfare effects are estimated to contribute less than 3% of the total benefits.

In the prospective study, from 1990 to 2010, total monetized benefits also exceeded costs although the ratio of benefits to costs was considerably lower than for 1970 to 1990 (period of the retrospective study). This difference suggests that the incremental gains to society from the 1990 Amendments are less than those achieved in the first 20 years of the Clean Air Act. Interestingly, the proportional contributions of the different categories of monetized benefits are roughly the same for the two time periods although benefits from avoided mortality actually account for a slightly higher proportion of the total benefits in 1990 to 2010 compared with 1970 to 1990.

Overall, on the basis of the prospective study, EPA concluded the following (EPA 1999, page v):

> While alternative choices for data, models, modeling assumptions, and valuation paradigms may yield results outside the range projected in our primary analysis, we believe based on the magnitude of the difference between the estimated benefits and costs that it is unlikely that eliminating uncertainties or adopting reasonable alternative assumptions would change the fundamental conclusion of … [the] study: the Clean Air Act(s') … total benefits to society exceed its costs.

Cost-benefit studies are probably of greater use for policy making if results can be linked directly to specific provisions of laws, regulations, or policies. Although EPA did not link these in the Section 812 studies, others have used the EPA analyses as a starting point and have developed more disaggregate estimates by title of the Clean Air Act. For example, Smith and Ross (1999) conducted an analysis for all titles, whereas Chestnut (1995) and Burtraw and colleagues (1998) considered the impact of Title IV on the electricity generation sector. In addition, EPA was able to develop separate benefit estimates for their new ozone and fine particulate NAAQS (EPA 1997b). The findings from these studies are presented in Appendix 2.A. These analyses suggest major differences in net benefits across titles of the Act. They also reveal that the new $PM_{2.5}$ NAAQS is likely to have, in most analyses, a better cost-benefit ratio than the new ambient ozone standards.

7. Interestingly, when the EPA adopted a life-years method as a form of sensitivity analysis—with a life-year valued at $293,000—the mortality benefits were considerably lower than shown in Table 1, although they still exceeded control costs by an order of magnitude.

8. Except for a supplementary calculation for avoided costs of nitrate reductions.

THE UK EXPERIENCE

HISTORY OF AIR QUALITY ISSUES AND CONTROL

The era of modern air pollution regulation began in the United Kingdom after the London smog of 1952 (Brimblecombe 1987). The subsequently convened Beaver Committee identified domestic coal smoke as the key source of urban air pollution. The Committee recommended introducing smokeless fuel for domestic heating, which in turn led to the UK Clean Air Act of 1956. The Act established smokeless zones in urban areas and provided financial assistance to local authorities for grants to support conversion of domestic grates to a type suitable for using smokeless fuel. At this time, gas and electricity were growing in popularity for domestic heating. Levels of smoke fell rapidly in many parts of the country.

A response to an all-too-obvious problem was thus relatively rapidly developed and adopted. (Success was not universally achieved across the United Kingdom, however. Some areas that lacked gas supplies continued to experience problems with air quality until the 1980s—although even in these areas, concentrations fell compared with the pre-1956 period.) Such progress was achieved through direct source reductions without Air Quality Standards. In the 1950s, both the policy and science communities recognized that too little was known about the effects of air pollutants on health. The need for evidence was addressed by the Air Pollution Research Unit, established by the Medical Research Council and other scientists. Studies were undertaken in London and other polluted cities (Waller 1971).

The concept of achieving threshold concentrations was developed for the then-predominant urban air pollutants: particles (measured as black smoke) and SO_2 (measured by a wet chemistry titration). These thresholds seemed within reach in urban areas, given the control policies being pursued with increasing vigor there. Levels fell below the thresholds in many areas, and some scientists became persuaded that at least the acute effects of air pollution were close to being mitigated (Holland et al 1979). Some scientists, notably professor PJ Lawther, cautioned against a single focus on the acute effects of large episodes. He argued that long-term exposure to air pollution might play a part in causing the then-prevalent disease chronic bronchitis (Lawther 1961).

During the early years of air pollution control in the United Kingdom, formal cost-benefit considerations were not part of policy-making deliberation. A policy of as much smoke reduction as could be afforded was followed. No attempt has yet been made to characterize the guidelines for expenditures. The public was rapidly persuaded of the dangers of smog, and they collaborated in reducing levels of key air pollutants. Other sources—especially motor vehicles—presented an increasing problem, but their impact on air quality was probably less apparent than it might have been because of the reductions in coal smoke pollution.

The UK Clean Air Act was extended in 1968. In 1970, the first European Commission (EC) directive relating to air pollution (EC directive 70/220/EC) was passed. Once again, the legislation focused on sources rather than levels: limits were established for emissions of CO and hydrocarbons from gasoline-powered cars.[9] The first Air Quality Standards applicable in the United Kingdom were called *EC Limit Values* and were developed by an EC directive in 1980. These dealt with black smoke and SO_2. The EC Limit Values were established on a basis of technical feasibility rather than quantified benefits to health. EC Limit Values for lead followed in 1982 and for nitrogen dioxide (NO_2) in 1985. All four EC Limit Values were formally adopted into UK legislation in 1989. Note that in addition to EC Limit Values, which had the force of law, more stringent Guide Values were also promulgated.

Since about 1990, the United Kingdom's domestic policy on air pollution has been active and has evolved in parallel with EC developments. In addition to the EC Limit Values, the United Kingdom has developed its own principles and guidelines for air pollution. These guidelines are described in a series of publications on air quality and health by the Committee on the Medical Effects of Air Pollution (COMEAP) and others (Appendix 2.B).

EVOLUTION TOWARD COST-BENEFIT ANALYSIS

The dramatic health effects of the London smog of 1952 launched the concern about air pollution in the United Kingdom that led to regulations. Concerns about the health effects of outdoor air pollution waned in the intervening years but are now increasing in response to growing evidence of adverse health effects at lower ambient concentration. Until the late 1980s, low levels (below threshold levels) of air pollutants were generally not considered to damage health, at least not acutely. Air Quality Guidelines published by the World Health Organization in 1987 set levels of air pollutants, below which effects on health were regarded as unlikely except possibly in a very small minority of the population (World Health Organization 1987). Such information fed into decision making in a straightforward way as target pollutant concentrations were proposed that implied safety.

9. Details of the development of EC legislation are available in the 1993 report *Urban Air Quality in the United Kingdom* (see Appendix 2.B).

Initially, policies for emissions reductions were developed to reduce air pollution to levels at which effects on health were thought to be unlikely. Feasibility of implementation was the guiding principle. As emissions control technology advanced, pollution levels were expected to continue to decrease. The health gain was not calculated in relation to cost of reducing levels of pollutants or emissions. Early legislation focused on pollution sources and followed a policy of continued reduction of source emissions (Brimblecombe 1987, pp 165–177). Decisions on acceptable levels of emissions tended to be based on best practice: For example, if one power station could achieve a particular level of SO_2 emissions per unit of power generated, others were assumed to be able to do so as well. Technology change was thus forced on older and dirtier industrial processes and motor vehicles. General improvement in air quality followed.

To some scientists the concept of exposure thresholds below which effects do not occur remained questionable. Many simply viewed air pollution as harmful and as something to be progressively reduced in line with technology advances. The United Kingdom's approach to regulation was essentially nonquantitative. No one calculated how many deaths or hospital admissions were attributable to the 1990 levels of air pollutants. Epidemiologic techniques then in use were not sufficiently sensitive to find adverse effects, except perhaps during air pollution episodes, which were less and less common.

Concern of the UK public health community was raised around 1993 and 1994 with publication of positive results of time-series studies (Schwartz and Dockery 1992; Dockery and Pope 1994). These findings persuaded regulators that even low levels of air pollutants were associated with adverse health effects. Findings with similar implications had been reported earlier for the United States by Lave and Seskin (1970, 1978) and others (Ozkaynak and Thurston 1987), but these studies had not motivated reconsideration of air quality standards. As the first time-series results were reported, questions were raised about their methods and interpretations, particularly with regard to whether the association between air pollution measures and health indicators was causal. Confounding by uncontrolled correlates of air pollution was of particular concern. For example, debate continues about the influence of temperature (Keatinge 2002). Thus the evidence of acute effects of air pollution on health has been accepted slowly, and first by regulatory and scientific communities. These communities recognized the evidence as indicating that emissions of air pollutants had not yet been sufficiently controlled. No emphasis was placed on quantifying the impact of air pollution control, however. As more studies from more areas with better explained methods added to the evidence, doubts about the validity of the evidence subsided.

With regard to quantifying the impact of air pollution regulation, the United Kingdom took a step forward in 1998 when COMEAP published *Quantification of the Effects of Air Pollution on Health in the United Kingdom* (UK Department of Health 1998). This first UK report of its kind was critically important in that it allowed policy developers to estimate the benefits to health that would—or at least might—be produced by reductions in air pollutant levels. The report offers options to be considered rather than final decisions, which has allowed industry and environmental groups to contribute to the debate on how to reduce air pollution levels. Comments on the report are posted on websites along with the original documents, discussed by the relevant expert committee, and incorporated into the report when appropriate.

ESTIMATING CLEAN AIR ACT BENEFITS AND COSTS

Design and Assumptions

In general, two approaches to conducting analyses were considered in the United Kingdom:

- use of quantified benefits to allow choice among competing policies; and
- use of monetized benefits to support decisions about the extent to which a policy that reduced pollutant levels should be pursued.

The former approach is comparatively easy; the latter very difficult.

Quantified Benefits The first efforts to quantify benefits were reported by COMEAP (UK Department of Health 1998). In that report, COMEAP provided two very different estimates for the effects of ozone during summer: one assumes no threshold of effect and the other assumes a threshold of effect at 50 ppb (8-hour average concentration). COMEAP briefly assessed the evidence of health effects due to long-term exposure to air pollutants but concluded that the evidence was insufficient for the United Kingdom to estimate the health effects quantitatively. COMEAP noted, however, that if the evidence were applicable in the United Kingdom, the impact of long-term exposure might be larger than the impact of short-term exposure (acute effects).

In 2001, COMEAP published a new report that described in detail the evidence that long-term exposure to PM affects health: *Statement and Report on Long-Term Effects of Particles on Mortality* (UK Department of Health 2001). Benefits

were estimated in terms of total life-years gained by the population of England and Wales alive in 2000 if the annual average $PM_{2.5}$ concentration were 1 μg/m^3 lower than that occurring in 2000. (The details of the calculation and its several assumptions are provided in the report.) Risk estimates were taken from Pope and coworkers' (1995) analysis of the ACS's Cancer Prevention II study. A life-table approach to convert changes in mortality rates to changes in life expectancy was developed by Hurley and Miller (Hurley et al 2000) at the Institute of Occupational Medicine in Edinburgh and Walton of the UK Department of Health (2001). The COMEAP report presents estimates of the health impact of a persistent 1 μg/m^3 reduction in annual average $PM_{2.5}$. The estimates span a 20-fold range of potential gain in life expectancy (0.2 to 4.1 million life-years).

These calculations of the impact of air pollution on life expectancy illustrate the benefit that might be produced by a small reduction in fine PM levels, although they are based on a hypothetical scenario of exposure. Such calculations have been difficult to explain to the public and the media; relative risk estimates from time-series studies have been more easily explained. We also do not know how gains in life expectancy due to a reduction in pollution would be distributed across the population. Individuals might value possibly substantial gains more than smaller gains.

Monetized Benefits The Interdepartmental Group on Costs and Benefits (of the National Air Quality Strategy) was established to consider these approaches. The group produced two reports (UK Department of the Environment 1999, 2001). The second report is especially valuable in that it includes the COMEAP analysis of benefits that might be produced by long-term changes in PM concentrations.

The Interdepartmental Group adopted an approach to cost-benefit analysis that is straightforward in concept but demanding in practice. It has the following elements:

- emissions mapping and modeling;
- assessing the costs of measures to reduce levels of particles;
- defining health benefits associated with specified reductions;
- examining nonhealth benefits; and
- balancing costs against benefits.

Throughout the 2001 report (UK Department of the Environment 2001), *additional* control measures are emphasized. (A substantial program is already in progress on the basis of this report; it is expected to lead to reduction of pollutant levels.) The analysis focuses on the costs and benefits of additional policies. Given the range of emissions sources, many combinations of measures could be devised to deliver a specified reduction in pollutant concentrations. A range of so-called policy packages is considered in the report and, for each, costs and benefits are estimated. Uncertainties, such as when the benefits of a reduction in levels of pollutant might appear, are also addressed. The report also considers the complexities of annualizing costs and discounting benefits.

Results

In its initial report quantifying the benefits of reduced short-term health effects as a result of reduced air pollution (UK Department of Health 1998), COMEAP concluded that current levels of air pollution have a considerable effect on health. Tables 2.2 and 2.3 present the estimated effects of PM_{10} and SO_2 and of ozone, respectively, on selected health outcomes.

After a further review of the existing evidence for long-term effects, COMEAP published a report on those effects and estimated possible population impact (UK Department of Health 2001). The estimated effects varied, depending on underlying assumptions (Table 2.4).

Table 2.2. Annual Deaths and RHAs in Urban Areas of Great Britain

Pollutant	Health Outcome[a,b]	Annual Total
PM_{10}	Attributable deaths (all cause)	8100
	Attributable RHAs	10,500
SO_2	Attributable deaths (all cause)	3500
	Attributable RHAs	3500

[a] Attributable = number attributable to a given pollutant as estimated by time-series studies.

[b] Attributable deaths also known as *deaths brought forward*; attributable RHAs also known as *RHAs brought forward and additional.*

Table 2.3. Annual Deaths and RHAs Affected by Ozone in Urban and Rural Areas of Great Britain[a]

Pollutant	Health Outcomes[b,c]	Ozone Threshold	
		50 ppb	0 ppb
Ozone	Attributable deaths (all cause)	700	12,500
	Attributable RHAs	500	9900

[a] During summer only.

[b] Attributable = number attributable to a given pollutant as estimated by time-series studies.

[c] Attributable deaths also known as *deaths brought forward*; attributable RHAs also known as *RHAs brought forward and additional.*

Table 2.4. Estimated Total Life Years to Be Gained[a] Due to Reductions in Mortality Rate After Air Quality Improvement[b]

Reduction in Mortality Rate	Total Life Years Gained (millions)	Comments
Estimate based on PM_{10} effect in time-series studies[c]	0.007–0.02	Estimate considered very likely to be at least this large. Time-series studies well replicated. Estimate suggests that the apparent long-term effect of PM is actually explained by unknown confounders.
0.1% (from lower adjusted relative risks in Krewski et al 2000)	0.2–0.5	Estimate considered accurate. It accounts for the few confounding factors that substantially reduced relative risks in Krewski et al 2000.
0.3% (from lower CI = 1.09 [Pope et al 1995])	0.6–1.4	Estimate considered reasonable but higher than predicted by some adjusted relative risks in Krewski et al 2000.
0.6% (from relative risk = 1.17 [Pope et al 1995])	1.2–2.8	Estimate considered less likely than others. Most factors examined in Krewski et al 2000 did not markedly affect relative risk, but some did and other unknown confounders are possible. Higher past exposures may also lead to an overestimate of relative risk at current levels.
0.9% (from upper CI = 1.26 [Pope et al 1995])	1.8–4.1	Estimate considered implausibly large for reasons given above and in comparison with other relative risks or total changes in life expectancy in recent years.

[a] Estimates are for the 2000 population of England and Wales followed to extinction with a range of reductions in hazard rates in those aged 30 years or more.

[b] Total effects immediate, phasing in gradually or stepwise after up to 40 years based on a 1 $\mu g/m^3$ drop in annual mean $PM_{2.5}$. Table adapted from UK Department of Health 2001.

[c] Estimate of effect in time-series studies based on a 1 $\mu g/m^3$ drop in annual mean PM_{10} assuming a coefficient of 0.075%, a loss of life expectancy of 2 to 6 months per death brought forward and a similar effect on all ages.

On the basis of these health impact analyses, the Interdepartmental Group on Costs and Benefits (of the National Air Quality Strategy) produced two reports summarizing both benefits and costs (UK Department of the Environment 1999, 2001). The 2001 report is of special interest because it includes the COMEAP analysis (UK Department of Health 2001) of benefits that might be produced by long-term changes in PM concentrations. The estimates from that report of the main costs and benefits of measures designed to reduce population-weighted PM_{10} levels by 0.751 $\mu g/m^3$ are summarized in Appendix 2.C.

EMERGING EUROPEAN AND CANADIAN TIME-SERIES ACCOUNTABILITY TECHNIQUES

Although many estimates of health benefits of pollution control have used risk estimates from studies of effects of long-term exposure (especially to PM using the results of the ACS study), risk estimates from time-series studies have also been used to calculate benefits of air pollution control (eg, to derive morbidity benefits in the EPA Section 812 studies)· Although time-series estimates of daily mortality capture only one dimension of health effects, a substantial body of evidence exists for short-term effects of air pollution on morbidity indicators. Two programs in Europe and Canada are using the time-series approach:

- The *Air Pollution and Health: A European Information System* (APHEIS) project. This project is funded by the EC public health agency DG-SANCO. APHEIS brings together 25 European research centers in 12 countries to develop, maintain, and regularly analyze a database with uniform health and air pollution measures. The project aims to facilitate the recurrent analysis of trends in air pollution and related health indicators to determine the magnitude of changes in health outcomes and whether those changes are attributable to air pollution exposure. APHEIS is now in its second year. Common protocols and approaches are now established and the databases are being developed (Medina et al 2001).
- Health Canada's *Canadian Progress Measures* project. This project is designed to collect and utilize time-series data to measure improvement of population health due to changes in air pollution over time. This project has begun testing pilot versions of a progress measure that includes temporal changes in location-specific ambient levels of air pollution and associations between those levels and daily deaths and hospitalizations for heart and lung disease.

Although the projects differ somewhat in approach, they both are using existing or readily available data sets to analyze periodically the relation between air pollution and certain health indicators. This approach is one of public health surveillance. Both projects aim to determine whether the *relative risk* of health effects is changing over time with changes in air pollution characteristics and levels and whether the *absolute level* of effects attributable to air pollution is changing as well. This general approach is feasible in Europe and Canada because of their routine collection of data on health and air pollutant concentrations. Although analytic methods remain to be established, these are attractive approaches that may provide the basis for relatively cost effective and sustainable systems for assessing the health impact of regulations over the long term.

CONCLUSION

Regulation of air pollutants in the United Kingdom began in the 1950s in response to highly visible air pollution incidents; this action resulted in identification and control of a key pollution source in urban areas: coal fires. Efforts to control air pollution in the United States began at about the same time in response to similar incidents and also to the problem of photochemical smog in Los Angeles. In both countries, the evident severity of the problems resulted in actions being taken without consideration of their costs or extensive analysis of their probable health benefits.

As levels of pollution have fallen, the benefits of further reduction have seemed less clear. The United Kingdom and United States, as well as the European Union and Canada, have been seeking to quantify the benefits of air pollution control and to compare those benefits to the costs. This effort has been encouraged by regulated industries that have become increasingly aware of the costs and increasingly sophisticated at challenging them. As a result, beginning in the 1980s in the United States and growing more widely throughout the world in the 1990s, techniques have been developed to analyze health benefits expected to be gained from new control actions and to compare those benefits to the costs. In recent years, these efforts have also included efforts to estimate the benefits of past actions retrospectively.

These efforts have demonstrated that epidemiologic methods can be applied with care in such studies. Within the inherent limits of such analyses, the overall benefits of reducing air pollution appear to have been substantial and, when monetized, appear to exceed the costs. In addition to the detailed efforts in the United States and United Kingdom to conduct these analyses, recent efforts in Canada and the European Union to use epidemiologic studies to measure progress have begun to show promise.

This experience has also highlighted several challenges that indicate future directions for research:

- **Estimation of baseline or counterfactual conditions.** Analyses of the health impact of air quality regulations must estimate baseline or counterfactual conditions: the air pollution concentrations that would have existed in the absence of the action or actions that were taken. These estimates can affect greatly the estimated benefits of the action. At a minimum, sensitivity analysis is required to determine the effects of different assumptions about baseline conditions. Better estimates will also probably require better information about factors that can affect emissions of pollutants in the absence of regulation.
- **Improved measurement of actual changes in health effects.** Although the strength of the existing data from epidemiologic studies allows some comfort in assuming that reducing pollution will reduce health effects (ie, produce health benefits), these studies require increased validation from carefully constructed prospective and retrospective analyses designed to actually measure the effects resulting from a regulatory action. Recent efforts in Dublin (Clancy et al 2002), Hong Kong (Hedley et al 2002), and elsewhere show promise in this direction.
- **Improved ability to measure and estimate health benefits of specific air pollution regulations.** To date, most retrospective efforts to estimate the health benefits of air quality regulations have been aggregate, measuring the benefits of overall reductions in pollution resulting from multiple actions. The costs of reduction, however, are most directly tied to specific regulatory measures (eg, a change in fuels and engines, a requirement for reduced emissions from electric power facilities). The few disaggregated analyses that have been conducted (eg, EPA 1999) suggest that all actions will not necessarily have the same benefits, for either emission reductions or health, or the same costs. Improved ability to estimate the benefits of specific measures will enhance future decision making.
- **Techniques to measure the benefits of incremental reductions accumulated over long time frames.** The largest estimates of health benefits from emissions reductions derive from studies of cohorts exposed to PM over the long term. But measuring actual health benefits over the long term is a challenge. Pollution reductions are likely to be accomplished in small increments over a long time, and other factors affecting health (eg, medical care) are likely to change over the same time frame. Improved techniques for identifying changes in health attributable to air pollution emission reductions over the long time frames would substantially improve analyses of health benefits.

Despite these limitations, the increasing sophistication of these techniques, particularly over the last decade, is impressive. Addressing these continuing challenges will further improve our ability to measure the benefits from actions to reduce air pollution in coming decades.

REFERENCES

Abbey DE, Nishino N, McDonnell WF, Burchette RJ, Knutsen SF, Lawrence Beeson W, Yang JX. 1999. Long-term inhalable particles and other air pollutants related to mortality in non smokers. Am J Respir Crit Care Med 159:373–382.

Ad-Hoc Group on the Economic Appraisal of the Health Effects of Air Pollutants (UK). 1999. EAHEAP report. HMSO, London, England.

American Thoracic Society. 2000. What constitutes an adverse health effect of air pollution? Am J Respir Crit Care Med 161:665–673.

Bell ML, Davis DL. 2001. Reassessment of the lethal London fog of 1952: Novel indicators of acute and chronic consequences of acute exposure to air pollution. Environ Health Perspect 109(Suppl 3):389–394.

Bobak M, Leon DA. 1999. The effect of air pollution on infant mortality appears specific for respiratory causes in the post-neonatal period. Epidemiology 10:666–670.

Brimblecombe P. 1987. The Big Smoke. Methuen and Co, London, England.

Brimblecombe P. 1999. Air pollution and health history. In: Air Pollution and Health (Holgate ST, Samet JM, Koren HS, Maynard RL, eds), pp 5–18. Academic Press, London, England.

Burtraw D, Krupnick AJ, Mansur E, Austin D, Farrell D. 1998. Costs and benefits of reducing air pollutants related to acid rain. Contemp Econ Policy 16(4):379–400.

Chestnut LG. 1995. Human Health Benefits from Sulfate Reductions under Title IV of the 1990 Clean Air Act Amendments. Final Report. Prepared for the EPA under contract 68-D3-0005. Office of Air and Radiation, Office of Atmospheric Programs, Acid Rain Division, Washington DC. Available from *http://yosemite.epa.gov/ee/epa/ria.nsf/377d5c0e0e55d70e85256b43007f9968/183623001d796c8c852564620063327f?OpenDocument*.

Ciocco A, Thompson DJ. 1961. A follow-up of Donora ten years after: Methodology and findings. Am J Public Health 51:155–164.

Clancy L, Goodman P, Sinclair H, Dockery DW. 2002. Effect of air pollution control on death rates in Dublin, Ireland: An intervention study. Lancet 360:1210–1214.

Department of Health (UK) (Committee on the Medical Effects of Air Pollutants). 1998. Quantification of the Effects of Air Pollution on Health in the United Kingdom. A Report of the Committee on the Medical Effects of Air Pollutants. HMSO, London, England.

Department of Health (UK) (Committee on the Medical Effects of Air Pollutants). 2001. Statement and Report on Long-Term Effects of Particles on Mortality. HMSO, London, England. Available from *www.doh.gov.uk/comeap/state.htm#state*.

Department of the Environment (UK) (Food and Rural Affairs). 2001. Interdepartmental Group on Costs and Benefits: An Economic Analysis to Inform the Review of the Air Quality Strategy Objectives for Particles. Food and Rural Affairs, London, England.

Department of the Environment (UK) (Transport and the Regions). 1999. An Economic Analysis of the National Air Quality Strategy Objectives: An Interim Report of the Interdepartmental Group on Costs and Benefits. HMSO, London, England.

Dockery DW, Pope CA III. 1994. Acute respiratory effects of particulate air pollution. Annu Rev Public Health 15:107–132.

Environmental Protection Agency (US). 1997a. The Benefits and Costs of the Clean Air Act, 1970 to 1990. EPA-410-R-97-002. Office of Air and Radiation, Washington DC.

Environmental Protection Agency (US). 1997b. EPA's regulatory impact analyses (RIA) for the revised ozone and PM NAAQS and proposed regional haze rule (last updated 7/11/2002). *www.epa.gov/ttn/oarpg/naaqsfin/ria.html*. Accessed 8/20/03.

Environmental Protection Agency (US). 1999. The Benefits and Costs of the Clean Air Act 1990 to 2010: EPA Report to Congress. EPA-410-R-99-001. Office of Air and Radiation, Washington DC.

Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. 2002. Cardiorespiratory and all-cause mortality after

restrictions on sulphur content of fuel in Hong Kong: An intervention study. Lancet 360:1646–1652.

Holland WW, Bennett AE, Cameron IR, Florey C du V, Leeder SR, Schilling RSF, Swan AV, Waller RE. 1979. Health effects of particulate pollution: Reappraising the evidence. Am J Epidemiol 110:525–659.

Hurley JF, Holland MR, Markandya A, Miller BG, Anderson HR, Ayres JG, Donnan PT, Harrison RM, King K, Stedman JR, Stevenson KJ. 2000. Towards assessing and costing the health impacts of ambient particulate air pollution in the UK. In: Joint Research Programmes on Outdoor and Indoor Air Pollution (Review of Progress, 1999), pp 102–103. MRC Institute for Environment and Health, Leicester UK.

Keatinge WR. 2002. Winter mortality and its causes. Int J Circumpolar Health 61:292–299.

Krewski D, Burnett RT, Goldberg MS, Hoover K, Siemiatycki J, Jerrett M, Abrahamowicz M, White WH. 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.

Krupnick A, Morgenstern RD. 2002. The future of benefit-cost analyses of the Clean Air Act. Annu Rev Public Health 23:427–448.

Lave LB, Seskin EP. 1970. Air pollution and human health. Science 169:723–733.

Lave LB, Seskin EP. 1978. Air Pollution and Human Health. Resources for the Future, Johns Hopkins University Press, Baltimore MD.

Lawther PJ. 1961. Urban air pollution and its effects on man. In: The Air We Breathe (Farber SM, Wilson RHL, eds). Charles C Thomas, Springfield IL.

Lipfert FW, Morris SC, Wyzga RE. 2000. Daily mortality in the Philadelphia metropolitan area and size-classified particulate matter. J Air Waste Manage Assoc 50:1501–1513.

Medina S, Plasència A, Artazcoz L, Quénel P, Katsouyanni K, Mücke HG, De Saeger E, Krzyzanowsky M, Schwartz J, and the contributing members of the APHEIS group. 2001. APHEIS: Monitoring the Effects of Air Pollution on Public Health in Europe. Scientific Report 1999–2000. Institut de Veille Sanitaire, Saint-Maurice, France.

Ministry of Health (UK). 1954. Mortality and Morbidity During the London Fog of December 1952. Reports on Public Health and Medical Subjects 95. HMSO, London, England.

National Research Council (US). 2002. Estimating the Public Health Benefits of Proposed Air Pollution Regulations. National Academy Press, Washington DC.

Ozkaynak H, Thurston G. 1987. Associations between 1980 US mortality rates and alternative measures of airborne particle concentration. Risk Anal 7:449–462.

Pope CA, Thun MJ, Namboodiri MM, Dockery DD, Evans JS, Speizer FE, Heath CW Jr. 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am J Respir Crit Care Med 151:669–674.

Schrenk H, Heimann J, Clayton G, Gafafer W, Wexler H. 1949. Air Pollution in Donora, PA: Epidemiology of the Unusual Smog Episode of October 1948. Preliminary report. Public Health Service Bulletin 36. US Public Health Service, Washington DC.

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

Smith A, Ross M. 1999. Benefit-Cost Ratios of the CAAA by CAAA Title: An Assessment Based on EPA's Prospective Study. Report for General Motors Corp. CRA D2050-00 (November). Charles River Associates, Boston MA.

Waller RE. 1971. Air pollution and community health. J R Coll Physicians Lond 5:362–368.

Woodruff TJ, Grillo J, Schoendort KC. 1997. The relation between selected causes of post-neonatal infant mortality and particulate air pollution in the United States. Environ Health Perspect 105:608–612.

World Health Organization. 1987. Air Quality Guidelines for Europe. WHO Regional Publications, European Series 23. WHO Regional Office for Europe, Copenhagen, Denmark.

APPENDIX 2.A. Detailed Results from EPA Section 812 Studies and Other Analyses

Table 2.A.1, taken directly from the prospective study, summarizes the central estimates on a present value basis by title of the Clean Air Act. For Titles I through V, present value estimates of benefits exceed those of costs approximately by a factor of four. About 90% of these benefits are associated with avoided mortality. The remainder are associated with avoided morbidity and with ecologic and welfare benefits. On the cost side, present values from the prospective analysis show that Title I accounts for almost half of the total cost of the first five titles. Title II accounts for another third, with the balance distributed among Titles III through V. Because of the long-term nature of the benefits of Title VI (stratospheric ozone), the results for this title are not fully integrated into the overall findings. However, the present value benefits of this title exceed costs by more than a factor of 20. Other studies have also estimated costs and benefits of the 1990 US Clean Air Act Amendments for 2010 (Table 2.A.2).

APPENDIX 2.B. UK Publications on Air Pollution and Health

- The Environment Act of 1991;
- Four reports of the Advisory Group on the Medical Aspects of Air Pollution Episodes:

  Department of Health (UK). 1991. Ozone. First Report of the Advisory Group on the Medical Aspects of Air Pollution Episodes. HMSO, London, England.

  Department of Health (UK). 1992. Sulphur Dioxide, Acid Aerosols and Particulates. Second Report of the Advisory Group on the Medical Aspects of Air Pollution Episodes. HMSO, London, England.

  Department of Health (UK). 1993. Oxides of Nitrogen. Third Report of the Advisory Group on the Medical Aspects of Air Pollution Episodes. HMSO, London, England.

  Department of Health (UK). 1995. Health Effects of Exposures to Mixtures of Air Pollutants. Fourth Report of the Advisory Group on the Medical Aspects of Air Pollution Episodes. HMSO, London, England.

Table 2.A.1. Summary of Primary Central-Estimate Benefits and Costs of 1990 US Clean Air Act Amendments

	Annual Estimates (millions of US dollars)		
	2000	2010	Present Value
Costs			
Title I	$8,600	$14,500	$85,000
Title II	$7,400	$9,000	$65,000
Title III	$780	$840	$6,600
Title IV	$2,300	$2,000	$18,000
Title V	$300	$300	$2,500
Total costs, Title I–V	$19,000	$27,000	$180,000
Title VI		$1,400[a]	$27,000[a]
Monetized Benefits			
Avoided mortality	$63,000	$100,000	$610,000
Avoided morbidity	$5,100	$7,900	$49,000
Ecological and welfare effects	$3,000	$4,800	$29,000
Total benefits, Title I–V	$71,000	$110,000	$690,000
Title VI		$25,000[a]	$530,000[a]

[a] Annual estimates for Title VI stratospheric ozone protection provisions are annualized equivalents of the net present value of costs from 1990 to 2075 (for costs) or 1990 to 2165 (for benefits). The difference in time scales for costs and benefits reflects the persistence of ozone-depleting substances in the atmosphere, the slow processes of ozone formation and depletion, and the accumulation of physical effects in response to elevated UV-B radiation levels. Table adapted from EPA 1999.

Table 2.A.2. Summary of Cost-Benefit Studies of the 1990 US Clean Air Act Amendments for 2010

Study	Benefits[a]	Costs[a]
Title IV		
Burtraw et al (1998)[b]	$25,000	$800
Chestnut (1995)	$35,277	NA
New NAAQS (EPA 1997)[c]		
Ozone (8-hour), partial attainment	$400–$2100	$1100
Ozone (8-hour), full attainment	$1500–$8500	$9600
Fine particulates, partial attainment	$19,000–$104,000	$8600
Fine particulates, full attainment	$20,000–$110,000	$37,000
Clean Air Act Amendments (Smith and Ross 1999)[d]		
Title I	$26,564	$14,500
Title II	$14,968	$9000
Title III	$1925	$840
Title IV	$69,297	$2000

[a] Estimates in millions of 1990 dollars.

[b] While this estimate is specific to the eastern United States, these benefits are expected to account for 98% of total US benefits.

[c] Partial attainment costs are incremental to partial attainment of current standards and reflect partial attainment of promulgated standards. EPA estimates 17 potential residual nonattainment areas for ozone and 30 potential residual nonattainment counties for fine particulates as of 2010. Full attainment costs, however, are incremental to full attainment of current standards.

[d] Total US Clean Air Act Amendments benefit estimate ($110 billion) and cost estimates by title are from EPA 1999.

- Four COMEAP reports:

  Department of Health (UK). 1995. Asthma and Outdoor Air Pollution. A Report of the Committee on the Medical Effects of Air Pollutants. HMSO, London, England.

  Department of Health (UK). 1995. Non-Biological Particles and Health. A Report of the Committee on the Medical Effects of Air Pollutants. HMSO, London, England.

  Department of Health (UK). 1998. Quantification of the Effects of Air Pollution on Health in the United Kingdom. A Report of the Committee on the Medical Effects of Air Pollutants. HMSO, London, England.

  Department of Health (UK). 2001. Statement and Report on Long-Term Effects of Particles on Mortality. HMSO, London, England. Available from *www.doh.gov.uk/comeap/state.htm#state*.

- Eleven reports of the Expert Panel on Air Quality Standards:

  Department of the Environment (UK) (Transport and the Regions). 1994. 1,3-Butadiene. Third Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

  Department of the Environment (UK) (Transport and the Regions). 1994. Benzene. First Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

  Department of the Environment (UK) (Transport and the Regions). 1994. Carbon Monoxide. Fourth Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

  Department of the Environment (UK) (Transport and the Regions). 1994. Ozone. Second Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

  Department of the Environment (UK) (Transport and the Regions). 1995. Particles. Sixth Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

  Department of the Environment (UK) (Transport and the Regions). 1995. Sulphur Dioxide. Fifth Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

  Department of the Environment (UK) (Transport and the Regions). 1996. Nitrogen Dioxide. Seventh Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

  Department of the Environment (UK) (Transport and the Regions). 1998. Lead. Eighth Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

  Department of the Environment (UK) (Food & Rural Affairs). 1999. Polycyclic Aromatic Hydrocarbons.

Ninth Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

Department of the Environment (UK) (Food & Rural Affairs). 2001. Airborne Particles: What is the Appropriate Measurement on Which to Base a Standard? Tenth Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

Department of the Environment (UK) (Food & Rural Affairs). 2002. Second Report on 1,3-Butadiene. A Report of the Expert Panel on Air Quality Standards. HMSO, London, England.

- The National Air Quality Strategy 1997.
- The Air Quality Strategy 1999.
- *Urban Air Quality in the United Kingdom* 1993:

  Quality of Urban Air Review Group (UK). 1993. Urban Air Quality in the United Kingdom. First Report of the QUARG prepared at the request of the Department of the Environment. QUARG, Birmingham, England.
- *Diesel Vehicle Emissions and Urban Air Quality* 1993:

  Quality of Urban Air Review Group (UK). 1993. Diesel Vehicle Emissions and Urban Air Quality. Second Report of the QUARG prepared at the request of the Department of the Environment. QUARG, Birmingham, England.
- *Airborne Particulate Matter in the United Kingdom* 1996:

  Quality of Urban Air Review Group (UK). 1996. Airborne Particulate Matter in the United Kingdom. Third Report of the QUARG prepared at the request of the Department of the Environment. QUARG, Birgmingham, England.
- *Source Apportionment of Airborne Particulate Matter in the United Kingdom* 1996:

  Airborne Particles Expert Group (UK). 1999. Source Apportionment of Airborne Particulate Matter in the United Kingdom. UK Department of the Environment, Transport and the Regions, London, England.
- A series of reports by the Photochemical Oxidants Review Group:

  Photochemical Oxidants Review Group (UK). 1993. Ozone in the United Kingdom. Third Report of the UK PORG. UK Department of the Environment, London, England.

  Photochemical Oxidants Review Group (UK). 1997. Ozone in the United Kingdom. Fourth Report of the UK PORG, prepared at the request of the UK Department of the Environment, Transport and the Regions. UK PORG, London, England.

APPENDIX 2.C. Cost and Benefit Estimates for Some PM Reduction Measures

Table 2.C.1 presents results of work by the UK Interdepartmental Group on Costs and Benefits (of the National Air Quality Strategy). The group's 2001 report (UK Department of Health 2001) estimated the main costs and benefits of a package of measures designed to reduce population-weighted PM_{10} levels by 0.751 μg/m^3. The table presents costs (but not health benefits) in monetary terms. Consequently, direct comparison of costs and benefits is

Table 2.C.1. Summary of Costs and Benefits of Additional Measures[a]

Costs	
Annualized costs of illustrative package of additional measures (2000 prices), 2010	£785–1115 million[b]
Health benefits[c]	
Total number of life years saved, 2010–2110	
Undiscounted	278,000–508,000
Discounted (by 1.5%)	81,000–212,000
RHAs	
Annual reductions in attributable RHAs, 2010	250
Additional health benefits from SO_2 reductions	
Annual reductions in attributable deaths, 2010	48
Annual reduction in attributable RHAs, 2010	36
Nonhealth benefits	
Reduction in building soiling damages: annualized benefit, 2010	£52 million
Sensitivity analysis only[c]	
Cardiovascular hospital admission (CVAs)	
Annual reductions in attributable CVAs, 2010	146
Additional health benefits from NO_2 reductions[d]	
Annual reductions in attributable RHAs, 2010	171

[a] Table adapted from UK Department of the Environment 2001.

[b] The range reflects uncertainty of costs of particulate traps transport.

[c] Attributable = number attributable to a given pollutant as estimated by time-series studies. The original table described attributable deaths as *acute deaths brought forward* and attributable RHAs and CVAs as *additional or brought forward*.

[d] Sensitivity analysis only, because whether this effect is additional to reduction in particles is uncertain.

difficult, and simple conversions of lost life expectancy to costs (eg, total years of life saved/average life expectancy × 1.5 million pounds = benefits in monetary terms) should be avoided.

Monetizing benefit estimates presents a considerable problem. In 1999, the UK Department of Health published a report by an expert group that included physicians and health economists. This group evaluated different ways in which health benefits could be expressed and recommended that the Willingness-to-Pay approach was the most suitable, although it had several problems. (This approach and its advantages and limitations are discussed in detail in the UK Economic Appraisal of the Health Effects of Air Pollution [EAHEAP] report [Ad-Hoc Group on the Economic Appraisal of the Health Effects of Air Pollutants 1999]).

The UK Department of Health Ministers decided, on the basis of the EAHEAP findings, that the currently available data were not sufficient to convert the benefits to health of reducing air pollution into monetary terms with adequate certainty. The department cautioned against using the EAHEAP results in a cost-benefit analysis of the National Air Quality Strategy. Because of this decision, cost calculations like those in Table 2.C.1 can be presented only for illustrative purposes. Nonetheless, at least for illustrative purposes, when costs for so-called "added life-year" are within the potential range of valuation of this benefit, the probable costs of the policy options do not grossly outweigh the benefits. This cost-benefit relation is encouraging and has strengthened the case for pursuing the suggested measures. A research project aimed at improving the assessment of willingness to pay for a specified reduction in health risk due to air pollution is currently being conducted in the United Kingdom.

ABBREVIATIONS AND OTHER TERMS

ACS	American Cancer Society
APHEIS	Air Pollution and Health: A European Information System
CI	confidence interval
COMEAP	Committee on the Medical Effects of Air Pollution (UK)
CVA	cardiovascular hospital admission
EAHEAP	Economic Appraisal of the Health Effects of Air Pollution (UK)
EC	European Commission
EO	Executive Order
EPA	Environmental Protection Agency (US)
GPRA	Government Performance and Results Act (US)
HMSO	Her Majesty's Stationery Office (UK)
NAAQS	National Ambient Air Quality Standard(s) (US)
NO_x	oxides of nitrogen
NO_2	nitrogen dioxide
OMB	Office of Management and Budget (US)
PM	particulate matter
PM_{10}	PM less than 10 μm in aerodynamic diameter
$PM_{2.5}$	PM less than 2.5 μm in aerodynamic diameter
RHA	respiratory hospital admissions
RIA	regulatory impact analysis
SO_2	sulfur dioxide
VOC	volatile organic compound

Chapter 3. FROM REGULATORY ACTION TO EXPOSURE AND DOSE

INTRODUCTION

This chapter focuses on the links in the chain of accountability that extend from emissions sources through the exposures of and doses to the population (Figure 3.1). The ultimate links in the chain are considered in Chapter 4. Figure 3.1 shows these links and notes some of the factors that influence the nature and strength of the connections between links. In applying the model depicted by this chain, evidence need not be available for each step; for example, data might connect a change in emissions directly to health effects, skipping the two links in between. For example, a recent study in Atlanta was able to track reductions in vehicular traffic, changes in air quality, and changes in health indicators during the Summer Olympic Games (Friedman et al 2001).

Interventions in the form of regulatory action are directed at the left side of the chain of accountability; therefore, cause and effect relations are most readily identified in these initial links in the chain. The difficulty of establishing causal relations increases as the chain extends to the right. To assess accountability at the links that precede adverse health consequences, indicators are needed to predict biologically relevant exposures and doses. In this chapter we evaluate the potential for such indicators to demonstrate the effectiveness of air quality regulations in reducing emissions, improving air quality, or reducing average exposures and doses of the population. We also assess the adequacy of ongoing pollutant monitoring to estimate exposure and dose for epidemiologic studies of accountability.

This chapter is one part of Health Effects Institute Communication 11, *Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research,* which also includes five other chapters and an Executive Summary of the project.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R82811201 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.

MEASURING CHANGES IN EMISSIONS AND AMBIENT CONCENTRATIONS

Measuring changes in pollutant emissions consequent to air pollution regulation can be straightforward, involving measurements from specific sources before, during, and after regulation implementation. The resulting data are direct and quantitative measures of accountability. However, changes in ambient concentrations after emissions reductions may also depend on factors other than specific emissions regulations. These other factors include the relative contribution of the regulated source to total emissions of the pollutant in question, concurrent changes in emissions of the same pollutant by other sources, and variation in meteorologic influences on atmospheric transport and chemical transformation. As a result, changes in ambient concentration resulting from a specific regulatory action may be difficult to detect in routinely collected data on ambient air quality.

Nonetheless, the approach of tracking the consequences of changes in sources across the left side of the chain of accountability has proven informative. One example is the decline of atmospheric lead concentrations in the United States following the phase-out of leaded gasoline. In this case, the concentration of lead in the atmosphere closely paralleled the use of leaded gasoline in automobiles, the principal source of airborne lead (US Environmental Protection Agency [EPA*] 1998). In contrast, however, changes in vehicle emissions cannot be so readily linked to ozone concentrations in urban areas, which are determined by multiple factors beyond emissions of ozone precursors.

The links from ambient air quality to exposures and doses in the population have multiple determinants. Pollutant concentrations cannot be directly mapped to either exposures or doses. Some factors that determine this link include time-activity patterns, penetration of outdoor pollutants indoors, and chemical transformations. For the exposure to dose step, physiologic factors at the individual level such as ventilation pattern, activity level, and lung geometry come into play. Additionally, some pollutant concentrations may vary across a relatively fine scale, complicating exposure estimates from centrally sited monitors. The link between ambient air quality and personal exposure or dose will be stronger for a pollutant such as

* A list of abbreviations and other terms appears at the end of this chapter.

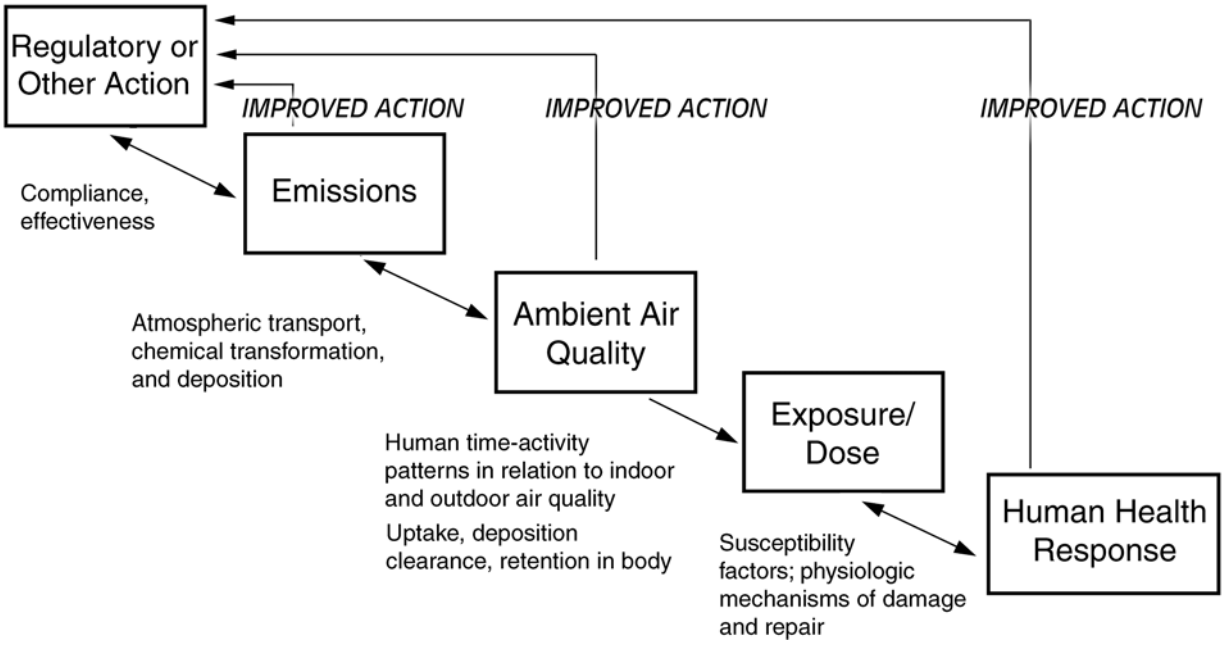


Figure 3.1. Chain of accountability. Each box represents a link between regulatory action and human health response to air pollution. Arrows connecting the links indicate possible directions of influence. Text below the arrows identifies general indices of accountability at that stage. At several stages, knowledge gained from accountability assessment can provide valuable feedback for improving regulatory or other action.

particulate matter (PM) less than 2.5 μm in aerodynamic diameter ($PM_{2.5}$), which exhibits relatively smooth spatial variation and penetrates the indoors with high efficiency, than for a pollutant such as coarse PM ($PM_{10–2.5}$), which has a larger gradient of spatial variation and does not penetrate indoors as efficiently.

Thus, the challenges of addressing accountability are likely to vary among sources and pollutants. Additionally, an assessment that encompasses the full chain of accountability will necessarily involve more assumptions and greater uncertainty than an assessment focused on the left-hand side of the chain. In the sections that follow we review the relatively well-developed tools that are available to characterize the effects of regulatory action on pollutant emissions and air quality as well as less well-developed tools for assessing the relation between ambient pollutant concentrations and the exposures and doses received by the population.

NORTH AMERICAN RESEARCH STRATEGY FOR TROPOSPHERIC OZONE

In order to evaluate the effectiveness of regulations, a process is needed that will demonstrate and quantitatively assess the progress of a regulatory action and, ultimately, its success in achieving its objectives and anticipated benefits. The North American Research Strategy for Tropospheric Ozone (NARSTO) Ozone Assessment (NARSTO Synthesis Team 2000) describes a process for accountability of air quality management. Three principal steps in this process are:

1. Verification that implemented emission controls have performed according to specifications and without unintended consequences.
2. Verification that environmental resources (eg, air, water, and soil) have responded as expected to changes in emission levels.
3. Verification that changes observed in the quality of these environmental resources have resulted in appropriate responses in public health and welfare.

Step 1 requires testing and evaluating the emission controls imposed by the regulatory system and verifying that they do in fact comply with specifications and established requirements. Some examples of federal requirements include engine certification, inspection and maintenance, and continuous emissions monitoring. Each requirement is designed to establish the efficacy of mandated control technologies at the source. The EPA compiles emissions data for a variety of chemical species and tracks and reports on the trends in these data on an annual basis (EPA 2002a). These annual estimates are based primarily on source activity and fuel-use patterns and are given for different source types at different scales. Emission inventories have been compiled in greater detail for space, time, and source in specific regions of the United States for use in the air quality models needed to address nonattainment of US National Ambient Air Quality Standards (NAAQS).

The step 1 requirements are not always sufficient for assessing regulation efficacy. For example, a combination of tunnel studies and roadside measurements provided the first compelling evidence that EPA's mobile source emissions models were seriously underestimating vehicular emissions of carbon monoxide (CO), volatile organic compounds (VOCs), and to a lesser extent oxides of nitrogen (NO_x) (US National Academy of Sciences 1991; NARSTO Synthesis Team 2000; Sawyer et al 2000). In addition, measurements of emissions in large samples of in-use vehicles have shown that a few so-called gross emitters can account for a large fraction of the total emissions from the sample (Zhang et al 1993). The problem of gross emitters is not necessarily addressed by an inspection and maintenance program, although some states have designed approaches to address the problem.

Another challenge to step 1 is that imposed emission controls can lead to unanticipated negative consequences. For examples: (1) introduction of vehicle fleets fueled by clean natural gas could increase formaldehyde and ultrafine particulate emissions (Holmén and Ayala 2002); or (2) introduction of MTBE (methyl *tert*-butyl ether) into the gasoline fuel supply to help reduce CO and other emissions could contaminate ground water (EPA 1999).

Step 2 requires demonstrating that air quality has changed as expected after emission reductions documented in step 1. Depending upon the specific air-quality goal, step 2 might involve monitoring chemical loading of pollutants in the atmosphere or deposition on the earth's surface and documenting trends in these data associated with emission changes. The nationwide network of air monitoring stations in the United States provides extensive data on concentrations of criteria pollutants. Limited temporal and spatial monitoring is conducted for noncriteria pollutants (EPA 2002a).

Step 3 involves demonstrating that the observed changes in environmental quality documented in step 2 result in the expected, quantitative benefit to health and welfare. This last step is the most difficult to accomplish; it could require many years of monitoring, depending on the health outcome of interest.

EMISSIONS REDUCTIONS AND ASSOCIATED AIR QUALITY

Even though the link between a new regulation and resulting changes in emissions is the most straightforward component of the chain of accountability, its assessment has a number of possible complications. For example, the distinction between efficacy (how well does the intervention work when tested under optimal circumstances?) and effectiveness (how well does the intervention work when actually applied?) must be considered. A control option may be efficacious but have only limited effectiveness in practice (eg, catalytic converters that do not perform to specification due to tampering or improper fuel use).

Emissions control programs often have multiple and incremental elements; so although evaluation of the full program is possible, assessment of each element may be difficult. Any decision to reduce emissions of a particular chemical is followed by a series of actions: (1) identifying all contributing sources of the chemical constituent; (2) assessing the cost and effectiveness of available control technology options; (3) establishing the objectives of the emission control program; and (4) setting in place procedures for tracking emissions reductions and evaluating effectiveness of the technology in meeting the objectives. This last step constitutes the emissions accountability component within an overall air quality management process. This key step has received too little emphasis to date.

Successful Cases

Lead The accountability process (as exemplified in NARSTO Synthesis Team 2000) can be illustrated with the reduction of lead in the environment in the United States. Figure 3.2 presents the trends in estimated lead emissions in the United States for 1970 to 1994 (EPA 1998a). The considerable decrease in emissions that resulted from control programs mandated by the US Clean Air Act was driven by phasing out leaded gasoline. Mobile sources were the main contributor to lead emissions. Unleaded gasoline was first introduced in the United States in 1975 for use in vehicles equipped with catalytic converters. The total market share of unleaded gasoline in 1975, 13%, climbed steadily to approximately 50% in 1982 and 99% in 1994. On January 1, 1996, leaded gasoline was prohibited for use in highway vehicles in the United States.

The connection between estimated lead emissions and air quality is shown in Figure 3.3. The trend in maximum observed quarterly lead concentrations measured at 122 sites from 1977 to 1985 and 208 sites from 1986 to 1996 shows a dramatic decline consistent with the emission estimates presented in Figure 3.2. Most measurement sites were located in urban areas representing the major US metropolitan areas.

Evidence that connects the changes in lead emissions to successive links in the chain of accountability are available. In Chapter 4 we discuss how observed reductions in ambient lead concentrations have been linked to decreases in blood lead levels in children aged 1 to 5 years. Associations between this biomarker of lead exposure and health effects are well-established and widely accepted. This example provides an empirical connection between mitigation efforts and the health benefits they were intended to produce.

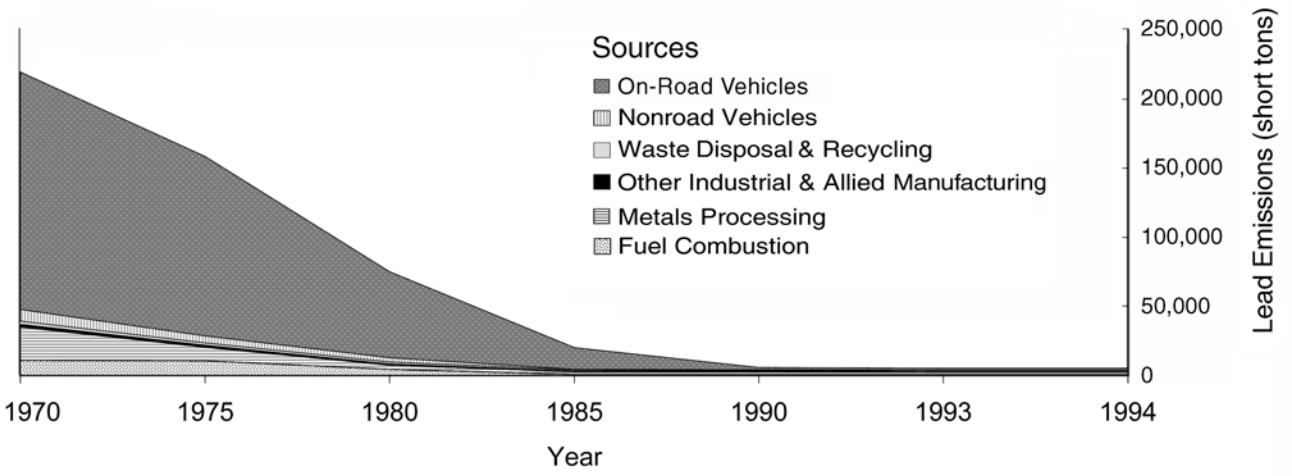


Figure 3.2. Trends in estimated US lead emissions. Adapted from EPA 1998a.

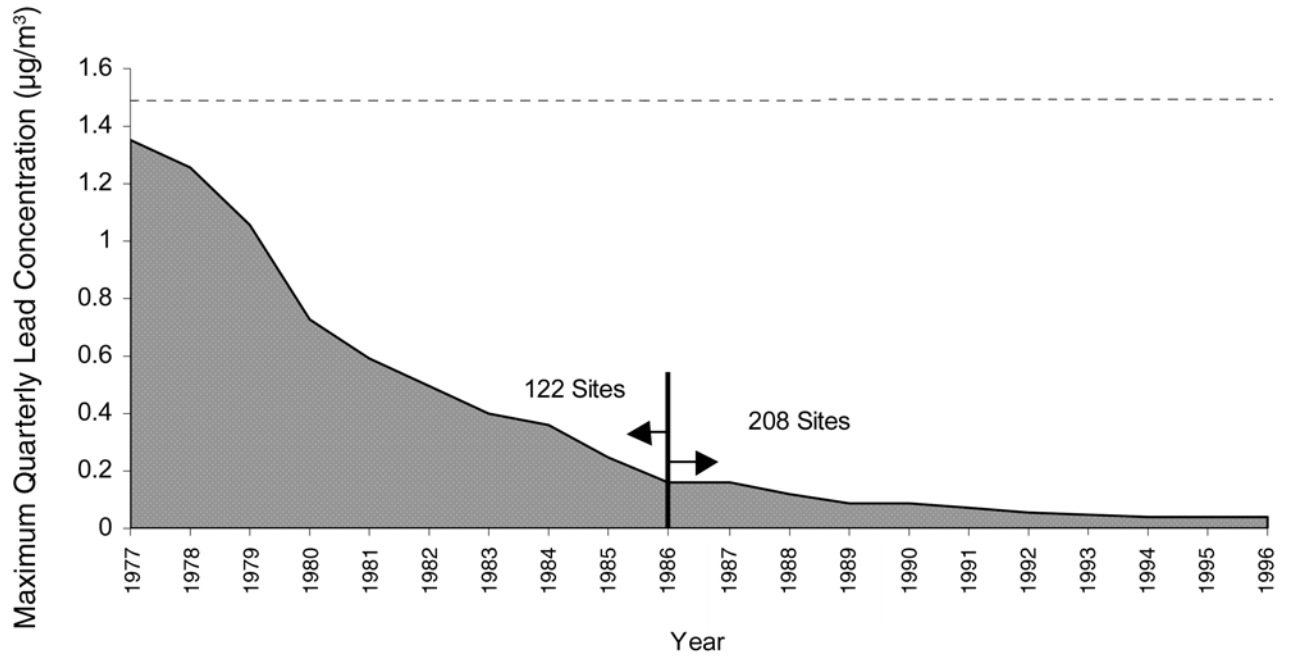


Figure 3.3. Trends in US airborne lead concentrations. Dashed line indicates the NAAQS for lead (1.5 µg/m^3). Adapted from NARSTO 2000.

Sulfur Dioxide Title IV of the 1990 US Clean Air Act Amendments includes provisions for reducing emissions from smokestacks and for uniform sampling of sulfur dioxide (SO_2) and NO_x emissions at all US electricity-generating utility plants. The resulting database, which documents a substantial drop in ambient levels of SO_2, is the most extensive that is currently available in the United States for emissions from a major source class; as such, it is useful for accountability assessment. Example of Title IV (phase I) SO_2 emission changes are shown in Figure 3.4, and the response in ambient SO_2 concentrations is shown in Figure 3.5.

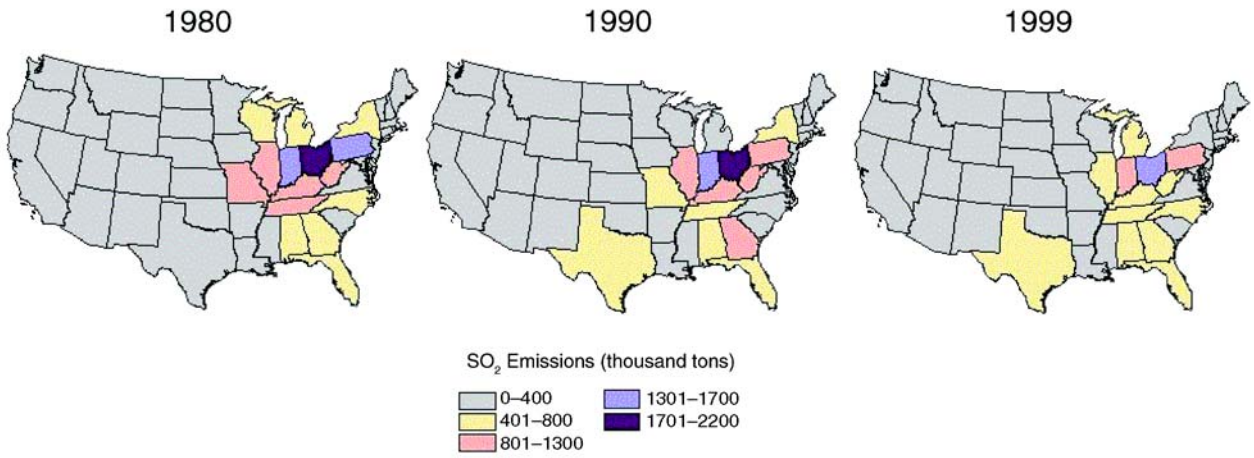


Figure 3.4. Total SO_2 emissions from electric utility plants over time. Title IV of the 1990 US Clean Air Act Amendments limits these SO_2 emissions. Adapted from EPA 2002b.

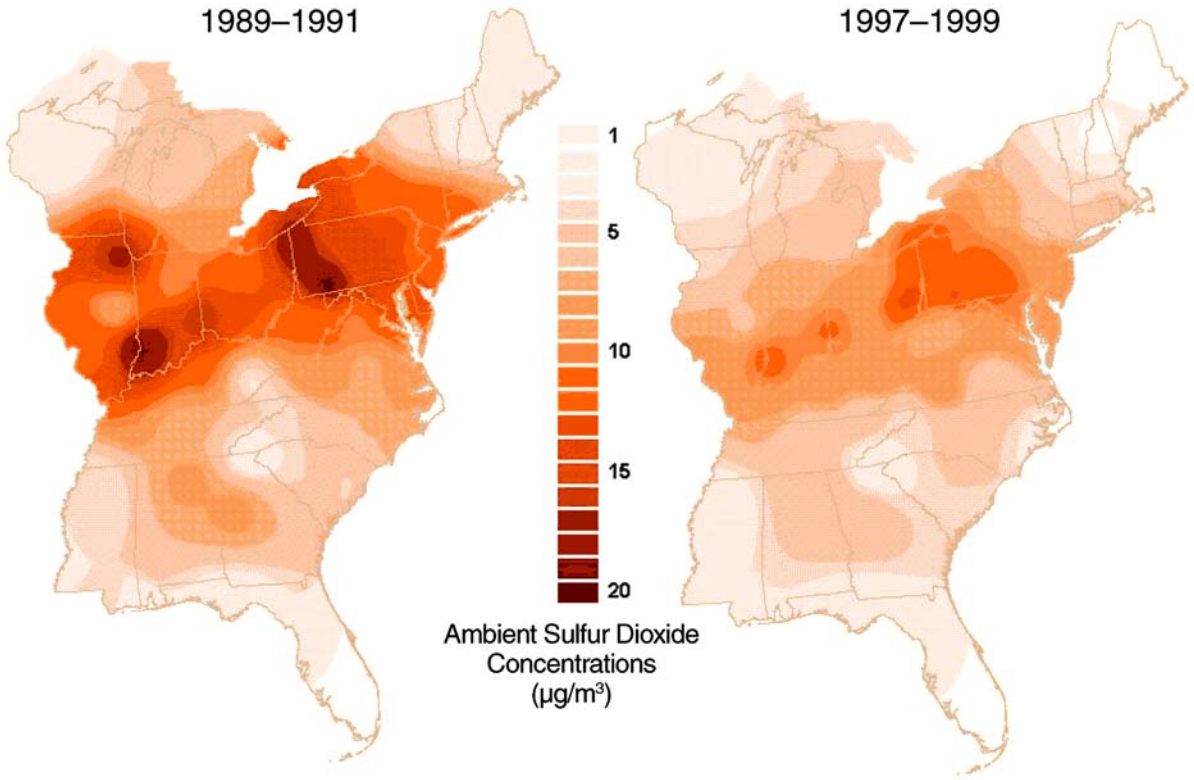


Figure 3.5. Ambient SO_2 concentrations in eastern United States. Adapted from EPA 2002c.

A Broader Challenge: PM and Ozone

The EPA compiles emissions data for a variety of chemical species and tracks and reports annually on the trends (EPA 2002a). These estimates are based primarily on source activity, fuel-use patterns, and source types, and they are given for different source types at different scales (national or state). With sufficient spatial and temporal disaggregation and routine validation, these data might aid analyses of the regulatory impact on source emissions.

Tracking the effect of changing emissions on air quality can be relatively straightforward for some primary pollutants, such as lead and SO_2. Tracking changes in secondary pollutants (eg, ozone and fine particles) is more challenging, however. Secondary pollutant concentrations are formed from precursor emissions via complex atmospheric chemistry and the precursors of secondary pollutants (such as ozone and fine particles) derive from diverse sources. Therefore, any single regulatory action may have relatively little leverage in changing ambient levels of secondary pollutants. Models and emissions inventories are needed to understand how precursor emissions relate to secondary pollutant concentrations. Emission inventories for ozone precursors that have detailed data for sources, geography, and time are available for US regions that are in nonattainment of the ozone NAAQS.

State Implementation Plans

The current US air quality management approach for regulating criteria pollutants defines a conceptual framework for linking pollutant emissions to ambient air quality. Each state provides the operational context for carrying out this management approach via a state implementation plan (SIP). SIPs offer opportunities for assessments relevant to accountability, and the modeling tools used to develop SIPs are also appropriate for forecasting accountability measures.

Although the air quality management approach can be based on elements other than models (eg, technology-based emissions controls or monitoring), modeling has been a critical part of the process for most criteria pollutants (eg, ozone, $PM_{2.5}$, SO_2, CO). Ozone has no direct emissions source; it is formed through a complex sequence of photochemical reactions involving emissions of VOCs and NO_x. A large fraction of fine-PM mass is formed via chemical reactions involving the same ozone precursors and sulfur oxides (SO_x). Models provide a means to quantify the complex chemical and meteorologic interactions that relate precursor emissions to production of secondary pollutants. To date, models have played a central role in air quality management for ozone, and they probably will for $PM_{2.5}$ as well.

As part of their administrative responsibility, states are required to monitor concentrations of ozone and PM to determine if and to what extent the NAAQS have been violated. In the event of a violation (ie, nonattainment), as part of the demonstration phase of the SIP process, states must acquire air quality, meteorologic, and emission data to support the operation of a modeling system and its verification for the base-case ozone and PM nonattainment year. States must then propose and evaluate control strategies to effectively mitigate the air quality violation. Typically, models are used to estimate the emission reductions needed to attain the NAAQS, taking into consideration emission projections. Once the control program for achieving attainment is stipulated and all supporting documentation is prepared, the SIP is then submitted for public comment and approval by EPA, which completes the SIP demonstration phase. After approval, the state must implement and enforce the control strategies and monitor the criteria pollutant(s) to ensure that progress is made toward meeting the NAAQS. Although the SIP system could provide a useful context for accountability assessments, it has been underutilized to date.

If air monitoring data later demonstrate that the NAAQS has been met, the area is redesignated to attainment status. If a state with nonattainment status does not develop a SIP, it is subject to federal sanctions (eg, restriction of development and federal funds) or imposition of a federal implementation plan. Failure to meet the NAAQS within the time frame designated in the SIP, as has been repeatedly the case for the ozone standard, triggers preparation of a revised SIP. To date, coordinated efforts to evaluate the monitoring data from the perspective of accountability have been few. Sanctions have not yet been imposed on any state, but some states have been threatened when their efforts were not clearly in good faith. The complex and difficult nature of ozone air quality management has required that SIPs be extended and revised many times since the first deadline (July 1975) set by Congress.

The SIP process timetable has not yet been finalized for the recently promulgated $PM_{2.5}$ standard (EPA 1997). Extensive data on nationwide $PM_{2.5}$ concentrations are becoming available from a new monitoring network that establishes baseline conditions against which future emissions reductions can be assessed. The SIP process for $PM_{2.5}$ could provide another opportunity for targeted accountability assessments, although its long duration will make it a challenge.

Air Quality Models

Models play a key role in several aspects of the current regulatory approach to air quality in the United States. EPA

uses models to predict improvements in air quality that may be expected after new regulations are implemented. These predicted reductions can in turn be translated into health benefits and compared to control costs in regulatory impact analyses, which EPA is required to conduct (US National Research Council 2002). Models also are an essential aspect of the SIP process to bring local air quality into compliance with the NAAQS. Thus, models are used in a variety of ways to predict future conditions and thereby justify regulatory and emissions policies. Surprisingly little effort has been devoted, however, to assessing how accurately models predict the changes observed after implementation of specific emissions policies.

The extent to which modeling systems (ie, models of air quality, emissions, and meteorologic data) adequately portray the dynamics of atmospheric processes that govern pollutant transport and transformation depends on the inclusion of key dynamic processes (eg, emissions, transport and dispersion, gas-phase chemistry, and dry deposition) of the most important sources and valid characterization of source emissions. Model development and application through the 1980s and 1990s have improved the modeling of these processes, as has the increased computing power now available to run larger and more complex models (Russell and Dennis 2000). Regional-scale modeling has addressed broader spatial and temporal scales.

With these improvements however, came new complications. For example, the incorporation of PM into a photochemical air quality model introduces several new levels of complexity to the modeling system. Theoretical modules (used to simulate chemical composition, size distribution, and ultimate fate of aerosols and required emissions estimates of primary particulate mass and composition) pose formidable challenges to the modeling community. Systemic limitations exist in the availability of information needed as inputs to photochemical models for ozone model application (including deficiencies in VOC and NO_x emissions estimates and meteorologic fields [eg, wind velocities and vertical exchange rates, boundary conditions, and aerometric data from the higher atmosphere]). Air quality models for PM will probably be even more uncertain. Such data gaps decrease modeling accuracy and the prospects for reducing or eliminating compensating errors.

In most modeling exercises, model parameters (predictors) are first adjusted so that model output corresponds closely with known pollution concentrations before a model is used to predict future concentrations. Then this tuned model is used to predict future conditions. A key test of performance is to evaluate the model for other situations without tuning adjustments, using the same rules for establishing model inputs for the second situation as for the first. The infrequent tests of this type have shown the performance of various models to be inconsistent (Roth 1999). For example, photochemical models have consistently overestimated the reductions in peak ozone concentrations that might be realized within a specified time period. Emissions reductions prescribed in the past and model-based projections of future air quality are documented in SIPs prepared from 1979 through the 1980s. In virtually all instances, past projections of improving air quality have not been achieved in the present. Although this failure in projection probably has many causes, a major one is the inaccuracy of modeling systems (ie, air quality, emissions, and meteorologic models and their inputs). Use of the projected concentrations to evaluate model performance could be informative with respect to the effectiveness of the SIP planning process. Complex socioeconomic forces drive population growth, industrial expansion, residential and commercial development, and emissions changes. In the United States, little effort has been put forth toward developing the capability to project emissions by quantitatively evaluating alternative future situations that reflect trends in these socioeconomic forces. Thus, uncertainties in emissions projections made 10 years or more in advance—for any hypothesized future achievement—can be considerable. This lack of prediction capability is likely to persist in the near future (US National Research Council 1999).

The current SIP process stipulates control programs based on so-called engineering estimates that are intended to achieve necessary reductions in precursor emissions as projected by model simulations. But the process lacks accountability procedures to ensure and demonstrate the achievement of goals (such as tracking progress toward and attaining emissions reductions and achieving planned air quality improvements). In addition, a post-SIP process should be instituted to determine cause(s) of modeling failures (such as flawed formulations, erroneous emissions representations, inadequate databases, and insufficiently effective implementation of emissions reductions) and to introduce improvements that will lead to more reliable estimation of pollution control needs.

Ambient Air Quality Monitoring

Air monitoring networks are designed and deployed to address specific objectives. These include providing a database for (1) determining air quality in major metropolitan areas; (2) determining pollution trends; (3) assessing compliance with or progress toward meeting air quality objectives or standards; (4) implementing emergency control plans to prevent or mitigate air pollution episodes; and (5) determining physical changes in the atmosphere

(eg, visibility) and in chemical loading of ecosystems through atmospheric deposition. Despite this broad set of objectives, networks have mostly been used to measure air quality to determine (1) maximum concentrations within specified spatial areas, (2) representative concentrations for use in assessing exposures, (3) source-receptor relations (mainly pertaining to the impact of specific sources on local air quality), and (4) background concentrations and temporal trends.

Because the SIP process determines air quality monitoring objectives, most available monitoring data are for the criteria pollutants (ie, ozone [O_3], SO_2, nitrogen dioxide [NO_2], CO, lead, and PM). Most network monitoring stations for criteria pollutants are located in urban areas that have multiple sources of air pollution and high population densities. Consequently, these stations are useful for assessing population exposures in epidemiologic studies or risk assessments. Some sites (eg, those near roadways at ground level) may be affected by local sources, possibly limiting the spatial representativeness of their data, whereas others (eg, those on tops of buildings) characterize air quality on more regional scales. Table 3.1 provides the design criteria for the US National Air Monitoring Stations (NAMS). Each year the EPA typically reports a variety of statistical measures for use in assessing ambient levels of criteria pollutants and their trends (EPA 2002a).

Continued operation of the extensive criteria pollutant network in the United States is essential for documenting long-term trends and spatial patterns in concentrations of pollutants known to have adverse human health effects. To the extent possible, discontinuities need to be avoided to minimize gaps in data. However, the two criteria pollutants of greatest current health concern, O_3 and $PM_{2.5}$, are often formed through secondary atmospheric reactions of precursor pollutants. Furthermore, $PM_{2.5}$ is a heterogeneous mix of particles of widely varying sizes, compositions, and sources. From an accountability perspective, a long-term trends database with sufficient spatial scope is essential to

Table 3.1. NAMS Design Criteria[a]

Pollutant	Population Size[b]	Station Density[b,c]	Site Selection
CO	> 500,000	≥ 2	Major traffic arteries and heavily traveled streets in downtown urban areas; neighborhood areas
NO_2	> 1,000,000	≥ 2	Neighborhood areas with highest NO_x emission
O_3	> 200,000	≥ 2	Urban areas (considering peak downwind ozone transport and population exposure); neighborhood areas on fringe of central business districts (considering peak downwind ozone transport and population exposure)
SO_2	> 1,000,000 500,000–1,000,000 250,000–500,000 100,000–250,000	H 6–10, M 4–8, L 2–4 H 4–8, M 2–4, L 1–2 H 3–4, M 1–2, L 0–1 H 1–2, M 0–1, L 0	Urban and neighborhood areas impacted by one or more point sources that cover broad geographic scales
PM_{10}	> 1,000,000 500,000–1,000,000 250,000–500,000 100,000–250,000	H 6–10, M 4–8, L 2–4 H 4–8, M 2–4, L 1–2 H 3–4, M 1–2, L 0–1 H 1–2, M 0–1, L 0	Urban and neighborhood areas affected by motor-vehicle diesel exhaust, industrial/combustion sources, and residential oil, coal, or wood burning for heat
$PM_{2.5}$[d]	PM_{10} criteria[d]	PM_{10} criteria[d]	PM_{10} criteria[d]

[a] The principal monitoring objective of NAMS is to measure pollutant concentrations in areas expected to have the highest concentrations and population exposures in terms of NAAQS averaging times. NAMS criteria identify two categories of stations: (a) stations located in areas of expected maximum concentration and (b) stations located in highly populated areas with poor air quality but not necessarily areas of expected maximum concentration. Source: Demerjian 2000.

[b] Selection of urban areas and station density are jointly determined by EPA and the state agency.

[c] Ranges for SO_2 and PM_{10} are based on expected pollutant levels as follows: H = exceeding NAAQS; M = exceeding 60% of primary or 100% of secondary NAAQS; and L = exceeding ≤ 60% of primary or 100% of secondary NAAQS.

[d] Federal reference method compliance (ie, mass monitoring at a minimum of 850 of 1050 sites required by 40 CFR 58 regulation) for $PM_{2.5}$ is similar to PM_{10} deployment. To date, ~300 chemical speciation sites have been allocated as follows: 54 trends sites (sampling 1 in every 3 days), ~40 sites to support Supersites (sampling 1 in every 3 days), 10 sites to support ongoing health studies (sampling daily), and ~200 sites to support SIPs and related work (sampling 1 in every 6 days).

document changes in precursor pollutants (such as reactive hydrocarbons) and specific $PM_{2.5}$ subcomponents that can be linked to specific source classes (such as size-classified counts, sulfate, nitrate, elemental and organic carbon, and trace elements). Such speciation data also would play a critical role in future epidemiologic studies directed at component-specific or source-specific analyses of health impact.

Data on a variety of these additional pollutant species are currently available from specialized monitoring networks. Photochemical assessment monitoring stations (PAMSs) produce data for a variety of VOCs that are O_3 precursors (EPA 1998b, 2003a); US National Atmospheric Deposition Program (NADP) has airborne acidity data (NADP 2002; also see *http://nadp.sws.uiuc.edu/*); Interagency Monitoring of Protected Visual Environments (IMPROVE) has visibility data (Malm et al 2000; also see *http://vista.cira.colostate.edu/improve/*); and PM Supersites have extensive data on PM subcomponents (Albritton and Greenbaum 1998; EPA 2003b; also see *www.epa.gov/ttn/amtic/files/ambient/pm25/casac/supersum.pdf*). These networks are much more limited in both temporal and spatial scopes than the criteria pollutant network, which limits their current utility for studies to assess effectiveness of control strategies aimed at specific source types or mitigating secondary pollutant production.

In recent years, the air quality management community has focused on the need for data to support the operation and evaluation of photochemical air quality simulation models (such as photochemical models). Typically, such intensive data sets were only available through special field measurement programs. In 1993, the EPA initiated the state-implemented PAMS program (EPA 1998b, 2003a). The full PAMS network is expected to provide the core of supplementary special field measurement programs to support model evaluation. Most importantly, PAMS deployment provides the first routine monitoring of pollutant-precursor data in the United States, thus making possible a variety of analysis opportunities that were previously impractical due to data limitations (Demerjian 2000).

Monitoring $PM_{2.5}$ and Its Components

The recently developed US $PM_{2.5}$ network has two major components: mass monitoring and routine chemical speciation. The basic network design (EPA 1998c, 2000a) consists of approximately 1300 sites, most located in populated regions but including several sites situated to characterize background levels, regional pollutant transport, and haze conditions. Sites established to collect data for NAAQS compliance are required to use federal reference method (FRM) samplers or federal equivalent method (FEM) samplers (samplers that pull air at a constant flow rate for 24 hours through a Teflon filter that is weighed before and after sampling).

Sampling using FRM or FEM methods starts at midnight and repeats at daily, three-day, or six-day intervals. Generally, 2 to 11 sampling sites are active in metropolitan areas with populations greater than 500,000. Continuous samplers (not FRM or FEM) provide real-time information, at hourly or shorter intervals. These samplers are required in each Metropolitan Statistical Area with a population of more than one million. Sites are chosen to ensure modeling of NAAQS compliance, characterization of regional and background exposures, and collection of other data.

Currently, 1050 FRM sites are planned for collection of compliance data, and approximately 210 continuous mass monitoring sites are planned for enhanced temporally resolved measurements. Another 110 IMPROVE sites are planned to characterize pollutant transport, background concentrations, and visibility. FRM and FEM monitors provide data on total $PM_{2.5}$ or PM_{10} (PM less than 10 μm in aerodynamic diameter) concentrations. These indicators are equivalent to those utilized in most epidemiologic studies that have provided evidence relevant to PM NAAQS.

Approximately 250 sites will provide data from the chemical speciation of PM samples collected every 1 in 3 or 1 in 6 days. These speciation sites will characterize the chemical composition of ambient aerosols, providing insight into potentially toxic components of PM, supporting source attribution analyses, and revealing potential positive or negative biases in bulk $PM_{2.5}$ data from FRM and FEM monitors and continuous monitors. Each site will be equipped with samplers using filters of multiple materials (Teflon, quartz, and nylon) to capture more representative samples of potentially volatile components and for their subsequent laboratory analysis.

Specifically, 54 of the speciation sites will sample trends to support ongoing health studies; 44 will sample every three days and 10 will sample daily. The remaining 200 sites will provide data for SIPs and other work; these will sample every six days. Chemical speciation of $PM_{2.5}$ filter samples will be analyzed at centralized, EPA-certified, analytical support service laboratories. The $PM_{2.5}$ samples will be analyzed for metals (such as iron, nickel, zinc, and lead), ions (such as chloride, sulfate, and nitrate), and organic and elemental carbon.

The current and planned deployment of urban and rural $PM_{2.5}$ speciation sites in the United States is shown by site type in Figure 3.6. In addition, EPA's recent Supersite program included time-resolved (averaged over ~10 minutes to 1 hour) $PM_{2.5}$ mass and chemical-speciated measurement data (eg, sulfate, nitrate, and organic compounds). As a result, several new continuous and semicontinuous PM

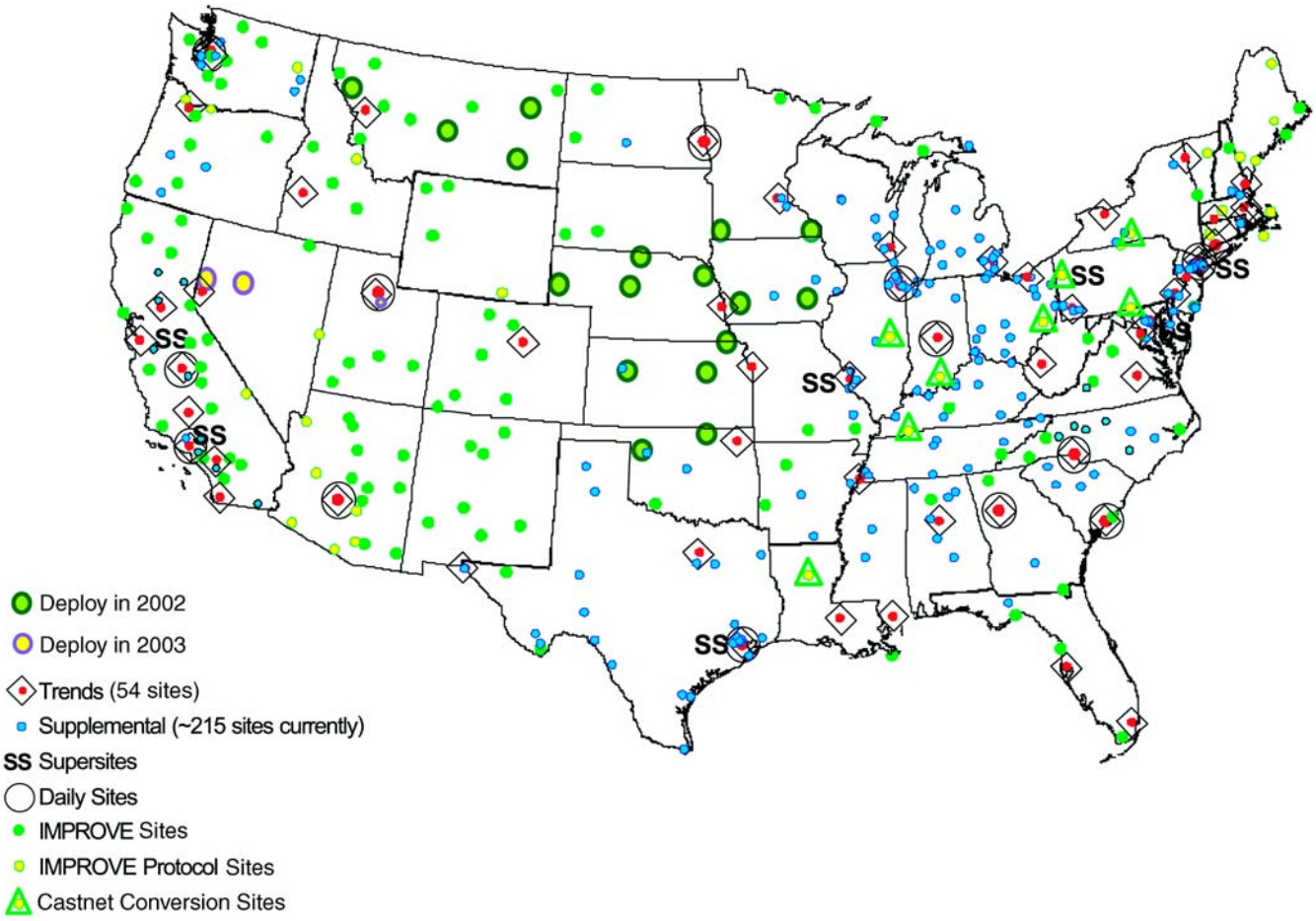


Figure 3.6. Current and planned speciation sites in the US $PM_{2.5}$ network. All sites were operating as of January 2002 except those marked "Deploy in 2002" or "Deploy in 2003". Adapted from EPA 2002d.

measurement technologies have been evaluated for possible application in routine monitoring networks (Drewnick et al 2003). Such time-resolved PM data capture diurnal patterns and extreme events, thus offering the possibility of enhancing estimates of population exposure.

These more sophisticated data facilitate rethinking the role that monitoring and analysis can play in the air quality management process and provides valuable new tools for accountability assessments. Some specific analyses of interest include (1) tracking ambient effects of changes in emissions for certain chemical species to judge the validity of modeled projections and the effectiveness of corresponding control programs; (2) corroboration of emission inventories; and (3) attributing source contributions to ambient levels of certain chemical species.

The expansion of speciation monitoring will provide improved tools for source attribution, or apportionment, analyses. Source apportionment involves applying multivariate statistical models to time-series records of ambient monitoring data to extract data signals that can be associated with specific sources. Source-apportionment analyses are improved when air monitoring data are available for a wider range of pollutant species (eg, elemental components of PM, VOCs, molecular markers) and when some of these species can be linked to the emissions signatures of specific source classes. Because source apportionment can identify and quantify proportional contributions of specific source types to ambient air quality, it holds promise as a tool for accountability studies as well as source-oriented studies of health effects.

FROM AMBIENT AIR QUALITY TO EXPOSURE OR DOSE

Policies to improve health of the population by controlling emissions will be successful only if the emissions reductions ultimately result in reduced population exposures and doses to the air pollutant of concern. To benefit public health, health-based air regulations must reduce not only ambient concentrations but also population exposures and doses.

It has long been recognized that ambient air pollution concentrations do not necessarily represent the levels of air pollution to which people are actually exposed. Factors that influence this relation include the spatial and temporal patterns of ambient concentrations, the ways in which these patterns intersect with time-activity patterns of individuals, and sources of pollution in indoor environments where people spend time. Most people in developed western countries spend the majority of their time indoors, where ambient pollutants may penetrate incompletely and where numerous indoor pollution sources exist (Sexton and Ryan 1988). In addition, activity patterns may change in response to improving air quality, with more people going outdoors and engaging in physical activity. Although reducing an individual's exposure always reduces dose, the dose (or body burden) of a pollutant that results from a given exposure level can be expected to vary among people due to differences in physical-activity levels, age, preexisting health status, physiology, and metabolic factors (Janssen et al 1999, 2000; Ebelt et al 2000; Sarnat et al 2000). Furthermore, the concentration of a chemical biomarker found in an accessible biological sample (eg, blood) may not represent the dose to the critical target tissue (eg, lung epithelial cells) (Xu and Yu 1986; Bennett et al 1996, 1997; Kim and Hu 1998).

Most ambient pollutants penetrate indoors only partially, with the penetration efficiency depending on pollutant type and building characteristics (eg, the use of air conditioning). Because of its reactive nature, O_3 is generally present at low concentrations indoors, with indoor/outdoor ratios toward the low end of the range of 0.1 to 0.8, depending on the degree of natural ventilation (penetration is greatest when windows are open). On the other hand, outdoor $PM_{2.5}$ particles appear to penetrate indoors more readily than O_3 (Ozkaynak and Spengler 1996). Penetration is greater with higher levels of building ventilation (eg, with open windows). In addition, the indoor environment contains sources of many pollutants, including $PM_{2.5}$, NO_x, and VOCs. As a result, higher concentrations of many pollutants occur indoors than outdoors, especially in buildings tightly sealed for energy efficiency.

For some pollutants, the usefulness of ambient monitoring data for representing population exposures is limited by the number of monitors and the difficulty of capturing small-scale patterns in ambient concentrations due to local sources (eg, the effect of heavy traffic roadways on nearby residents). The extent to which a central-site monitor can characterize individual or population-average exposures in an area depends on spatial variability in ambient concentrations, time-activity patterns, building ventilation characteristics, and other factors. In order to fully characterize exposures of the population, ambient monitoring data can be supplemented with personal monitoring (ie, monitoring exposure of individual people).

Basing accountability analyses on ambient data may not be appropriate in some situations, such as when estimating risks for subgroups of a diverse population that varies in time-activity patterns, housing characteristics, and exposures to indoor air pollution sources. Estimates of exposures or doses based on ambient data may not capture the full range of doses; some individuals may be subjected to levels of risk for health outcomes generally considered unacceptable, even though the average risk for the population appears acceptable. For some pollutants, measurements of personal exposure may be possible and may provide a better picture of the distribution of exposure than estimates based on central-site monitoring data. For example, convenient personal samplers are available for some pollutants (eg, passive samplers for NO_2 and VOCs); for other pollutants, biomarkers of dose can be measured (eg, lead in blood, VOCs in blood and exhaled breath). For still other pollutants, including O_3 and $PM_{2.5}$, biomarkers are not available; personal-exposure measurements for these require equipment that is somewhat expensive or cumbersome, limiting widespread use.

$PM_{2.5}$ as a bulk pollutant appears to have several unique properties that tend to improve the correlation between ambient and personal exposures. Spatial variation in ambient $PM_{2.5}$ concentrations within urban areas is relatively smooth, so that a central site is reasonably representative of ambient concentrations throughout a city (Suh et al 1995; Kinney et al 2000). In addition, $PM_{2.5}$ penetrates indoors more readily than other pollutants such as O_3 (Wilson and Suh 1997; Abt et al 2000). Despite many indoor sources of $PM_{2.5}$, studies have demonstrated high correlations over time between ambient $PM_{2.5}$ and personal exposures to $PM_{2.5}$ (Janssen et al 1999, 2000; Abt et al 2000; Sarnat et al 2000). Finally, health risks associated with $PM_{2.5}$ have been characterized mainly by epidemiologic studies that used ambient concentrations to derive concentration-response functions. In calculating risks based on the epidemiologic evidence, ambient concentrations provide the most appropriate measure of population exposure associated with undifferentiated $PM_{2.5}$ mass and some of its main components (eg, sulfate and nitrate). Exposures to

other components of $PM_{2.5}$ (such as elemental carbon) that exhibit small-scale spatial variation in ambient concentrations in urban areas may not be as well characterized by ambient concentrations measured at a single location (Kinney et al 2000).

Exposure models can be used if neither measurements nor simple assumptions are sufficient to link ambient concentrations to exposure and dose. These models incorporate empirically based assumptions regarding activity patterns, outdoor to indoor penetration factors, decay of indoor concentrations, human breathing rates, lung deposition, and other factors. The models can then be used to simulate the population-wide distribution of exposures or doses to a particular pollutant. For example, EPA has used the ASPIN model to estimate census-tract level ambient VOC concentrations, personal exposures, and doses across the US (Rosenbaum et al 1999; EPA 2000b).

Input data on population-activity patterns are available from several large-scale time-activity surveys, including the National Human Activity Pattern Survey (NHAPS) that was conducted with EPA support in the early 1990s. In this survey, data were collected on activity patterns for 9386 subjects over 24 hours (EPA 1996).

Assessing Population Exposure and Dose

Several general strategies can be used to measure exposure and dose. One involves the use of large-scale, periodic, random monitoring surveys of the general population to document long-term trends in exposure and dose. On such a large scale, personal monitoring would be constrained by cost considerations as well as the need to keep the burden on participants as low as possible. When practical, blood or other biospecimens could be collected for analysis. Another strategy involves smaller-scale studies of specific subpopulations to document changes in exposure and dose before and after specific interventions. This strategy could yield richer data on personal exposure and biological dose measurements, but in a smaller population.

An excellent example of the large-scale survey design is the recent US National Health and Nutrition Examination Survey (NHANES) (US National Center for Health Statistics 2003). The most recent NHANES included measurement of personal exposures to and blood levels of VOCs among a stratified random sample of 1000 US adults (20–59 years of age). Personal exposures were measured using small, lightweight, passive-diffusion badges worn for 48 hours. The badges were analyzed for several VOC compounds (Table 3.2). Before and after the sampling period, a small blood sample was collected for analysis of a wide range of VOCs (Table 3.2). Tap water samples were also collected. According to the investigators (US Centers for Disease Control and Prevention 2003),

> the data will be used to: (1) characterize the distribution of personal exposures to selected volatile organic compounds; (2) characterize the distribution of blood levels of selected volatile organic compounds; (3) characterize the distribution of levels of selected volatile organic compounds in home tap water samples; (4) examine the relationship between personal exposure measures and blood levels and the relationship between water levels and blood levels of selected volatile organic compounds; (5) examine the relationship between measures of volatile organic compounds and demographic, economic, and behavioral characteristics; and (6) investigate possible associations between measures of volatile organic compounds and selected measures of health status.

Because of its large size, national scope, and representative sampling design, NHANES represents a milestone in

Table 3.2. VOC Analytes in Air and Blood Collected in NHANES

Air Analytes	Blood Analytes
Benzene	1,1,1-Trichloroethane
1,3-Butadiene	1,1,2,2-Tetrachloroethane
Carbon tetrachloride	1,1,2-Trichloroethane
Chloroform	1,1-Dichloroethane
1,4-Dichlorobenzene	1,1-Dichloroethene
Ethyl benzene	1,2-Dichlorobenzene
Methylene chloride	1,2-Dichloroethane
Methyl tertiary-butyl ether	1,2-Dichloropropane
Styrene	1,3-Dichlorobenzene
Tetrachloroethylene	1,4-Dichlorobenzene
Toluene	2-Butanone
Trichloroethylene	Acetone
Xylenes	Benzene
	Bromodichloromethane
	Bromoform
	Carbon Tetrachloride
	Chlorobenzene
	Chloroform
	cis-1,2-Dichloroethene
	Dibromochloromethane
	Dibromomethane
	Ethylbenzene
	Hexachloroethane
	m-/*p*-Xylene
	Methylene chloride
	Methyl tertiary-butyl ether
	o-Xylene
	Styrene
	Tetrachloroethene
	Toluene
	trans-1,2-Dichloroethene
	Trichloroethene

exposure and dose assessment in the United States. Studies with the same design repeated over time could provide valuable information on trends in exposure and dose, which could be analyzed with respect to emission regulations. Pollutants for which new, lightweight personal monitors are developed could be measured in future surveys. In analyzing data from national surveys, population subgroups that have high exposures or doses should be identified. This information could justify oversampling of appropriate subgroups in future surveys to identify those who may be subject to unacceptable risk.

Another possible context for large-scale population surveys is phase 2 of the US National Human Exposure Assessment Study (NHEXAS) (EPA 2003c). The phase 1 pilot studies, completed in 1998 (Whitmore et al 1999), collected extensive data on blood, urine, personal air, soil, dust, food concentrations of several VOCs, pesticides, and particulate metals for over 500 subjects across three regions (Arizona, the upper midwest, and the Baltimore, Maryland, area). Phase 2, intended to be a national exposure survey, is currently being designed, pending further analysis of the data and changes to the protocol based on lessons learned in phase 1.

An optimal strategy would be to combine the technical knowledge obtained from NHEXAS phase 1 with the survey capabilities demonstrated in NHANES to design and carry out a comprehensive national survey not only of VOCs but of other agents. Given the importance of PM in current assessments of the health impact of air pollution, $PM_{2.5}$ and chemical-speciation measurements should be incorporated into personal monitoring surveys. This step first requires investing in improved technology for monitoring personal exposures to particles, including development of miniature, lightweight, battery-powered samplers.

Carefully designed and targeted small-scale field studies of personal exposures or dose have the potential to play a critical role in assessing interventions directed at reducing pollution and improving health. Such small-scale studies would complement large-scale population surveys that aim to track long-term trends. In contrast to national surveys, which require long-term, centralized, stable staffing and funding (perhaps best provided by federal agencies), smaller targeted studies can be designed to address specific interventions and can be conducted with more modest time and funding resources. These studies can be carried out more quickly in response to emerging needs. Studies that target specific interventions must, however, incorporate a sufficiently extensive and specific set of pollutant measurements so that consequences of the intervention can be separated from changes in exposure due to other, uncontrollable factors.

Behavioral Responses to Air Quality

Regulatory action to reduce air pollution concentrations may induce changes in behavior that affect exposure and dose. For example, if people perceive that the air is cleaner, they may spend more time outside or become more active. Such behavioral changes could either augment the direct effects of ambient air quality improvements on exposure (eg, by reducing exposure to indoor air pollution), or increase air pollution exposure or dose (eg, by promoting greater activity outdoors; by increasing pulmonary ventilation in conditions of reduced, but still elevated, pollution concentrations; or by encouraging driving for leisure and vacations [which would increase mobile-source emissions]). Over extended periods of time, real or perceived changes in air quality may induce demographic changes that also affect air quality, such as increased migration to or development of so-called clean areas. Air pollution alert programs that are implemented as part of regulatory action to improve air quality clearly have the potential to alter exposure to air pollution, as indeed they are intended to do.

Surveys such as those discussed above may provide some of the data that will be necessary to assess the effect of behavioral changes induced by regulatory activity on exposure.

PLANNING AHEAD: OPPORTUNITIES FOR ACCOUNTABILITY

Retrospective estimation of the health impact of regulations implemented over long time periods (such as the US Clean Air Act) or over shorter intervals (such as banning the sale of coal in Dublin [Clancy et al 2002]; see Chapter 4) are of necessity based on whatever data are available after the fact. In our view, accountability studies will be facilitated by anticipating promulgation of regulations and including prospective data collection in anticipation of possible accountability studies.

LONG-TERM INITIATIVES

Several major regulatory initiatives are planned or have recently been initiated on a national level in the United States. These initiatives are promising targets for accountability research because years will pass before they are implemented and their effects on air quality are manifest. Future research planning will need to evaluate how these

opportunities can best be exploited to address situations in which the greatest uncertainties intersect with the largest potential health risks.

Heavy Duty Diesel/Low Sulfur Fuel Rule

EPA recently promulgated regulations to reduce heavy-duty diesel vehicle emissions via reductions beginning in 2006 and 2007 in fuel sulfur combined with particle trap technology. The rulemaking involved extensive analyses of technological feasibility and costs of the proposed regulations and their probable benefits (in terms of reduced $PM_{2.5}$ and ozone levels) and associated health risks over the 30-year period after implementation. The rule therefore could provide a useful context in which accountability could be judged. Both costs and benefits of regulations could be tracked over time during the implementation phase and could then be compared against the costs and benefits predicted by regulatory impact analysis. The extent to which future benefits would indeed be measurable (against shifting air quality baselines) is not at all clear. However, certain links in the chain of accountability (Figure 3.1) would probably be amenable to analysis, especially the link between regulatory action and emissions.

California Diesel Risk Reduction Plan

The state of California aims to reduce diesel PM emissions by 75% by the end of the decade by using particle traps, low-sulfur fuels, advanced engine technologies, and alternative fuels (California Air Resources Board 2000). A number of these approaches to diesel PM reduction (such as retrofits of the existing fleets, buses fueled by clean natural gas, and roadside inspections) involve discrete changes in emissions over relatively short time frames and are being implemented in poor communities that currently experience the greatest exposures.

Implementing New $PM_{2.5}$ and O_3 Standards

The United States is beginning the implementation phase of the new $PM_{2.5}$ and O_3 NAAQS. Each state has begun to develop a SIP designed to bring ambient concentrations into compliance with the annual and daily PM standards (15 and 65 $\mu g/m^3$, respectively) and the new 8-hour ozone standard (0.08 ppm) over several years (EPA 1997). This phase represents an opportunity for both large-scale surveys and target studies aimed at documenting changes in emissions, ambient concentrations, and exposures or doses resulting from these new policies. The SIP process is now in its initial stages. Meanwhile, extensive data on nationwide $PM_{2.5}$ and O_3 concentrations are now available from a new $PM_{2.5}$ monitoring network and the existing ozone monitoring network. These are baseline data against which future emissions reductions can be assessed. Additional monitoring undertaken as part of studies intended to measure the health impact of SIPs must be in accord with the temporal and spatial scales of the SIPs themselves.

SIP NO_x Reduction Plan (NO_x SIP Call)

A 25% reduction in NO_x is planned for approximately 19 states in the eastern and midwestern United States: regulations are set to be implemented in 2004 with a target for compliance by 2007. This reduction of NO_x emissions from major point sources is directed mainly at mitigating regional ozone but will also reduce acid deposition and PM-nitrate (NO_3). The resulting regional perturbations in air quality (in this case, ozone, nitric acid, and PM-NO_3) can and should be tracked, but changes in specific health outcomes may be subtle and difficult to measure. Opportunities to correlate air quality changes to specific ecosystem responses may also be a possibility.

EPA Air Toxics Control Plan

Hazardous air pollutants, also termed *air toxics*, are pollutants that cause or may cause cancer or other serious health problems (such as reproductive disorders or birth defects) or adverse environmental and ecologic effects. EPA is required to control 188 hazardous air pollutants, according to approaches recommended by Congress in the 1990 Clean Air Act and its Amendments and as part of EPA's mandate to reduce the daily emissions of air toxics, EPA first uses a technology-based approach to reduce emissions of air toxics from major sources of air pollution, followed by a risk-based approach to address any remaining, or residual, health risks. Under the technology-based approach, EPA is required to develop standards, known as *maximum achievable control technology* (MACT) standards, for controlling routine emissions of air toxics from each major type of facility within an industry group (or source category). MACT standards are based on emissions levels that are already being achieved by the better-controlled and lower-emitting sources in industry. Eight years after each MACT standard is issued, EPA must assess the remaining health risks from source categories and implement additional MACT standards necessary to address any considerable remaining risk.

Over the past seven years, EPA has issued MACT standards covering over 80 categories of major industrial sources (such as chemical plants, oil refineries, aerospace manufacturers, and steel mills) as well as categories of

smaller sources (such as dry cleaners, commercial sterilizers, secondary lead smelters, and chromium electroplating facilities). When fully implemented, these standards are projected to reduce annual air toxics emissions by about 1.5 million tons—15 times the reductions achieved before 1990 (EPA 2000c).

EPA has also put into place important controls for fuels and vehicles that are expected to reduce selected motor vehicle air toxics by more than 75% (relative to 1990 levels) by 2020. Finally, EPA has implemented programs that reduce indoor air toxics (EPA 2000c).

SHORT-TERM INITIATIVES

Relatively rapid changes in ambient concentrations may occur in a localized area as a result of a major change in local source emissions due to regulatory action (eg, closing downtown streets to traffic, installing new emission controls on a fleet of diesel trucks, converting a bus-fueling depot from diesel to natural gas, closing down or imposing strict controls on a large power plant). Such interventions present opportunities for studies aimed at documenting cause-effect relations between emissions changes and exposure or health changes. The relatively compressed temporal and spatial scales that characterize many of these local interventions make assessment studies both economically and logistically feasible.

Designing and implementing studies of short-term initiatives is challenging, however. A control group is desirable, to ensure that any observed temporal changes in exposure are related to the specific emission change under study rather than to changes in other factors (including seasonal changes). Control for other factors could be achieved by an assessment carried out concurrently in another area in which emission changes are not implemented. Another challenge is to design an exposure measurement study to capture the impact of the intervention. In urban areas, where city-wide concentrations of many air toxics are likely to be elevated (Kinney et al 2002), the impact on exposure of a change in one emission source may be difficult to measure against the urban background. Ideally the study design would include documenting the distribution of ambient impact as a function of distance from the source, with measurements conducted both before and after intervention. This design is complicated further for personal exposures, for which population mobility (among geographic regions, indoor to outdoor) may smooth the impact of ambient changes induced by the intervention but also may introduce other, undocumented sources of exposure. Still, in spite of these challenges, well-designed studies can provide a cost-effective mechanism for documenting the effectiveness of a wide range of small-scale air quality interventions.

Numerous opportunities for such studies exist worldwide. For example, in New York City, the Metropolitan Transit Authority plans to convert bus fueling and storage depots from diesel to natural gas, which could reduce neighborhood levels of elemental carbon and other diesel-related particle components (Office of the Governor 1997). Similarly, low-sulfur diesel fuel and particle trap control technology is being implemented in fleets of diesel vehicles in many locations (including garbage trucks and school buses) (California Air Resources Board 2000). Recently, municipal authorities in London have proposed a series of interventions designed to reduce the impact of vehicular traffic on the general urban environment (and on air pollution levels in particular) (Mayor of London 2002). The most recent in this series involved the levying of a tax on any vehicle entering central London at certain times. Studies that document changes in ambient concentrations, personal exposures, and even health status in connection with such interventions could be quite informative. To successfully carry out such studies, however, these actions need to be anticipated and almost certainly the studies need to be planned in cooperation with the authorities who are implementing the intervention.

SPECIFIC RESEARCH NEEDS AND OPPORTUNITIES

In the US regulatory and public health programs have led to the creation of extensive national databases. These databases could be more effectively exploited to better document the performance of regulatory interventions along the chain of accountability.

Tracking Changes in Emissions in Response to Regulation

The EPA compiles source-specific emission factor data for a variety of chemical species and estimates and reports on the trends in these data on an annual basis (EPA 2002a). Studies are needed to assess the relations between the implementation of regulations and the measured emissions changes that result.

The SIP process has many elements that mesh with the goals of accountability analyses, although its potential has been largely untapped to date. These elements include modeled predictions of ambient air quality under planned emissions control programs and extensive ambient monitoring over space and time. An historical record of prescribed emissions reductions, model-based projections of future air quality, and observed air quality is now available. A focused

effort to analyze such data could be informative with respect to the effectiveness of the SIP planning process.

Tracking Changes in Nature of Air Pollution

Speciated monitoring data for O_3 and PM precursor pollutants and PM subcomponents are increasingly abundant (Albritton and Greenbaum 1998; EPA 2003b; also see *www.epa.gov/ttn/amtic/files/ambient/pm25/casac/supersum.pdf*). Accountability studies can and should be designed to take advantage of these data. Examples of specific analyses of interest include: (1) tracking the ambient impact of emissions changes for specific chemical species to judge the validity of modeled projections and the effectiveness of corresponding control programs; (2) corroborating emission inventories; and (3) attributing source contributions to ambient levels of certain species.

Tracking Changes in Human Exposure

Because of its large size, national scope, and representative sampling design, NHANES represents a milestone in exposure and dose assessment in the United States. Development of new exposure measurement tools is encouraged for incorporation into future NHANESs. Given the importance of PM in current assessments of the health impact of air pollution, incorporating $PM_{2.5}$ and chemical speciation data into future personal monitoring surveys like NHANES would strengthen assessments of accountability. Doing so would first require a pilot demonstration of new technology for personal particle monitoring, including development of miniature, lightweight, battery-powered samplers capable of accumulating sufficient material for speciation analyses.

REFERENCES

Abt E, Suh HH, Catalano P, Koutrakis P. 2000. The relative contribution of outdoor and indoor particle sources to indoor concentrations. Environ Sci Technol 34:3579–3587.

Albritton DL, Greenbaum DS, cochairs. 1998. Atmospheric Observations: Helping Build the Scientific Basis for Decisions Related to Airborne Particulate Matter; Report of the PM Measurements Research Workshop, Chapel Hill NC, July 22 and 23, 1998. Aeronomy Laboratory of the National Oceanic and Atmospheric Administration, Boulder CO, and Health Effects Institute, Cambridge MA.

Bennett WD, Zeman KL, Kim C. 1996. Variability of fine particle deposition in healthy adults: Effect of age and gender. Am J Respir Crit Care Med 153:1641–1647.

Bennett WD, Zeman KL, Kim C, Mascarella J. 1997. Enhanced deposition of fine particles in COPD patients spontaneously breathing at rest. Inhalation Toxicol 9:1–14.

California Air Resources Board. 2000. Risk Reduction Plan to Reduce Particulate Matter Emissions from Diesel-Fueled Engines and Vehicles. California Air Resources Board, Sacramento CA. Available from *www.arb.ca.gov/diesel/documents/rrpFinal.pdf.*

Centers for Disease Control and Prevention (US). 2003. Volatile organic compounds exposure monitoring: Public health objective (last updated 1/1/03). *www.cdc.gov/nchs/data/nhanes/specstudy.pdf.* Accessed 8/29/03.

Clancy L, Goodman P, Sinclair H, Dockery DW. 2002. Effect of air pollution control on death rates in Dublin, Ireland: An intervention study. Lancet 360:1210–1214.

Demerjian KL. 2000. A review of national monitoring networks in North America. Atmos Environ 34: 1861–1884.

Drewnick F, Schwab JJ, Hogrefe O, Peters S, Husain L, Diamond D, Weber R, Demerjian KL. 2003. Intercomparison and evaluation of four semi-continuous $PM_{2.5}$ sulfate instruments. Atmos Environ 37:3335–3350.

Ebelt ST, Petkau AJ, Vedal S, Fisher TV, Brauer M. 2000. Exposure of chronic obstructive pulmonary disease patients to particulate matter: Relationships between personal and ambient air concentrations. J Air Waste Manage Assoc 50:1081–1094.

Environmental Protection Agency (US). 1996. Analysis of the National Human Activity Pattern Survey (NHAPS) Respondents from a Standpoint of Exposure Assessment. EPA/600/R-96/074. National Exposure Research Laboratory, Research Triangle Park NC.

Environmental Protection Agency (US). 1997. National Ambient Air Quality Standards for Particulate Matter; Final Rule. 40 CFR, Part 50. Fed Regist 62:38651–38760.

Environmental Protection Agency (US). 1998a. National Air Pollutant Emission Trends. Procedures Document 1900–1996. EPA-454/R-98-008. Office of Air Quality Planning and Standards, Research Triangle Park NC.

Environmental Protection Agency (US). 1998b. Analysis of Photochemical Assessment Monitoring Station (PAMS) Data to Evaluate Reformulated Gasoline (RFG) Effect. Final Report STI-997350-1774-FR2. Office of Mobile Sources, Washington DC.

Environmental Protection Agency (US). 1998c. Implementation Plan: $PM_{2.5}$ Monitoring Program. Emissions, Monitoring and Analysis Division, Research Triangle Park NC. Available from *www.epa.gov/ttn/amtic/files/ambient/pm25/pmplan3.pdf*.

Environmental Protection Agency (US). 1999. Achieving Clean Air and Clean Water: The Report of the Blue Ribbon Panel on Oxygenates in Gasoline. EPA/420/R-99-021. Blue Ribbon Panel for Reviewing Use of MTBE, Office of Transportation and Air Quality, Washington DC.

Environmental Protection Agency (US). 2000a. Update: $PM_{2.5}$ Monitoring Implementation. Emissions, Monitoring and Analysis Division, Research Triangle Park NC. Available from *www.epa.gov/ttn/amtic/files/ambient/pm25/exec.pdf*.

Environmental Protection Agency (US). 2000b. User's Guide for the Assessment System for Population Exposure Nationwide (ASPEN, Version 1.1) Model. EPA-454/R-00-017. Office of Air Quality Planning and Standards, Research Triangle Park NC.

Environmental Protection Agency (US). 2000c. Taking Toxics Out of the Air. EPA-452/K-00-002. Office of Air Quality Planning and Standards, Research Triangle Park NC. Available from *www.epa.gov/oar/oaqps/takingtoxics/airtox.pdf*.

Environmental Protection Agency (US). 2002a. Latest Findings on National Air Quality: 2001 Status and Trends. EPA 454/K-02-001. Office of Air Quality Planning and Standards, Research Triangle Park NC. Available from *www.epa.gov/air/aqtrnd01/*.

Environmental Protection Agency (US). 2002b. Trends in Sulfur Dioxide Emissions Following Implementation of Phase I of the Acid Rain Program: Total State-level Utility SO_2 (1980, 1990, 1999). Clean Air Markets Division, Washington DC. Available from *www.epa.gov/airmarkt/cmap/mapgallery/mg_so2_before_and_aft.html*.

Environmental Protection Agency (US). 2002c. Comparison of Ambient Sulfur Dioxide Concentrations in the Eastern United States from CASTNet Monitoring Data (1989–1991 vs. 1997–1999). Clean Air Markets Division, Washington DC. Available from *www.epa.gov/airmarkt/cmap/mapgallery/mg_so2.html*.

Environmental Protection Agency (US). 2002d. Monitor Network and IMPROVE Protocol Network Maps. Office of Air and Radiation, Washington DC. Available from *www.epa.gov/ttn/amtic/pmspec.html*.

Environmental Protection Agency (US). 2003a. Enhanced ozone monitoring (PAMS) (last updated 7/7/03). *www.epa.gov/oar/oaqps/pams/*. Accessed 8/22/03.

Environmental Protection Agency (US). 2003b. PM Supersites information (last updated 5/15/03). *www.epa.gov/ttn/amtic/supersites.html*. Accessed 8/22/03.

Environmental Protection Agency (US). 2003c. Human exposure measurements: National Human Exposure Assessment Survey (NHEXAS) (last updated 7/21/03). *www.epa.gov/heasd/edrb/nhexas.htm*. Accessed 8/21/03.

Friedman MS, Powell KE, Hutwagner L, Graham LM, Teague WG. 2001. Impact of changes in transportation and commuting behaviors during the 1996 Summer Olympic Games in Atlanta on air quality and childhood asthma. JAMA 285:897–905.

Holmén BA, Ayala A. 2002. Ultrafine PM emissions from natural gas, oxidation-catalyst diesel, and particle-trap diesel heavy-duty transit buses. Environ Sci Technol 36:5041–5050.

Janssen NAH, De Hartog JJ, Hoek G, Brunekreef B, Lanki T, Timonen KL, Pekkanen J. 2000. Personal exposure to fine particulate matter in elderly subjects: Relation between personal, indoor, and outdoor concentrations. J Air Waste Manage Assoc 50:1133–1143.

Janssen NAH, Hoek G, Harssema H, Brunekreef B. 1999. Personal exposure to fine particles in children correlates closely with ambient fine particles. Arch Environ Health 54:95–101.

Kim CS, Hu SC. 1998. Regional deposition of inhaled particles in human lungs: Comparison between men and women. J Appl Physiol 84:1834–1844.

Kinney PL, Aggarwal M, Northridge ME, Janssen NAH, Shepard P. 2000. Airborne concentrations of $PM_{2.5}$ and diesel exhaust particles on Harlem sidewalks: A community-based pilot study. Environ Health Perspect 108:213–218.

Kinney PL, Chillrud SN, Ramstrom S, Ross J, Spengler JD. 2002. Exposures to multiple air toxics in New York City. Environ Health Perspect 110(Suppl 4):539–546.

Malm WC, Pitchford ML, Scruggs M, Sisler JF, Ames R, Copeland S, Gebhart KA, Day DE. 2000. Spatial and Seasonal Patterns and Temporal Variability of Haze and its

Constituents in the United States. Report III. ISSN 0737-5352-47. Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins CO.

Mayor of London. 2002. Cleaning London's Air: The Mayor's Air Quality Strategy. Greater London Authority, London, United Kingdom. Available from *www.london.gov.uk/approot/mayor/strategies/air_quality/index.jsp*.

National Academy of Sciences (US). 1991. Rethinking the Ozone Problem in Urban and Regional Air Pollution. National Academy Press, Washington DC.

National Atmospheric Deposition Program (US). 2002. National Atmospheric Deposition Program 2001 Annual Summary. NADP Data Report 2002-01. Illinois State Water Survey, Champaign IL.

National Center for Health Statistics (US). 2003. National Health and Nutrition Examination Survey (last updated 8/19/03). *www.cdc.gov/nchs/nhanes.htm*. Accessed 8/21/03.

National Research Council (US). 1999. Research Priorities for Airborne Particulate Matter II: Evaluating Research Progress and Updating the Portfolio. National Academy Press, Washington DC.

National Research Council (US). 2002. Estimating the Public Health Benefits of Proposed Air Pollution Regulations. National Academy Press, Washington DC.

North American Research Strategy for Tropospheric Ozone (NARSTO) Synthesis Team. 2000. An Assessment of Tropospheric Ozone Pollution. NARSTO Management Office (Envair), Pasco WA. Available from *www.cgenv.com/Narsto/*.

Office of the Governor (New York State). 1997. Governor Pataki announces clean fuel bus plan for NYC [press release]. Office of the Governor, Albany NY. Available from *www.state.ny.us/governor/press/jan29_3.html*.

Ozkaynak H, Spengler J. 1996. The role of outdoor particulate matter in assessing total human exposure. In: Particles in Our Air: Concentrations and Health Effects (Wilson R, Spengler JD, eds), pp 63–84. Harvard University Press, Cambridge MA.

Rosenbaum AS, Axelrad DA, Woodruff TJ, Wei YH, Ligocki MP, Cohen JP. 1999. National estimates of outdoor air toxics concentrations. J Air Waste Manage Assoc 49:1138–1152.

Roth PM. 1999. A qualitative approach to evaluating the anticipated reliability of a photochemical air quality simulation model for a selected application. J Air Waste Manage Assoc 49:1050–1059.

Russell A, Dennis R. 2000. NARSTO critical review of photochemical models and modeling. Atmos Environ 34:2283–2324.

Sarnat JA, Koutrakis P, Suh HH. 2000. Assessing the relationship between personal particulate and gaseous exposures of senior citizens living in Baltimore, MD. J Air Waste Manage Assoc 50:1184–1198.

Sawyer RF, Harley RA, Cadle SH, Norbeck JM, Slott R, Bravo HA. 2000. Mobile sources critical review: 1998 NARSTO assessment. Atmos Environ 34: 2161–2181.

Sexton K, Ryan PB. 1988. Assessment of human exposure to air pollution: Methods, measurements, and models. In: Air Pollution, the Automobile, and Public Health (from the Health Effects Institute; Watson AY, Bates RR, Kennedy D, eds), pp 207–238. National Academy Press, Washington DC.

Suh HH, Allen GA, Koutrakis P, Burton RM. 1995. Spatial variation in acidic sulfate and ammonia concentrations within metropolitan Philadelphia. J Air Waste Manage Assoc 45:442–452.

Whitmore RW, Byron MZ, Clayton CA, Thomas KW, Zelon HS, Pellizzari ED, Lioy PJ, Quackenboss JJ. 1999. Sampling design, response rates, and analysis weights for the National Human Exposure Assessment Survey (NHEXAS) in EPA Region 5. J Expos Anal Environ Epidemiol 9:369–380.

Wilson WE, Suh HH. 1997. Fine particles and coarse particles: Concentration relationships relevant to epidemiologic studies. J Air Waste Manage Assoc 47:1238–1249.

Xu GB, Yu CP. 1986. Effects of age on deposition of inhaled aerosols in the human lung. Aerosol Sci Technol 5:349–357.

Zhang Y, Stedman DH, Bishop GA, Guenther PL, Beaton SP, Peterson JE. 1993. On-road hydrocarbon remote sensing in the Denver area. Environ Sci Technol 27:1885–1891.

ABBREVIATIONS AND OTHER TERMS

CO	carbon monoxide
EPA	Environmental Protection Agency (US)
FEM	federal equivalent method
FRM	federal reference method
IMPROVE	Interagency Monitoring of Protected Visual Environments
MACT	maximum achievable control technology
NAAQS	National Ambient Air Quality Standard(s) (US)
NAMS	National Air Monitoring Stations (US)
NARSTO	North American Research Strategy for Tropospheric Ozone
NHANES	National Health and Nutrition Examination Survey (US)
NHEXAS	National Human Exposure Assessment Study (US)
NO_2	nitrogen dioxide
NO_x	oxides of nitrogen
O_3	ozone
PAMS	photochemical assessment monitoring station
PM	particulate matter
PM_{10}	PM less than 10 μm in aerodynamic diameter
$PM_{2.5}$	PM less than 2.5 μm in aerodynamic diameter
SIP	state implementation plan
SO_2	sulfur dioxide
VOC	volatile organic compound

Chapter 4. HEALTH OUTCOMES FOR MEASURING ACCOUNTABILITY

SELECTING OUTCOMES

Air quality regulations are established with the primary purpose of protecting the public's health. Regulatory action is taken on the basis of evidence that indicates a causal association between exposure to air pollution and health risk. Thus, studies that provide more definitive evidence concerning the causality of an association between exposure and a specific health outcome strengthen the rational basis for policy making. The outcomes considered in assessments of accountability should reflect the evidence on which regulation is based. For estimating the impact of specific regulations, however, certain outcomes may be preferred.

GENERAL CONSIDERATIONS

Consider Goals of the Regulation

The first and most obvious step in selecting outcomes for evaluating the impact of a regulation is to consider the objectives of the regulation itself. For the pollutants for which US National Ambient Air Quality Standards (NAAQS*) are set in the United States, the language of the Clean Air Act specifies these objectives. Demonstrating that the public health objectives of air quality regulations have been met requires estimating changes in health outcomes that are often difficult to measure and whose specificity may be low due to the many factors besides air pollution that cause them. A range of measurable health outcomes is described in subsequent sections. Few, if any, of these outcomes are specific to a single pollutant class or even to air pollution generally. Therefore, isolating the burden of the total outcome (ie, burden of an adverse health outcome) attributable to a pollutant, or changes in outcome attributable to changes in pollutant exposure, is difficult. This lack of specificity contributes to a principal difficulty of studying any outcome: disentangling the beneficial effect of reduced pollutant exposure from the net sum of the beneficial and detrimental effects of other factors that also determine the risk of the outcome.

Consider Outcomes That Drove Promulgation of Regulation

Evidence-based regulations have their primary rationale in health-related studies that point to associations consistent with a causal relation between pollution and risk of the health outcome. Therefore, a key step in selecting outcomes for accountability purposes is identifying those outcomes that drove promulgation of the regulation. For example, time-series epidemiologic studies of mortality and morbidity and intercity comparisons of life span largely drove promulgation of the 1997 $PM_{2.5}$ (particulate matter [PM] less than 2.5 µm in aerodynamic diameter) NAAQS (US Environmental Protection Agency [EPA] 1996b). Although some key studies have recently undergone further analyses because of newly discovered difficulties with software used in some studies, these new analyses corroborated the association of short-term exposure to PM with daily morbidity and mortality (HEI 2003).

Consider Outcomes Identified by Regulators

Promulgation of regulations may be accompanied by an analysis of anticipated benefits, with emphasis on health outcomes most relevant to the specific pollutant(s). The outcomes thus identified would certainly be key targets for accountability assessments. The EPA develops a regulatory impact analyses (RIAs) when promulgating a new or revised standard. Each RIA includes a cost-benefit analysis, which necessitates stating the nature of the projected impact and, implicitly, the outcomes by which the impact can be measured. The RIA developed for the 1997 revisions of ozone and PM NAAQS and the proposed regional haze rule is an example (EPA 1997b). Several types of health outcomes were used in the PM RIA cost-benefit analysis, including short- and long-term average mortality, hospital admissions for certain disorders, incident cases of certain diseases and symptoms, and lost work days. Although an accountability assessment for PM need not be restricted to these outcomes, investigating them is an appropriate starting point because it affords an opportunity to validate the RIA.

* A list of abbreviations and other terms appears at the end of this chapter.

This chapter is one part of Health Effects Institute Communication 11, *Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research,* which also includes five other chapters and an Executive Summary of the project.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R82811201 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.

Consider Methodologic Practicalities

A number of practical considerations necessarily enter into the selection of outcomes for assessing the impact of air quality regulations. An obvious consideration is the availability of data needed to assess adequately the success or benefit of a regulation. Data needed to establish a preregulation baseline may not have been collected, or a baseline may need to be established, in part to determine initial compliance. For example, in implementing the new NAAQS for $PM_{2.5}$, a national monitoring network was initiated in 1999 (EPA 2002). The initial data collected will establish a baseline for $PM_{2.5}$ and also will identify those places not in compliance with the new standard. In implementing the NAAQS, however, no consideration has been given to how the health impact will be tracked, either in the general population or susceptible subpopulations.

Other practical considerations include costs of data acquisition and issues related to privacy and confidentiality. For some health outcomes (as mentioned in subsequent sections and in Chapter 5), regional or national strategies for recording, collecting, and assimilating data may be necessary to acquire sufficiently high-quality and specific data. In addition, collection and analysis of surveillance data will need to be longer term than is typical in research. New or expanded nationwide tracking systems for certain health outcomes might have to be implemented to obtain the type(s) and completeness of information required. Establishing such systems may be costly and may also raise broader issues regarding government access to personal information. Establishing surveillance systems will require considerable support from the public and political will from multiple levels of government.

Assessing consequences of an implemented regulation is necessarily bounded by the constraints of available technologies. Sensitive biomarkers of injury that might be used in the laboratory may not be suitable for use in the population and appropriate subpopulations may not be readily available for study. For example, there is no PM-specific biomarker to measure DNA adduction or oxidative damage that may connect PM exposure and cancer. Additionally, any system for data collection needs to be sufficient in scope to track anticipated impact. Methods for data analysis also need to be sufficiently powerful and capable of separating the effect of environmental changes associated with the intervention from the effects of other factors.

IMPACT ASSESSMENT

Direct Observations

Evidence of the health impact of air quality regulations is the cornerstone of accountability assessment. Although evaluating whether regulations have improved air quality is itself a key component of accountability assessment (as discussed in Chapter 3), an assessment that does not offer evidence of the regulation's intended health impact is unlikely to be widely accepted. If a regulation is justified on the basis of reducing pollutant-related daily mortality, for example, most stakeholders will expect an assessment to include evidence that reduced pollutant levels were in fact accompanied by reduced mortality. Even though direct observation of such reductions may be difficult or even impossible, regulators and the public will expect that an attempt will be made to directly measure progress following a regulation.

Indirect Support

The credibility of claims about the consequences of changes in air quality can also be bolstered by indirect evidence. This evidence might enhance understanding of the mechanisms involved in the chain of accountability, thereby supporting the plausibility of claimed benefits. For example, clinical or toxicologic studies might demonstrate that mechanisms by which exposures were hypothesized to cause the effects do indeed occur, that mechanisms not previously hypothesized underlie the effects, that effects occur due to exposures to pollutants at ambient levels, or that thresholds for effects are below ambient levels.

For some environmental agents, such mechanistic information has been critical in guiding the development of regulatory standards with confidence in the anticipated benefits. For exposure to alpha radiation from radon decay, for example, understanding the mechanisms supports the conclusion that no level of exposure is biologically safe and that a public health benefit can be reasonably expected from reducing radon levels in homes (US National Research Council 1999). We propose that evidence, even indirect, that is relevant to all links of the chain of accountability should be considered.

For many environmental regulations, turning to indirect evidence to document accountability may be unavoidable. Regulations and standards have often been motivated by findings of adverse effects, and dose-response relations have been used to develop standards. Lowering population exposure would be expected to reduce risks, as described by the dose-response curve. Unless exposure is lowered substantially, however, a change in disease frequency or severity may not be detectable, and moreover, the risk per unit exposure may be unchanged if the dose-response relation is approximately linear.

One justification for the new US $PM_{2.5}$ standard was the risks implied by the exposure-response relations, or slopes, indicating certain magnitudes of health outcomes per unit of ambient mass concentration (EPA 1997a). Moreover, con-

cern for public health was heightened by apparent nonthreshold risk, either because thresholds do not exist or because studies could not identify them. Reducing population exposure by reducing ambient total $PM_{2.5}$ mass concentrations, without changing the composition of the mass, would be expected to reduce the health burden resulting from exposure but would not affect the slope describing the exposure-response relation. The most appropriate indicator of benefit from the new NAAQS standard would be an estimate of the health burden avoided by reducing population exposure, rather than an estimate of the change in risk per unit exposure.

Using the term *outcomes* in the broadest sense, even efforts that yield improved tools for assessing either the risks of exposure or the impact of reduced exposure can contribute indirect evidence to accountability assessments. Examples include improved experimental tools (such as molecular, cellular, or animal models of target populations), development of improved biomarkers of exposure or effects in the laboratory or in exposed populations, and new statistical strategies for separating the effects of exposure from effects of other factors.

Supporting Roles of Human and Animal Toxicology

Observation of changes in the frequency of relevant health outcomes after implementation of a regulation would provide the most convincing direct evidence of benefit. Epidemiologic approaches to making such observations include diverse study designs that can target many health outcomes and focus on either the general population or susceptible subpopulations. Epidemiologic observations may also add new outcome measures for accountability research (eg, adding concern about cardiovascular outcomes to the preexisting concern about respiratory outcomes of PM exposure), adding new susceptible groups (eg, demonstrating the impact on lung growth in children), or refining exposure-response estimates (eg, defining differences in potency estimates among cities). Broad-scale surveillance data collection may have an unanticipated impact.

Clinical studies involving experimental exposures of selected human groups can bridge epidemiologic studies of populations and toxicologic studies of experimental biological systems. The outcomes measured in clinical studies overlap with those used in both epidemiologic and toxicologic studies. Although measurements are necessarily noninvasive or minimally invasive, subjects in clinical studies can be more extensively tested and intensively followed and a greater range of more detailed responses can be measured than in epidemiologic studies. Because more invasive data collection is possible in the laboratory than within a free-living population, mechanisms of injury can be more readily explored. In this translational role, the clinical-study design can be used to test hypotheses in humans that have been generated by toxicologic experiments or observational epidemiology studies. Clinical studies may also be a critical source for evidence that can come only from controlled human exposures. For example, experimental exposures of humans with different respiratory tract characteristics (eg, gender, age, breathing patterns) and diseases (eg, asthma, chronic obstructive pulmonary disease [COPD]) have found differences in the total fractional deposition of inhaled PM (Kim and Kang 1997). Although toxicologic evidence indicated that such variations might exist, the initial exploratory and confirmatory work was done in humans rather than in animals (Mauderly et al 1990).

In general, contributions of toxicologic research (using biological systems other than humans) to accountability assessments parallel its contributions to promulgation of air quality regulations. Toxicologic tools encompass myriad types of biological systems (eg, animals, tissues, cells, biomolecules), methods of controlled exposure, doses, exposure patterns and times, and measurements of response. The outcomes measured may include some of those used in epidemiologic or clinical studies but also more detailed and invasive measurements. Toxicologic research gains considerable specificity and precision by predetermining study conditions, designing protocols, and limiting the variables that contribute to observed outcomes. For example, a single species or cell line can be used and dose can be carefully controlled. Toxicologic studies have the disadvantages of uncertain extension of results to humans, limited statistical power that results from small populations, and artificial conditions that result from constraining experimental variables. Exposures to real-world air pollution mixtures are not easily replicated (HEI 2002b).

Toxicologic evidence has had an important supporting role in identifying and defining relations between air quality and health in support of regulations for criteria pollutants. It will probably have a more indirect role for assessments of regulatory impact. Currently no NAAQS indicator, concentration level, or averaging time is based solely on results from toxicologic studies. Toxicologic research (especially in humans) has, however, identified and characterized hazards, suggested the nature of the dose-response curve, confirmed the biological plausibility of exposure-response associations, refined our understanding of causal mechanisms, and provided markers of exposure and effects. Studies of animals and humans exposed directly to ambient air can demonstrate effects, but this approach is unlikely to play a significant role in demonstrating the impact of regulatory actions. Toxicology has also contributed directly to regulation of Hazardous Air

Pollutants (or Air Toxics). Regulations for these pollutants are sometimes promulgated on the basis of animal studies when hazards are identified but few or no exposure-response data from humans exist or when the data are uncertain. Still, direct evidence for health benefits of regulations is much more likely to come from studies of humans than from studies of animals.

Biomarkers as Outcomes

Markers of physiologic response measured in biological samples to changes in air pollution could provide early evidence of changes in health risks and thus be used to track the impact of actions to reduce air pollution. Biomarkers have been anticipated for use in gaining insights into risks from environmental agents without needing to track consequences for the occurrence of actual disease (US National Research Council 1989), although successful examples are few. Identifying biomarkers that are specific for the exposure or health outcome of interest has been a major challenge.

Biomarkers of exposure, particularly if measured within defined population samples, have the potential to track changes in population exposures over time. For example, comparison of serum cotinine levels in two national survey populations (National Health and Nutrition Examination Surveys [NHANESs] III and IV) showed a dramatic drop in exposure of nonsmokers to tobacco smoke across the 1990s (US Centers for Disease Control and Prevention 2003a). Serum lead provides another excellent example of the effectiveness of using a validated marker (Cohen et al 1990). Markers such as DNA or protein adducts have been used as indicators of both exposure and effect, but these can rarely be interpreted as having resulted from exposure to emissions from a specific source or to specific compounds. The recent explosion of genomic and proteomic technologies has renewed hopes that readily obtained samples can provide high specificity. Because these technologies are still immature, however, their impact is unclear. Recently, analysis of constituents of exhaled breath (often collected as condensate) has once again been used to detect respiratory disorders (Paredi et al 2002). This approach could provide biomarkers that are useful for accountability purposes.

Validated biomarkers of health outcomes have the potential to predict the health impact of regulations before disease is evident. Validation has proved challenging, however, and ultimately interpretation requires relating biomarkers to the course of disease. Findings from studies using these intermediate endpoints (as well as some less adverse endpoints, such as change in forced expiratory volume in one second [FEV_1] or bronchial hyperresponsiveness) need to be interpreted in the context of what comprises an adverse effect of air pollution (American Thoracic Society 2000) and whether they function as adequate surrogates for endpoints of greater interest. One example of using a biomarker to directly evaluate the impact of an air quality regulation was reported by Wong and colleagues (1998). They evaluated changes in bronchial hyperreactivity in children after regulatory action to reduce sulfur in fuel in Hong Kong.

SELECTING OUTCOMES FOR ACCOUNTABILITY RESEARCH: REALISTIC EXPECTATIONS

Expected Versus Actual Detection and Magnitude of Impact

Despite evidence of adverse effects supporting regulation, public health surveillance may lack the sensitivity to detect the impact of regulation on those adverse effects. Impact assessments often project seemingly easy-to-detect consequences of exposure reduction, which are sometimes based on simplifying assumptions. Real-world accountability assessments need to contend with nonspecific outcomes and possible changes in other factors that affect outcomes of interest. The public, legislators, and other stakeholders may expect accountability assessment to yield informative results, but these expectations may not be readily met.

Misplaced expectations may reflect misunderstanding of the epidemiologic evidence that has motivated air quality standards or insufficient understanding of the expectations of scientific research. For PM, epidemiologists estimate the relation between concentration and risk by fitting statistical models to public health data. The resulting risk coefficients, which describe the predicted change in risk as changes in PM concentration, can then be used in a risk assessment to project the burden of disease attributable to air pollution. Although such projections are based in real-world data, they come from a model of how pollution affects health that probably does not represent faithfully how pollution actually affects health or what impact will occur if PM concentrations are lowered. These projections are also subject to uncertainty, as are any studies intended to address accountability. Projections of burden must consider the range over which an accountability study has reasonable statistical power to detect change.

The magnitude of impact from regulations estimated from toxicologic, rather than epidemiologic, findings may be even more uncertain and no easier to detect. Toxicologic studies, regardless of the biological system used, usually employ higher exposure concentrations or tissue-specific doses than occur naturally, even compared with those before regulations. Indeed, the lower bounds of experimental exposures often exceed even those received by the

most heavily exposed occupational groups. Projection of exposure-response functions down to the level of environmental exposures is tentative, even though the nature or magnitude of effects demonstrated by such projection may warrant regulation. In the absence of data from humans, however, estimates of human environmental exposure-response functions and expected magnitude of impact in humans is also uncertain. Understanding the biological mechanisms can enhance the degree of certainty.

Uncertainty in Measuring Impact

Few (if any) health outcomes relevant to accountability are affected by single factors; thus, a pollutant, pollutant class, or even air pollution in aggregate, is not likely to be the sole determinant of health outcomes used in accountability assessments. The multicausal nature of many health outcomes is well recognized. Researchers take this fact into account through design or by using proper analytic strategies to control for possible confounding or to identify changes in effects. The relative effects of other factors that also determine risk may exceed those of the air pollutants of interest. Thus, detecting the effect of a decline in exposure may be difficult if potential confounding and modifying factors are also changing.

Changes at the left side of the chain of accountability might be counterbalanced by other, simultaneous changes. An emission reduction might be countered by an increase in vehicle miles traveled, for example, or emissions controls to power plants might be instituted at the same time as coal combustion for generating power is increased. In such instances, demonstrating a net benefit of the regulation would require counterfactual evidence (ie, that which estimates how much greater the health burden would have been had the regulation not been implemented). Taken broadly, even no reduction in disease burden might be considered a positive consequence in the face of increasing population and economic productivity.

Delayed or Nonlinear Changes in Health

The ability of studies to detect the impact of air quality regulations may also be blunted by the time course of change in air quality after implementation. Substantial time for compliance may be needed, as is likely for the new NAAQS for $PM_{2.5}$. Over that same interval, changes in other pollutants and other risk factors are likely. Up to 20 years may be needed for all regions of the United States to comply with the $PM_{2.5}$ NAAQS. Time is required to establish a monitoring system, determine compliance, develop and approve state implementation plans (SIPs), implement control strategies, and assess the impact.

Substantial time may also be needed for source-based pollutant-reducing technology to affect the operation of some pollution sources. For example, a more stringent emission standard for new vehicles may meet its primary goal of reducing emissions from those new vehicles, but the rate at which human exposures are reduced will depend on the rate at which the vehicle fleet turns over. Further complicating assessment of accountability is the uncertain temporal relation between a change in exposure and a change in disease risk, particularly for chronic, sustained exposures. For example, the lung cancer risk associated with outdoor air pollution probably reflects exposure across a lifetime; a reduction in emissions of airborne carcinogens would not affect lung cancer rates for decades. A more intermediate marker, more to the left on the chain of accountability (such as mutagenicity of airborne particles), would be a more temporally sensitive indicator. A critical example is the presumed shortening of life associated with long-term exposure to PM. Any gain from reduced PM exposures may be relatively small, come into play slowly, and possibly be undetectable among the myriad factors affecting longevity.

Nonlinear relations on the left side of the chain of accountability may complicate assessments of outcomes at the right side. The relation of changes in emissions to changes in exposures is not necessarily linear. For some pollutants, the atmospheric chemistry determining relations among emissions, meteorology, and pollutant concentrations is complex and nonlinear. One well-known example is the complex relations among emissions of volatile organic compounds (VOCs) and oxides of nitrogen (NO_x), meteorology, and ozone formation. Regulations successful in achieving reductions in emissions of either VOCs or NO_x do not necessarily lead to a proportional reduction in ozone (EPA 1996c). Exposure and health outcomes may also not be linearly related, so consequences of actions taken at the left side of the chain of accountability for outcomes on the right side may not be predictable with great certainty.

HEALTH OUTCOMES

The specific health outcomes chosen for studies of the health impact of air quality regulations and the inferences that can be drawn from these studies depend on the temporal relation between the change in exposure induced by a regulatory intervention and the outcome(s) presumed to result from it. Numerous epidemiologic studies have addressed health outcomes in relation to air pollution exposure using time periods that reflect the assumed course of underlying biological mechanisms and the availability of exposure data. These studies provide a range of health outcomes to consider.

SHORT-TERM TEMPORAL EFFECTS

One of the most obvious and relevant dimensions of exposure variability is short-term (1–5 day) temporal variability. Short-term episodes of high air pollution concentrations have had disastrous effects in the past, and episodic increases continue to trigger warnings in parts of the United States. The majority of epidemiologic studies of air pollution conducted in the past decade, including many studies of PM and ozone, have evaluated health effects associated with short-term temporal variation in ambient pollutant concentrations. Control measures that reduce average as well as peak exposures would be expected to reduce effects over short time scales. However, the relation between effects associated with short-term and long-term variability is uncertain. For example, substantially greater effects on mortality are estimated for long-term than for short-term differences in ambient concentrations of PM.

Differences in short-term temporal variation in exposure may be important to health effects of long-term exposure for several reasons. First, health effects of long-term exposure to PM and to oxidants such as ozone (EPA 1996c) probably are cumulative effects of repeated short-term exposure. Second, mean or median exposure over years or decades may not account fully for observed health effects. Differences in patterns of day-to-day variability of exposure may also be relevant. Living in a community with moderate mean PM or ozone pollution levels but with some short-term periods of very high levels may have different health consequences than living with moderate mean pollution levels with little day-to-day variability. How to quantify, model, or describe the relation between cumulative short-term effects and effects of long-term exposure remains unclear. But the short-term temporal dimension of exposure variability may be important even when studying effects of long-term exposure.

Mortality Counts

Mortality has long been a key health endpoint in epidemiologic studies. It is a distinct and discrete health outcome; mortality data are routinely collected and readily available for epidemiologic analyses. Vital-statistics records can be obtained from national databases and can be used to generate daily death counts in specific cities or communities. Mortality counts can be stratified by cause of death (respiratory, cardiovascular, or other), sex, age, and various other factors. In some studies of air pollution episodes, pollutant concentrations that varied temporally over a brief period were the source of exposure variability and mortality counts were the health outcome. These studies include the early reports of severe air pollution episodes in Meuse Valley, Belgium (Firket 1931), Donora, Pennsylvania (Ciocco and Thompson 1961), and London, England (Logan and Glasg 1953), as well as more recent studies of more moderate episodes (Wichmann et al 1989; Anderson et al 1995).

To some degree, accountability has been partially addressed using these relatively simple episode studies. Mortality counts become elevated during episodes of high pollution exposure and then return to normal afterward. Also, policies that have eliminated extreme pollution episodes in the United Kingdom and the United States have also resulted in the elimination of such extreme mortality episodes.

Daily time-series studies also consider short-term temporal variation as the source of exposure variability and mortality counts as the health outcome (Vedal 1997; Pope and Dockery 1999). These studies, which previously involved one or at most a few cities, now often involve many cities (Samet et al 2000a,b; Katsouyanni et al 2001). Overall, this design may be difficult to use for accountability assessment because statistical control for long-term time trends and seasonality is employed in models used to estimate effects. With sufficiently long time series, however, changes in magnitude of effect over time might be detectable (Burnett et al 2003).

Hospitalizations and Access to Medical Care

Similar to mortality counts, counts can be compiled for hospitalizations and other indicators of use of clinical medical care. These data are often available from computerized hospital records, Medicaid, Medicare, Canadian provincial or national health insurance databases, or other sources such as health maintenance organization (HMO) records. These counts can be generated for different diseases (such as asthma or other respiratory or cardiovascular diseases) and stratified by sex, age, and other factors. Various studies of short-term air pollution episodes have evaluated hospitalizations and related medical care endpoints (Logan and Glasg 1953; Wichmann et al 1989; Anderson et al 1995). These studies found that respiratory and cardiovascular events increased during pollution episodes and then returned to normal levels afterward. Also, dozens of daily time-series studies have similarly evaluated the association of PM with daily counts of hospitalizations and other health care endpoints (eg, Pope and Dockery 1999).

Medication Use, Symptoms, and Subclinical Physiologic Changes

Various other health endpoints such as daily medication use, daily symptoms, or short-term subclinical physiologic changes (eg, spirometric changes) could be evaluated in relation to short-term temporal changes in air pollution.

These health outcomes are not routinely collected and can only be feasibly collected on a daily basis via surveillance of panels of subjects. In fact, many panel studies have evaluated air pollution–related changes in lung function, respiratory symptoms, medication use, cardiac autonomic function (using heart rate variability [HRV]) and related health endpoints, exploiting short-term temporal variability in air pollution (eg, Pope and Dockery 1999). The use of such health outcomes in planned, prospective accountability assessments would be most feasible in well-defined situations where abrupt changes in pollution exposure are anticipated.

LONG-TERM TEMPORAL VARIABILITY

A second dimension of exposure variability for accountability research is long-term temporal variability. Examples include: (1) changes in ambient pollutant concentrations over years or decades due to changes in emissions or other factors (EPA 2002); (2) changes in indoor air pollution levels (eg, PM levels) due to alterations in space heating (eg, conversion from coal to natural gas), cooking appliances (eg, conversion from gas-powered to electric-powered), and housing characteristics and penetration of outdoor pollutants (eg, more sealed and insulated homes, greater use of air conditioning) (Abt et al 2000); and (3) long-term changes in activity patterns (such as outdoor recreation, work outdoors, and commuting times) (EPA 1996a). Long-term exposure variability provides the most direct opportunities for accountability assessment studies of long-term consequences of exposure, which may be irreversible and contribute to chronic disability and reduced life expectancy.

Mortality Counts or Rates

In order to control for season and various potential time-dependent confounders, time-series studies of daily mortality eliminate this dimension of exposure variability by using analytic approaches (such as filtering data or nonparametrically smoothing time). Yet, long-term changes in mortality associated with changes in pollution can be assessed. For example, Archer (1990) utilized long-term temporal changes in pollution in three Utah counties (partially due to a steel mill opening in one of these counties) to explore the long-term effects on malignant and nonmalignant respiratory deaths. More recently, investigators have attempted to evaluate the health impact of long-term improvements in air quality in the United States by conducting analyses of county-level mortality rates from 1960 to 1997 (Lipfert and Morris 2002). An assessment of mortality in relation to decades-long trends in air pollution levels in California's South Coast area is currently being conducted by investigators at University of California at Berkeley with support from the California Air Resources Board (California Air Resources Board 2001; HEI 2002a). Interpreting the results of these studies hinges on how they have accounted for effects of secular trends in other, possibly more powerful, determinants of mortality (such as smoking, diet, exercise, and factors related to social class).

More recently, restrictions were imposed on use of bituminous coal in Dublin, Ireland, and on sulfur content of fuel for power generation and transportation in Hong Kong, China. Both restrictions were instituted over short time intervals in 1990, providing opportunities for researchers to measure directly the impact on mortality of actual air quality interventions (Clancy et al 2002; Hedley et al 2002). In both locales, investigators were able to document changes in ambient air quality subsequent to the restrictions and declines in long-term average mortality rates from cardiovascular and respiratory diseases associated with those changes (see sidebar 4.1). Although both studies attempted to account for secular changes in other mortality risk factors that could have produced the observed declines, they acknowledged that they may not have been completely successful.

Nonetheless, these studies and the studies of infant mortality discussed in the next section provide important information on the health impact of air quality interventions and offer models showing that studies can measure the mortality impact of interventions imposed over relatively short time periods.

Infant Mortality and Other Adverse Perinatal Events

Infant mortality and other adverse perinatal events, such as low birthweight, have been associated with prolonged exposure to air pollution (Woodruff et al 1997; Bobak and Leon 1999; Ritz et al 2002). These outcomes may be of particular interest for evaluating the health impact of regulations, especially those regulations that take effect over a relatively short time, because the outcomes must occur shortly after a change in presumably causal factors. Two studies recently described changes in infant mortality associated with reductions of industrial emissions due to a recession and reductions in pollution mandated by the 1970 Clean Air Act Amendments (Chay and Greenstone 1999, 2001). Between 1970 and 1990, air pollution (measured as total suspended particles [TSP]) and infant mortality both fell markedly in the United States. The TSP reductions were not, however, uniform in time or degree: US counties with the highest TSP levels experienced larger and more rapid declines during two periods in the early 1970s and 1980s. From these county-level data, Chay and Greenstone estimated that 4 to 8 infant deaths per 100,000 live births were prevented for each microgram per cubic meter reduction in TSP.

SIDEBAR 4.1. TWO RECENT STUDIES OF HEALTH IMPACT OF AIR QUALITY REGULATIONS

Recently, mortality impact of air quality regulations instituted over relatively short times was estimated in two locations nearly a world apart: Dublin, Ireland, and Hong Kong, China.

Ban on Coal in Dublin

Clancy and colleagues (2002) estimated the impact on mortality of a ban on marketing, sale, and distribution of coal in Dublin. The ban, instituted on September 1, 1990, was intended to improve air quality in Dublin. The city's air quality had declined markedly after a decade in which coal had, for reasons of cost and availability, increasingly replaced oil as the fuel of choice for domestic use.

To estimate the ban's impact on mortality, Clancy and colleagues (2002) compared air pollution levels and rates of mortality between 6-year periods before and after the ban: September 1, 1984 to August 31, 1990, and September 1, 1990 to August 31, 1996. They documented long-term trends in ambient concentrations of black smoke (BS) and sulfur dioxide (SO_2) collected daily in Dublin from a six-monitor network maintained by the Dublin County Borough. Rates of mortality were calculated for deaths due to all natural causes (ie, nontrauma deaths) and subgroups thereof: deaths due to respiratory disease, cardiovascular disease, and other causes.

In the period after the ban compared with that before the ban, concentrations of BS and SO_2 fell 35.6% and 11.3%, respectively, with the largest reductions in autumn and, especially, winter (Figure 4.1). Age-standardized mortality from natural causes declined by 5.7%. These declines were more substantial for people less than 60 years of age (7.9%), for mortality due to respiratory causes (15.5%) and cardiovascular causes (10.3%) as opposed to other causes (1.7%), and in winter (Figure 4.2). These estimates, which were adjusted for weather, influenza epidemics, and secular changes in countrywide mortality, were

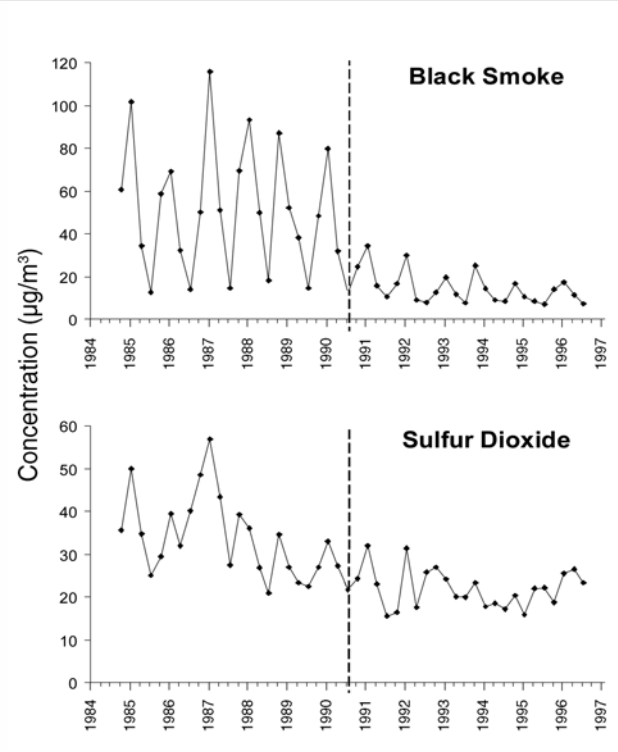


Figure 4.1. Seasonal average air pollutant concentrations for specific pollutants in Dublin. Dotted line indicates when coal sales were banned (from Clancy et al 2002).

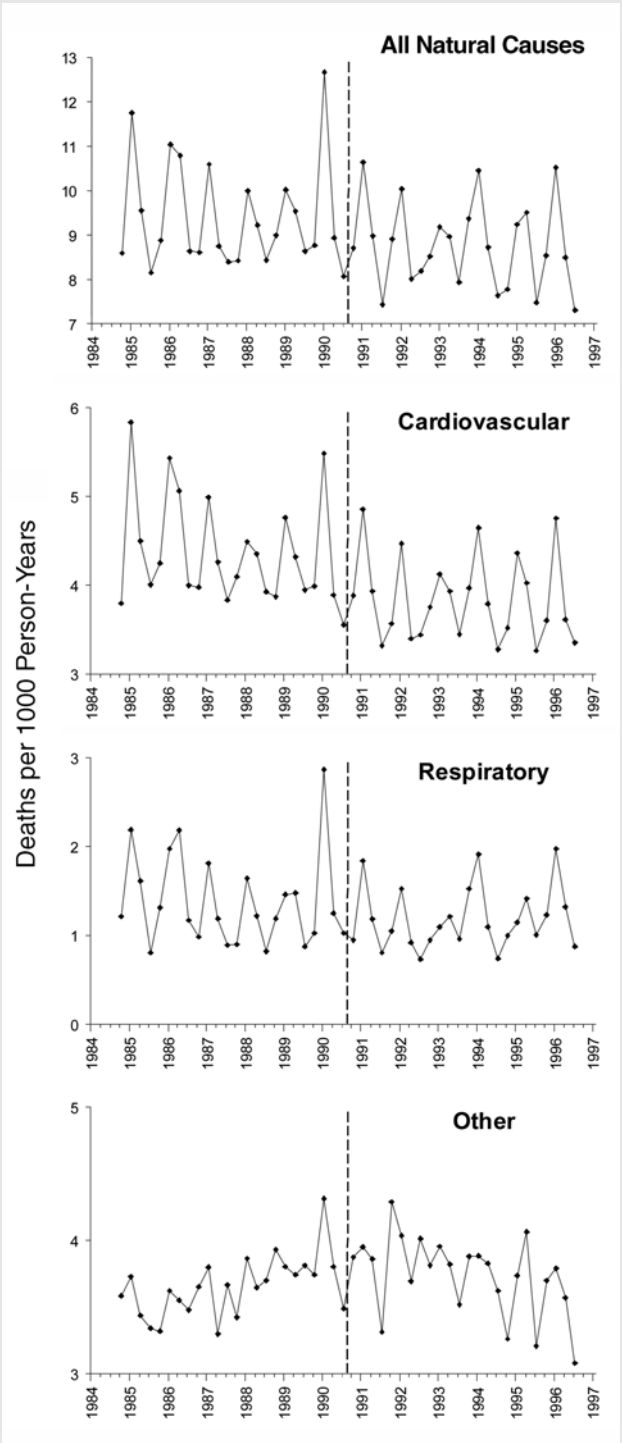


Figure 4.2. Seasonal average rates of mortality from various causes in Dublin. Dotted line indicates when coal sales were banned (from Clancy et al 2002).

generally lower than the corresponding unadjusted estimates (eg, for natural-cause mortality: −5.7% adjusted vs −8.0% unadjusted).

The investigators were aware that mortality risk factors besides air pollution levels might have changed before and after the ban, so they adjusted their estimates to take these factors into account. Weather is a strong risk factor for mortality that is related to air pollution. Clancy and colleagues (2002) used routinely collected data on temperature and relative humidity to account for effects of weather on mortality rates. Influenza epidemics, which occur regularly in winter, are an even greater risk factor. Lacking ongoing virologic surveillance data for influenza in Dublin, the investigators used countrywide information on deaths from pneumonia and influenza, combined with influenza surveillance data from the United Kingdom, to identify presumed influenza epidemic periods. Finally, to account for declines in mortality from cardiovascular and respiratory disease due to documented changes in known risk factors (such as hypertension and cigarette smoking) that occurred throughout Ireland during the study, the investigators adjusted their mortality risk estimates for Dublin using mortality rates from the rest of Ireland.

The investigators noted that this impact was qualitatively similar but quantitatively larger than what one would have predicted from existing European studies of air pollution and mortality. The results of the multicity Air Pollution and Health: A European Approach (APHEA) study (Katsouyanni et al 1997), when applied to the air pollution reductions in Dublin, predict only 2.1% and 0.7% reductions in natural-cause mortality for BS and SO_2, respectively, versus the 5.7% reduction actually observed.

Removal of Sulfur from Fuel in Hong Kong

In July 1990 authorities in Hong Kong regulated the sulfur content of fuel oil for power generation and road transport to be 0.5% or less by weight, a change that was implemented over a single weekend. Researchers at the University of Hong Kong have studied the impact of this regulatory action on ambient air quality and several health endpoints in the years since.

The regulation produced large reductions in ambient concentrations of routinely monitored pollutants that derive most directly from combustion of sulfur-containing fuels (Hedley et al 2002) (Figure 4.3). Over the following year ambient concentrations of SO_2 measured at multiple sites throughout the city fell by an average of 53% relative to baseline levels measured in the 2 years before. SO_2 levels remained 45% below baseline 5 years after regulation. Two years after regulation, levels of respirable sulfate (SO_4) particles fell by as much as 23% relative to the baseline period but rose again over the next 3 years and eventually exceeded baseline levels. Ozone levels rose in the 5 years after regulation, and no major changes in levels of either particulate matter less than 10 μm in aerodynamic diameter (PM_{10}) or nitrogen dioxide (NO_2) were observed.

Investigators at the University of Hong Kong measured the regulation's impact on indices of respiratory health in children living in two areas of Hong Kong that differed markedly in levels of ambient air pollution. Before the regulation, Peters and colleagues (1996) had documented a higher prevalence of respiratory symptoms such as cough, wheeze, and sore throat in primary-school children living in the more-polluted area compared with those in the less-polluted area. After the reduction in fuel sulfur content the ambient levels of both SO_2 and SO_4 fell markedly in the

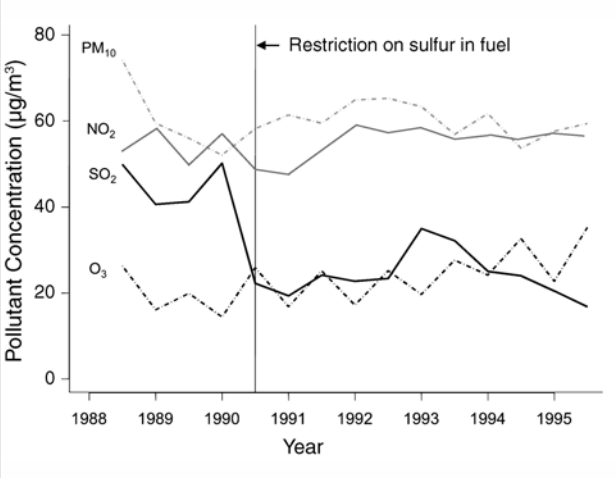


Figure 4.3. Average concentrations for specific air pollutants at five Hong Kong monitoring stations over the course of the study (data taken from Hedley et al 2002).

Hospitalizations and Access to Medical Care

In the mid to late 1980s several studies evaluated the health effects of a 13-month closure of the Utah Valley steel mill (eg, Pope 1996) (sidebar 4.2). During the winter of 1986/1987 a labor dispute and change in ownership resulted in a 14-month closure of the local steel mill, the largest single source of particulate air pollution in the valley. The 13-month reduction in pollution due to the closure of the steel mill resulted in marked reductions in pediatric respiratory illness in the community (Pope 1989, 1991). As in this study, hospitalization records collected over prolonged periods of time can be used in studies that assess accountability, although the recent Health Insurance Portability and Accountability Act (HIPAA) poses additional barriers to such studies (US Department of Health and Human Services 2002).

Medication Use, Symptoms, and Subclinical Physiologic Changes

Few studies have used endpoints such as daily medication use, daily symptoms, or short-term subclinical physiologic changes in assessing changes in long-term temporal variability. Such health outcomes would be most feasibly collected as part of a prospective panel surveillance effort or perhaps from databases of large health-care organizations.

SPATIAL VARIABILITY

Another important dimension of exposure variability is spatial variability. Studies that evaluate mortality effects of long-term exposure to PM have relied on long-term spatial variability in exposure: primarily changes in ambient concentrations across metropolitan areas. Changes in spatial

more-polluted area (80% and 38%, respectively), but major changes were not observed in the less-polluted area. The improvement in air quality was associated with larger declines in respiratory symptoms among primary-school children in the more-polluted area than in the less-polluted area (Peters et al 1996) (Figure 4.4). Wong and colleagues (1998) observed a similar pattern for changes in bronchial reactivity among primary-school children without asthma or wheezing in the year after the regulation. Children in both areas showed improved bronchial reactivity; those in the more-polluted area improved more markedly. Two years after regulation, improvements in bronchial hyperreactivity were evident only in the more-polluted area.

Most recently the University of Hong Kong researchers estimated the impact of the regulation on mortality of the Hong Kong population (Hedley et al 2002). The investigators studied changes in monthly average mortality rates for 1985 through 1995, 5 years before and after regulation. The peak in cool-season deaths apparent in the preceding 5 years was markedly reduced in the first year after fuel sulfur reduction but returned to preregulation levels over the next 4 years. The regulation was also associated with declines (2.1%, 2.0%, and 3.9%) in the upward trend of annual average natural cause, cardiovascular and respiratory disease mortality, respectively, that had been observed in the preceding 5 years (Figure 4.5). Annual rates of mortality due to other causes were not markedly affected. Reductions in mortality were more pronounced in the most highly polluted areas that had experienced the largest improvements in air quality. Hedley and coworkers estimated that the regulation resulted in gains in average life expectancy per year of exposure to the lower levels of air pollution: 20 days for women, 41 days for men.

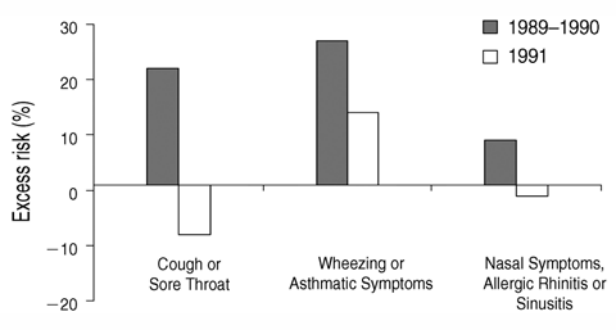


Figure 4.4. Excess risk for respiratory symptoms among primary-school children in Hong Kong before and after intervention to reduce sulfur content of fuel oil. Excess risk percentage calculated by comparison of children in a more-polluted area with children in a less-polluted area. For all symptoms, excess risk dropped to nonsignificant levels in 1991, after intervention. Data taken from Peters et al 1996.

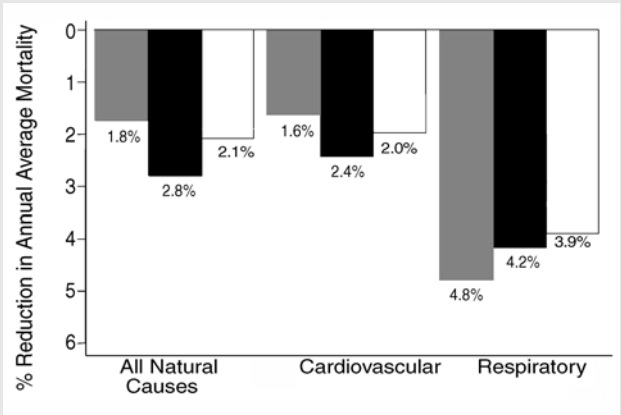


Figure 4.5. Percent reduction in annual average mortality due to all causes, cardiovascular disease, or respiratory disease after intervention to reduce sulfur content of fuel oil. Gray bars indicate people 15 to 64 years of age; black bars indicate people 65 years or more of age, white bars indicate people of all ages. Data taken from Hedley et al 2002.

differences in pollution levels over time provide some opportunities for accountability assessment studies.

Mortality Rates

Disease mortality rates offer an obvious health endpoint to use because of readily available data from vital statistics surveillance systems that exist in the United States and other countries. Since an early study by Lave and Seskin (1970), several population-based studies (ie, ecologic or aggregate level) have evaluated the association of cross-sectional spatial differences in annual PM concentrations with annual mortality rates across US metropolitan areas (Evans et al 1984; Lipfert 1984; Ozkaynak and Thurston 1987; Lipfert et al 1988). Overall conclusions from these studies were that mortality rates are associated with air pollution, most strongly with fine or sulfate PM. In terms of assessing accountability, these studies are limited. Nevertheless, accountability can be partially assessed by evaluating spatial differences in mortality that is associated with pollution after a policy leads to reduced pollution in higher-pollution areas and thus reduced spatial variability in pollution generally.

Survival and Mortality Risk

In addition to relatively simple population-based mortality rates, another mortality-related health endpoint that can and should be used in accountability assessment studies is loss of life expectancy. This outcome can be estimated from cohort studies, but such studies are expensive to conduct and are currently few in number.

A few prospective cohort mortality studies have been conducted that evaluated effects of long-term pollution exposure by using spatial differences in average exposures.

SIDEBAR 4.2. HEALTH IMPACT OF CLOSURE OF A UTAH VALLEY STEEL MILL

From 1985 to 1988, Utah Valley provided a unique opportunity to study health effects of particulate air pollution. A steel mill that employed only about 2% of the local work force contributed approximately 50% of the Valley's respirable PM. The mill closed in August 1986 due to a labor dispute and reopened in September 1987. During winters when the steel mill was open, PM levels were nearly double those observed when the mill was closed. An opportunistic study of this natural experiment (Pope 1989) revealed that elevated PM levels were associated with increased hospitalizations for pneumonia, asthma, bronchitis, and pleurisy. Figure 4.6 illustrates that hospital admissions of children due to respiratory disease were two to three times higher during years the mill was open relative to the year it was closed (Pope 1989). Critics of this early study suggested that apparent air pollution effects were due to epidemics of respiratory syncytial virus that were coincidental to operation of the steel mill (Lamm et al 1991). Subsequent epidemiologic research (Pope 1991, 1996) did not support this explanation and, furthermore, reported that day-to-day changes in PM were also associated with various other health endpoints (including lung function, respiratory symptoms, school absences, and deaths due to cardiovascular and respiratory disease).

More recently, several highly innovative toxicologic studies (Kennedy et al 1998; Frampton et al 1999; Ghio et al 1999; Soukup et al 2000; Dye et al 2001; Ghio and Devlin 2001; Wu et al 2001; Molinelli 2002) have reported analogous findings. Researchers analyzed extracts collected from filters used to monitor PM levels in the Utah Valley. These studies demonstrated that the extracts caused increases in oxidant generation, airway inflammation, cytokine release, and other indicators of lung injury; these effects were larger for particles collected when the steel mill was operating than when it was closed. For example, Figure 4.7 illustrates that extracts from filters collected while the mill was open caused an increase in cell counts in bronchoalveolar lavage (BAL) fluid of rats exposed for 24 hours relative to the effect of saline or extracts from when the mill was closed. Figure 4.8 shows that extracts from when the mill was open also increased proinflammatory cytokine concentrations relative to the effect of saline in BAL fluid of healthy human volunteers.

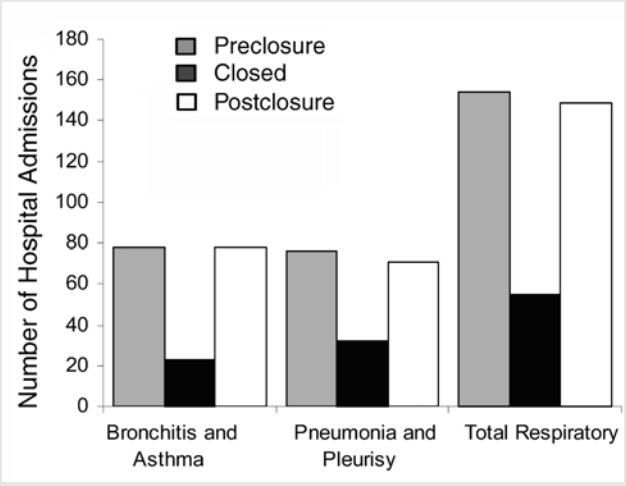


Figure 4.6. Winter hospital admissions in Utah Valley due to various respiratory diseases for children ages 0–17. Data taken from Pope 1989.

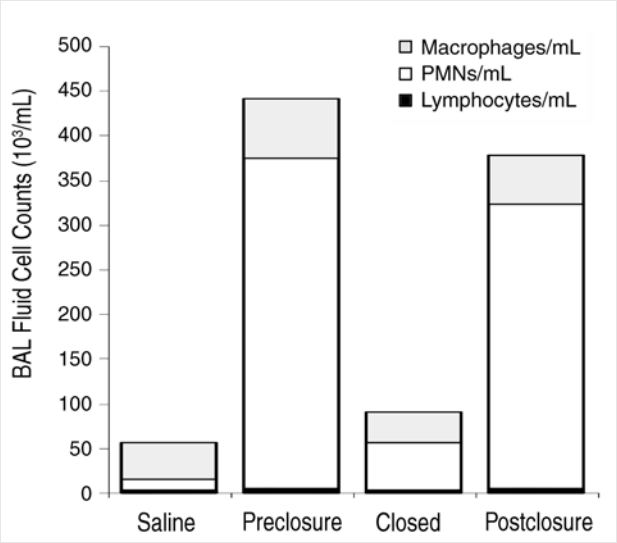


Figure 4.7. Cell counts in rat BAL fluid 24 hours after exposure to saline or saline plus PM extracts from years before, during, or after mill closure (2.5 mg extract/rat). Data taken from Dye et al 2001.

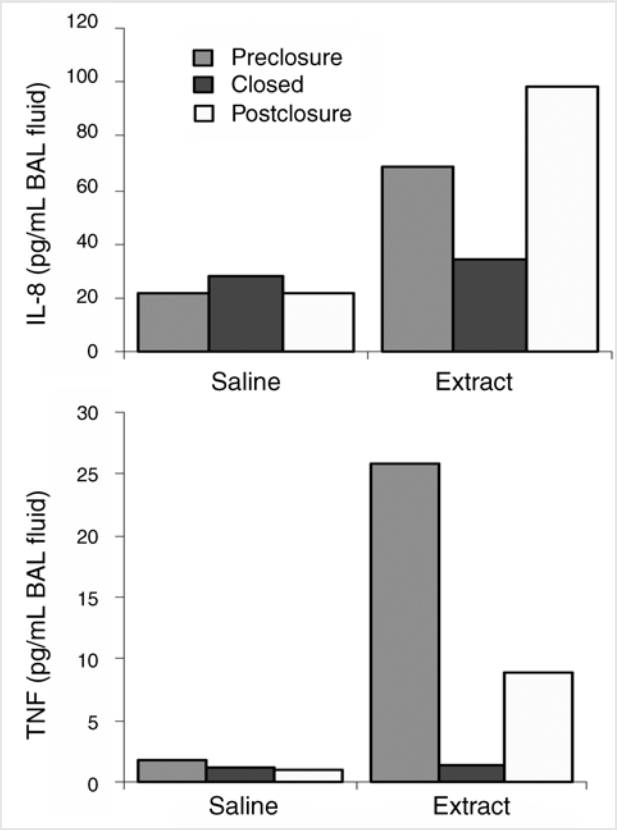


Figure 4.8. Proinflammatory cytokine concentrations in BAL fluid of human volunteers exposed to saline or saline plus PM extracts from years before, during or after mill closure. Data taken from Ghio and Devlin 2001.

The first of these studies, often called the *Harvard Six Cities study* (Dockery et al 1993), involved a follow up of more than 8000 adults living in six US cities. It controlled for individual differences in age, sex, cigarette smoking, education levels, body mass index, and other risk factors. Cardiopulmonary mortality was significantly associated with mean sulfate and fine PM concentrations during the study. Originally this study was designed to include a long-term temporal dimension by taking advantage of declines in pollution expected to occur in Steubenville and St Louis. Large declines in TSP but much smaller temporal changes in fine particles and sulfate particles were observed. These data made the initially reported analysis of mortality risks with respect to fine or sulfate PM primarily an analysis of spatial differences in exposure. More recent analysis (F Laden, personal communication, 2002) using a longer follow-up period included recent time periods when Steubenville had substantially lower levels of fine PM; results suggest that mortality risk did drop in association with the drop in pollution levels.

A second study, termed the *ACS* (American Cancer Society) *study*, was conducted to link individual risk factor data from the ACS Cancer Prevention Study II (CPS-II) with ambient air pollution data (Pope et al 1995, 2002). The results were similar to those observed in the Harvard Six Cities study: risk of cardiopulmonary mortality as well as lung cancer mortality were significantly associated with mean sulfate PM and fine PM concentrations.

These studies have provided surprising and powerful evidence for effects of long-term exposure to air pollution. Existing cohorts may offer even better opportunities for accountability assessments if the period of observation allows preintervention baselines to be established and duration of follow up is long enough to allow the observation of intervention-related long-term effects. However, in general, cohort-study designs are unique and thus do not offer a readily available model for accountability assessment.

Incidence and Prevalence of Disease

The incidence and prevalence of cardiac and respiratory disease could also be important health endpoints in accountability studies. Studies of the impact of reductions in sulfur in fuel in Hong Kong (see sidebar 4.1) on respiratory symptoms in children provide recent examples of how such endpoints can be used (especially as part of a larger suite of health endpoints, such as pulmonary function and mortality). In that situation, symptom-prevalence studies were possible because earlier work had established a preintervention baseline for comparison. Wider use of such endpoints as indicators of accountability would require a more systematic approach to tracking of such health endpoints. To date, however, tracking these diseases has proved difficult.

Other Biomarkers

Various other endpoints such as lung growth, lung and cardiac function decline, and markers of pulmonary or systemic inflammation might also be used in studies of spatial variability. Such studies would probably be costly but could be included in ongoing health surveillance efforts such as NHANES.

SUMMARY AND RECOMMENDATIONS

Air quality regulations are based largely on evidence interpreted as indicating a causal association between exposure and risk for the outcome. They are also based on the assumption that reduction of that risk is consequent to any control program (see Chapter 2 and above). Stakeholders therefore reasonably expect accountability studies to provide evidence of some improvement in outcomes that provided evidence to support promulgation of regulations. Therefore, for example, a reduction in daily numbers of deaths, or improvements in life expectancy, associated with air pollution will be critical endpoints for accountability studies.

The most serious health effects (such as mortality and increased morbidity from cardiovascular and respiratory diseases) are associated with not just one but several pollutants as well as other behavioral and environmental factors. Thus, although researchers planning studies of the health impact of air quality regulations have a variety of possible health endpoints from which to choose, none are associated uniquely with air pollution. Relevance and feasibility will therefore be the key determinants in choice of health endpoints. There is a critical need to account for causes of health endpoints besides air pollution that may be correlated with regulatory interventions (see Chapter 5).

A range of practical considerations will determine the feasibility of using specific endpoints for accountability research. National databases currently exist for some endpoints of interest (eg, mortality via the US National Centers for Health Statistics, hospitalization via the US Health Care Finance Administration). Some data, including baseline (preintervention) rates of some endpoints may be unavailable or of limited quality, however. For example, data on asthma prevalence in major cities across the United States is not yet collected in a consistent or reliable fashion. Although some biomarkers may be suited to addressing some questions of interest, developing and validating biomarkers from laboratory studies to use in public health surveillance will be challenging.

Toxicologic evidence, including that from human studies, plays a critical role in causal thinking about health effects of air pollution. In contrast, toxicology has played a less central role, relative to epidemiology, in promulgation of air quality regulations. For this reason, it will also likely play a lesser but important supporting role in accountability assessments.

Researchers and other stakeholders need to consider several caveats with regard to the ability to assess directly the health impact of air quality regulations. First, the magnitude of the expected benefit may be smaller, and thus the ease of its detection may be more difficult, than anticipated. Some stakeholders' expectations of observable benefits may reflect misunderstanding of the inevitable uncertainties in epidemiologic or toxicologic evidence advanced in support of the regulation. Describing such uncertainties as clearly and quantitatively as possible is, therefore, a major responsibility of health researchers and risk assessors (US National Research Council 2002). Second, the regulation may have a beneficial but immeasurable effect on the outcome. This caveat is, due, for example, to influence of causes besides air pollution on occurrence of the outcome. Also, although regulatory action may produce intended reductions in emissions per vehicle or from power plants, these may not lead to decreased exposure if source utilization increases. Finally, anticipated changes in health may not be sudden and may be nonlinear, due to time lags between promulgation of the regulation and its effects on emissions, ambient concentrations, and health effects.

Choosing health endpoints for use in assessments of the health impact of air quality regulations will depend critically on temporal relations among changes in pollutant emissions, concentrations, and exposure and development of a detectable endpoint. Endpoints that might be detectable shortly after exposures change are counts of daily deaths and hospitalizations, certain clinical endpoints such as medication use, and subclinical indices (such as changes in pulmonary function that can be linked to adverse clinical conditions). Biomarkers of health response have the potential to predict the health impact of regulations without waiting for disease outcomes. However, the challenges to using biomarkers are considerable: for instance, relations between biomarkers and health endpoints must be demonstrated, and biomarkers must be validated under field conditions.

Endpoints that might be appropriate targets for assessments of the long-term impact of air quality regulations include long-term average rates of adult and infant mortality, effects on population average lifespan, incidence of chronic cardiovascular and respiratory disease, and biomarkers such as age-related growth and decline of lung function. Studies of long-term impact will probably need to use information on spatial variation in exposure induced by regulatory interventions, as has been done in earlier epidemiologic studies. Existing cohorts may offer some opportunities for accountability assessments if the period of observation allows preintervention baselines to be established and the duration of follow up is long enough to allow observation of intervention-related long-term effects.

IDENTIFY ACCOUNTABILITY HEALTH ENDPOINTS WHEN REGULATIONS ARE PROMULGATED

Promulgation of regulations should be accompanied by analysis of impact, including specific outcomes to be measured. Outcomes thus identified would become main targets for accountability assessments.

INVENTORY AVAILABLE DATA RESOURCES

An inventory of data that could be used for assessing the health impact of air quality interventions would be of great value to scientists and research funding agencies by identifying both immediate opportunities and critical gaps in data. An inventory should comprise, at a minimum, the following elements:

- Systematic description of national, state, or local databases on health outcomes that might be used for studies of accountability. A catalog of data elements, strengths, and weaknesses would facilitate accountability assessments. Recent efforts to implement nationwide environmental health tracking in the United States and Canada have begun to identify and evaluate potential outcomes (Canadian Institute for Health Information 1999; US Centers for Disease Control and Prevention 2003b).
- A similar listing of databases of factors that could confound assessments of the health impact of air quality regulations. These listings should include data systems that track factors such as smoking behavior, medication use, and nutritional data at a population level over time.

ASSESS POTENTIAL BIOMARKERS OF RESPONSE

Biomarkers of response would be invaluable, particularly if they could be obtained as part of national or local health surveys. The current state of potentially useful biomarkers should be assessed critically with specific reference to their application to accountability research. This assessment would enable researchers and funding agencies to accurately gauge the feasibility of using available markers and the effort and cost required to improve promising markers for use in health-impact assessments.

REFERENCES

Abt E, Suh HH, Catalano P, Koutrakis P. 2000. The relative contribution of outdoor and indoor particle sources to indoor concentrations. Environ Sci Technol 34:3579–3587.

American Thoracic Society. 2000. What constitutes an adverse health effect of air pollution? Am J Respir Crit Care Med 161:665–673.

Anderson HR, Limb ES, Bland JM, Ponce de Leon AA, Strachan DP, Bower JS. 1995. Health effects of an air pollution episode in London, December 1991. Thorax 50:1188–1193.

Archer VE. 1990. Air pollution and fatal lung disease in three Utah Counties. Arch Environ Health 45:325–334.

Bobak M, Leon DA. 1999. The effect of air pollution on infant mortality appears specific for respiratory causes in the postneonatal period. Epidemiology 10:666–670.

Burnett RT, Cakmak S, Bartlett S, Stieb D, Jessiman B, Raizenne M, Blagden P, Brook JR, Samson PR, Dann T. 2003. Measuring progress in the management of ambient air quality: The case for population health. J Toxicol Environ Health. In press.

California Air Resources Board. 2001. Request for Proposals: Health Benefits of Incremental Improvements in Air Quality. California Air Resources Board, Sacramento CA.

Canadian Institute for Health Information. 1999. National Consensus Conference on Population Health Indicators. Final Report. Canadian Institute for Health Information, Ottawa ON, Canada. Available from *http://secure.cihi.ca/cihiweb/en/downloads/infostand_ihisd_e_phi.pdf*.

Centers for Disease Control and Prevention (US). 2003a. Second National Report on Human Exposure to Environmental Chemicals. National Center for Environmental Health, Atlanta GA. Available from *www.cdc.gov/exposurereport/tobacco/*.

Centers for Disease Control and Prevention (US). 2003b. Environmental public health indicators project (last updated 4/10/03). *www.cdc.gov/nceh/indicators/default.htm*. Accessed 8/18/03.

Chay KY, Greenstone M. 1999. The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. Working Paper 17 (unpublished). Center for Labor Economics, University of California, Berkeley CA.

Chay KY, Greenstone M. 2001. Air quality, infant mortality, and the Clean Air Act of 1970. Working Paper 42 (unpublished). Center for Labor Economics, Berkeley CA.

Ciocco A, Thompson DJ. 1961. A follow-up of Donora ten years after: Methodology and findings. Am J Public Health 51:155–164.

Clancy L, Goodman P, Sinclair H, Dockery DW. 2002. Effect of air pollution control on death rates in Dublin, Ireland: An intervention study. Lancet 360:1210–1214.

Cohen J, Brion G, Haines J. 1990. Review of the National Ambient Air Quality Standards for Lead: Assessment of Scientific and Technical Information. EPA 450/2-89/022. NTIS PB91-206185. US Environmental Protection Agency, Research Triangle Park NC. Available from *www.ntis.gov*.

Department of Health and Human Services (US). 2002. Standards for privacy of individually identifiable health information: Final rule. Fed Regist 67:53182–53273.

Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG Jr, Speizer FE. 1993. An association between air pollution and mortality in six US cities. N Engl J Med 329:1753–1759.

Dye JA, Lehmann JR, McGee JK, Winsett DW, Ledbetter AD, Everitt JI, Ghio AJ, Costa DL. 2001. Acute pulmonary toxicity of particulate matter filter extracts in rats: Coherence with epidemiological studies in Utah Valley residents. Environ Health Perspect 109(Suppl 3):395–403.

Environmental Protection Agency (US). 1996a. Analysis of the National Human Activity Pattern Survey (NHAPS) Respondents from a Standpoint of Exposure Assessment. EPA/600/R-96/074. National Exposure Research Laboratory, Research Triangle Park NC.

Environmental Protection Agency (US). 1996b. Review of the National Ambient Air Quality Standards for Particulate Matter: Policy Assessment of Scientific and Technical Information. OAQPS Staff Paper. EPA/452/R-96-013. Office of Air Quality Planning and Standards, Research Triangle Park NC.

Environmental Protection Agency (US). 1996c. Tropospheric ozone and its precursors. Chapter 3 in: Air Quality Criteria for Ozone and Related Photochemical Oxidants. EPA/600/P-93/004aF. Office of Research and Development, Washington DC.

Environmental Protection Agency (US). 1997a. National Ambient Air Quality Standards for Particulate Matter; Final Rule. 40 CFR, Part 50. Fed Regist 62:38651–38760.

Environmental Protection Agency (US). 1997b. Regulatory Impact Analyses (RIA) for the Particulate Matter and Ozone National Ambient Air Quality Standards and Proposed Regional Haze Rule. Office of Air Quality Planning and Standards, Research Triangle Park NC. Available from *www.epa.gov/ttn/oarpg/naaqsfin/ria.html.*

Environmental Protection Agency (US). 2002. Latest Findings on National Air Quality: 2001 Status and Trends. EPA 454/K-02-001. Office of Air Quality Planning and Standards, Research Triangle Park NC. Available from *www.epa.gov/air/aqtrnd01/.*

Evans JS, Tosteson T, Kinney PL. 1984. Cross-sectional mortality studies and air pollution risk assessment. Environ Int 10:55–83.

Firket J. 1931. The cause of the symptoms found in the Meuse Valley during the fog of December, 1930. Bull Acad R Med Belg 11:683–741.

Frampton MW, Ghio AJ, Samet JM, Carson JL, Carter JD, Devlin RB. 1999. Effect of aqueous extracts of PM_{10} filters from the Utah Valley on human airway epithelial cells. Am J Physiol 277:L960–L967.

Ghio AJ, Devlin RB. 2001. Inflammatory injury after bronchial instillation of air pollution particles. Am J Respir Crit Care Med 164:704–708.

Ghio AJ, Stonehuerner J, Dailey LA, Carter JD. 1999. Metals associated with both the water-soluble and insoluble fractions of an ambient air pollution particle catalyze an oxidative stress. Inhalation Toxicol 11:37–49.

Health Effects Institute. 2002a. Request for Applications: Winter 2002 Research Agenda. Health Effects Institute, Boston MA.

Health Effects Institute. 2002b. Understanding the Health Effects of Components of the Particulate Matter Mix: Progress and Next Steps. HEI Perspectives. Health Effects Institute, Boston MA.

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

Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. 2002. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: An intervention study. Lancet 360:1646–1652.

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, Bacharova L, Schouten JP, Ponka 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. BMJ 314:1658–1663.

Kennedy T, Ghio AJ, Reed W, Samet J, Zagorski J, Quay J, Carter J, Dailey L, Hoidal JR, Devlin RB. 1998. Copper-dependent inflammation and nuclear factor-κB activation by particulate air pollution. Am J Respir Cell Mol Biol 19:366–378.

Kim CS, Kang TC. 1997. Comparative measurement of lung deposition of inhaled fine particles in normal subjects and patients with chronic obstructive airway disease. Am J Respir Crit Care Med 155:899–905.

Lamm SH, Hall TA, Engel A, White LS, Rueter FH. 1991. Assessment of viral and environmental factors as determinants of pediatric lower respiratory tract disease admissions in Utah County, Utah (1985–1989) (unpublished report). Consultants in Epidemiology and Occupational Health, Washington DC.

Lave LB, Seskin EP. 1970. Air pollution and human health. Science 169:723–733.

Lipfert FW. 1984. Air pollution and mortality: Specification searches using SMSA-based data. J Environ Econ Manage 11:208–243.

Lipfert FW, Malone RG, Daum ML, Mendell NR, Yang CC. 1988. A statistical study of the macroepidemiology of air pollution and total mortality. Report BNL 52112. Brookhaven National Laboratory, Upton NY.

Lipfert FW, Morris SC. 2002. Temporal and spatial relations between age specific mortality and ambient air quality in the United States: Regression results for counties, 1960–97. Occup Environ Med 59:156–174.

Logan WPD, Glasg MD. 1953. Mortality in London fog incident, 1952. Lancet 1:336–338.

Mauderly JL, Bice DE, Cheng YS, Gillett NA, Griffith WC, Henderson RF, Pickrell JA, Wolff RK. 1990. Influence of pre-existing pulmonary emphysema on susceptibility to chronic inhalation exposure to diesel exhaust. Am Rev Respir Dis 141:1333–1341.

Molinelli AR. 2002. Effect of metal removal on the toxicity of airborne particulate matter from the Utah Valley. Inhalation Toxicol 14:1069–1086.

National Research Council (US). 1989. Biological Markers in Pulmonary Toxicology. National Academy Press, Washington DC.

National Research Council (US). 1999. Health Effects of Exposure to Radon: BEIR VI. National Academy Press, Washington DC.

National Research Council (US). 2002. Estimating the Public Health Benefits of Proposed Air Quality Regulations. National Academy Press, Washington DC.

Ozkaynak H, Thurston GD. 1987. Associations between 1980 US mortality rates and alternate measures of airborne particle concentrations. Risk Anal 7:449–461.

Paredi P, Kharitonov SA, Barnes PJ. 2002. Analysis of expired air for oxidation products. Am J Respir Crit Care Med 166:s31–s37.

Peters J, Hedley AJ, Wong CM, Lam TH, Ong SG, Liu J, Spiegelhalter DJ. 1996. Effects of an ambient air pollution intervention and environmental tobacco smoke on children's respiratory health in Hong Kong. Int J Epidemiol 25:821–828.

Pope CA III. 1989. Respiratory disease associated with community air pollution and a steel mill, Utah Valley. Am J Public Health 79:623–628.

Pope CA III. 1991. Respiratory hospital admissions associated with PM_{10} pollution in Utah, Salt Lake, and Cache Valleys. Arch Environ Health 46:90–97.

Pope CA III. 1996. Particulate pollution and health: A review of the Utah Valley experience. J Expos Anal Environ Epidemiol 6:23–34.

Pope CA III, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287:1132–1141.

Pope CA III, Dockery DW. 1999. Epidemiology of particle effects. In: Air Pollution and Health (Holgate ST, Samet JM, Koren HS, Maynard R, eds), pp 673–705. Academic Press, London, England.

Pope CA III, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, Heath CW Jr. 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am J Respir Crit Care Med 151:669–674.

Ritz B, Yu F, Fruin S, Chapa G, Shaw GM, Harris JA. 2002. Ambient air pollution and risk of birth defects in southern California. Am J Epidemiol 155:17–25.

Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000a. The National Morbidity, Mortality, and Air Pollution Study, Part 1: 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. 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.

Soukup JM, Ghio AJ, Becker S. 2000. Soluble components of Utah Valley particulate pollution alter alveolar macrophage function in vivo and in vitro. Inhalation Toxicol 12:401–414.

Vedal S. 1997. Ambient particles and health: Lines that divide. J Air Waste Manage Assoc 47:551–581.

Wichmann HE, Mueller W, Allhof P, Beckmann M, Bocter N, Csicsaky MJ, Jung M, Molik B, Schoeneberg G. 1989. Health effects during a smog episode in West Germany in 1985. Environ Health Perspect 79:89–99.

Wong CM, Lam TH, Peters J, Hedley AJ, Ong SG, Tam AYC, Liu J, Spiegelhalter DJ. 1998. Comparison between two districts of the effects of an air pollution intervention on bronchial responsiveness in primary school children in Hong Kong. J Epidemiol Community Health 52:571–578.

Woodruff TJ, Grillo J, Schoendorf KC. 1997. The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States. Environ Health Perspect 105:608–612.

Wu W, Samet JM, Ghio AJ, Devlin RB. 2001. Activation of the EGF receptor signaling pathway in airway epithelial cells exposed to Utah Valley PM. Am J Physiol Lung Cell Mol Physiol 281:L483–L489.

ABBREVIATIONS AND OTHER TERMS

ACS	American Cancer Society
BAL	bronchoalveolar lavage
BS	black smoke
EPA	Environmental Protection Agency (US)
IL-8	interleukin 8
NAAQS	National Ambient Air Quality Standard(s) (US)
NHANES	National Health and Nutrition Examination Survey (US)
NO_x	oxides of nitrogen
PM	particulate matter
$PM_{2.5}$	PM less than 2.5 μm in aerodynamic diameter
PMN	polymorphonuclear leukocyte
RIA	regulatory impact analysis
SO_2	sulfur dioxide
SO_4	sulfate
TNF	tumor necrosis factor
TSP	total suspended particles
VOC	volatile organic compound

Chapter 5. MODEL DESIGN AND DATA ANALYSIS

MEASURING HEALTH BENEFITS OF AIR QUALITY REGULATIONS: EXPECTATIONS AND EVIDENCE

The scientific and public health communities face a remarkable challenge in evaluating accountability for air quality regulations. Stakeholders' expectations are high and, as a result of past regulatory successes, current regulations pertain to ambient concentrations at which health effects are expected to be small. Some evidence of reductions in pollutant emissions, concentrations, and exposures could occur soon after implementation of regulations. But health benefits (especially effects on mortality and morbidity from chronic cardiovascular and respiratory diseases) may only accrue and become measurable over time—as full implementation is achieved, exposures are reduced, and the biological effect of those reductions occurs. A strategy for both short-term and long-term approaches to gauging accountability will, therefore, be needed. This chapter begins by considering research concepts, designs, and methods that contribute to estimating health benefits.

SCIENTIFIC APPROACH

Evidence for accountability is of interest when a policy change (an intervention) is made with the expectation that health benefits will result. Data for assessing accountability might be obtained by collecting information before and after intervention or, absent a baseline, after intervention only. These observational approaches are sometimes referred to as *quasi-experimental*. Such interventions are not eligible for randomized clinical trial design, which involves random assignment of the treatment of interest (eg, to control and experimental groups). This design, generally used for individuals and therapeutic interventions, has been applied at the community level to assess the impact of health interventions such as tobacco control programs.

Because randomization leads to statistical comparability of treatment and control groups, a key strength of the randomized clinical trial is the strength of its evidence in causal inference. In observational studies, the treated groups (for example, those who experience higher air pollution exposures) often differ from the control groups (for example, those who experience lower air pollution exposures). These a priori differences between the groups may introduce confounding and result in incorrect estimates of the effect of intervention. In designs that track changes after an intervention, confounding occurs if other factors associated with the intervention, and independently with the outcome, change along with air pollution exposure. In observational studies of real-world exposures, findings may also be biased by people choosing to be exposed or not on the basis of factors related to disease risk. For example, families with children with severe asthma may live in or move to less polluted locations. A further complication in observational studies that may also affect randomized clinical trials is measurement error (ie, imperfect assessment of exposures, outcomes, and other relevant factors).

Criteria have been developed for evaluating a body of research findings to reach conclusions about the evidence it provides for causal relations. Historically, somewhat distinct but overlapping approaches have been used in observational and experimental research. The Koch postulates, developed in the 19th century, were used to determine whether the link between an infecting organism and a disease was causal. The postulates include replication of human disease in an animal experiment using the suspected microorganism. This approach is not, however, directly applicable to observational studies or most experimental research in humans.

The criteria now widely used in public health are commonly termed the *Hill criteria* (articulated as viewpoints by Sir Austin Bradford Hill, a British medical statistician [Hill 1965]) or the *Surgeon General's criteria* (due to their landmark application in the 1964 Surgeon General's Report on Smoking and Health) (US Public Health Service 1964). The criteria include strength, consistency, specificity, and temporality of the association (the effect must

This chapter is one part of Health Effects Institute Communication 11, *Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research,* which also includes five other chapters and an Executive Summary of the project.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R82811201 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.

follow the cause in time); the presence of dose-response gradient, plausibility, and coherence; and secondarily, availability of experimental evidence and applicable analogies. These criteria have been applied to evidence concerning air pollution and health to guide judgments about causality of associations (Ostro 1993; Gamble 1998). Rothman has argued that, apart from temporality (ie, cause precedes effect), no Hill criteria are necessary; that is, an association might be causal and yet not meet some of the Hill criteria (Rothman and Greenland 1998a). Nonetheless, as the strength of evidence mounts for each criterion, the overall level of confidence for a causal relation increases.

In biomedical research, a hierarchy of evidence types has been proposed that grades the relative weight of studies and their findings with regard to causal inference (Guyatt et al 1995). In this hierarchy, findings of randomized clinical trials are given greater credibility than results of observational studies. Further, studies are more credible if their findings are in agreement with other studies (corresponding to Hill's notion of consistency).

The randomized clinical trial is at the top of this evidential hierarchy; it provides the most convincing evidence for causality (and therefore accountability) for comparisons for which validity is protected by the randomization of treatment. Such experiments are considered the gold standard in biomedical research if they have sufficiently large samples, highly structured protocols, and limited variation in participants and treatments (eg, exposures, interventions). Internal validity, clear and a priori definitions, and standardized measurement are the strengths of this approach. Even when the mechanism of action is unknown and potential confounders are not measured, causality can be inferred because randomization increases the probability that treated and comparison groups have comparable distributions of potential confounders. That is, when study participants or other units are randomly assigned to various exposures and exposure mixtures, factors that might bias interpretation of results (eg, confounders) are stochastically balanced over exposure groups even when potential confounders are not measured. However, selection bias and lack of follow up can degrade the protection afforded by randomization and increase vulnerability to biases that affect purely observational studies (particularly if the bias is informative, that is, dependent on exposure and susceptibility to the health outcome).

Randomized interventions for ambient air pollution exposures at the population level are not generally possible: studies of people moving to areas of higher or lower pollution, cohort studies, and time-series studies have only some required features. However, randomized studies in clinical and laboratory settings can be informative. All studies strike a trade-off between direct relevance to public health issues and internal precision and credibility. For example, animal studies provide tightly controlled experimental evidence for the species under study (within the dose range administered and for the endpoints considered) but the validity of extrapolations to other doses and to humans depends on assumptions that cannot usually be verified.

Even though randomized studies are in many ways the gold standard for hypothesis testing, tight inclusion standards (the environmental equivalent of such clinical criteria as "50–59-year-old women with newly discovered stage I breast cancer") limit generalizability. And in a randomized study, only a small fraction of relevant information is protected by randomization; much of it is observational. Of course, many of these features apply to any study used to assess accountability. The randomized trial remains a useful benchmark for considering the possible weaknesses of observational study.

CAUSAL MODEL

The chain of accountability (Figure 5.1) is a simplified causal model of the way air quality regulation affects health. It may be useful for designing and evaluating studies to address accountability.

The chain depicts possible interrelations between actions and their effects on pollutant emissions, concentrations, and exposures and health effects. The arrows are bidirectional in some instances. For example, air pollution regulations aim to reduce air pollution emissions, but sources and patterns of air pollution emissions also affect the thrust of regulations. Emissions clearly affect concentrations, but measured air pollution concentrations also influence emissions, most often by catalyzing more stringent regulation. Air pollution concentrations are a key determinant of personal exposure, which is associated with adverse health effects. Evidence of adverse health effects can in turn influence exposure to air pollution through personal avoidance strategies and changes in regulations. At least a partial understanding of these links and feedback loops is necessary to structure an accountability assessment and to understand how accountability can be addressed through research.

In biomedical research, the concepts of efficacy and effectiveness are used in considering consequences of interventions. *Efficacy* refers to the effect of an intervention when it is tested in a trial or another setting that is likely to yield an optimum effect. *Effectiveness* refers to the effect of an intervention in a real-world setting. In an accountability framework, assessments of anticipated benefits resulting

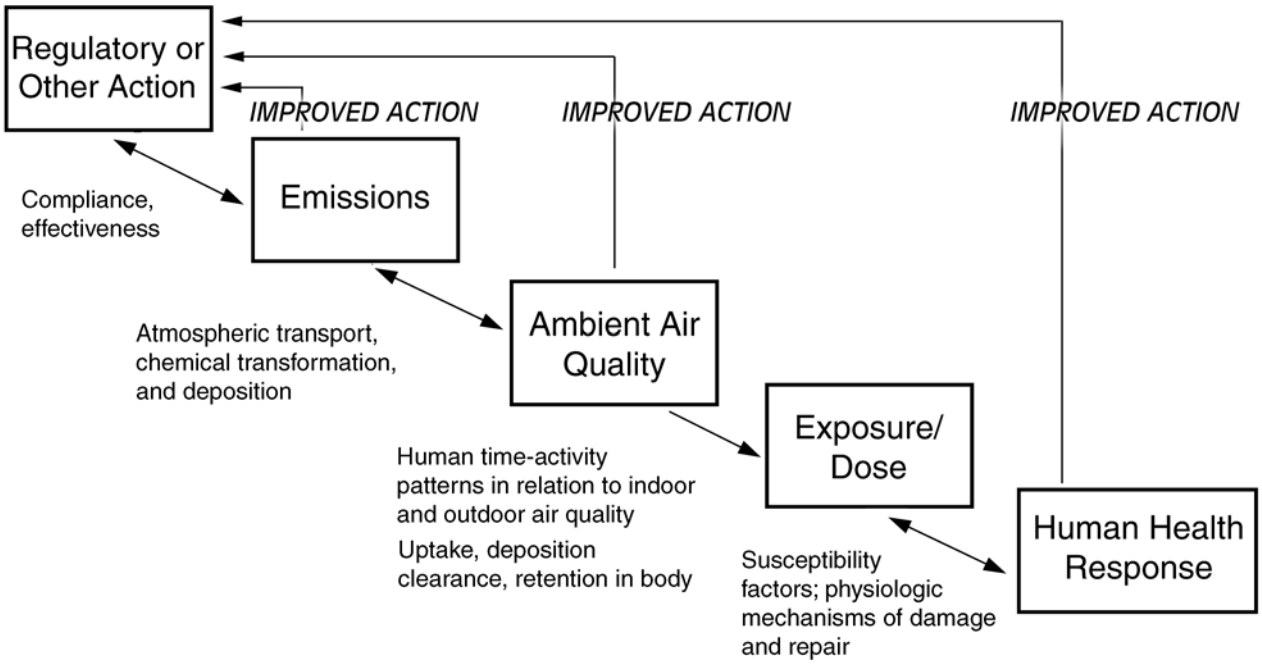


Figure 5.1. Chain of accountability. Each box represents a link between regulatory action and human health response to air pollution. Arrows connecting the links indicate possible directions of influence. Text below the arrows identifies general indices of accountability at that stage. At several stages, knowledge gained from accountability assessment can provide valuable feedback for improving regulatory or other action.

from changing regulations may be viewed as assessments of efficacy (eg, US Environmental Protection Agency [EPA*]'s Prospective 812 Report [EPA 1999]), whereas estimation of actual benefits resulting from changing ambient concentrations or exposures may be viewed as assessments of effectiveness (eg, the study of Dublin's coal ban [Clancy et al 2002]).

Although efficacy typically exceeds effectiveness in biomedical research and clinical practice (Reigelman and Hirsch 1996), Clancy and colleagues (2002) noted that the decline in observed mortality after air pollution reduction in Dublin was *greater* than what would have been predicted from results of epidemiologic time-series studies (see Chapter 4). Regardless, the Dublin study underscores the fact that a change in regulations might be expected to have different impact on health than a change in emissions, concentrations or exposures because regulatory changes may have downstream effects other than those intended. For example, they may affect exposure through mechanisms unrelated to change in concentrations. Therefore, measurement of effects at each step of the causal pathway may be necessary to understand why an intervention was or was not successful in improving a health outcome. In circumstances in which concentrations were actually reduced as a result of regulatory intervention, any real effect on health could be larger than a global, unconditional assessment of changes in regulation that produced health benefits.

Ambient air pollution levels do influence personal exposure, but they do so in a complicated, dynamic, interactive manner that varies from pollutant to pollutant. People may modify their behavior in response to high ambient pollutant levels so that the delivered dose may be lower for those people than for individuals not taking such avoidance actions. Further, people susceptible to air pollution (eg, those with asthma) may be more likely to take avoidance action. The possible role of behavior as a determinant of exposure highlights the need for an appropriate list of ultimate endpoints (some of which are lifestyle and quality-of-life endpoints rather than health endpoints) and a modeling framework that incorporates feedback loops. To date, however, neither the models nor the input data they require are available for this broad approach.

Studies that consider all or only part of the chain of accountability may be warranted and informative. Immediately after an intervention, elements on the left side of the chain may be appropriate targets for investigation even before anticipated health benefits are expected to occur. Over time, studies of the entire chain may be informative. More generally, building a relational model, bringing information to bear on the links in the model,

* A list of abbreviations and other terms appears at the end of this chapter.

and conducting sensitivity analyses will provide information on the expected range of causal effects.

HISTORICAL PRECEDENTS

Even with strong and unbiased evidence, accountability assessment is challenging. Decades were needed for the causal relations between cigarette smoking and lung cancer and other diseases to be fully accepted by the most skeptical stakeholder in that evaluation process, the tobacco industry. Early resistance was based on the absence of randomized, experimental evidence of long-term effects in humans and their long latency, coupled with vested economic interests (White 1990).

Accepting this type of causal relation—chronic exposure with health effects that occur long after the initial exposure—is a prerequisite for credible evaluation of the long-term health impact of environmental regulations. Climate change resulting from greenhouse gases is another, possibly more relevant problem with similar features. For climate change, of course, the time frame is even longer than for air pollution and the experiment cannot be repeated, even in principle. The lessons from smoking and its health effects predict difficulties for accountability assessment of air pollution: complete certainty is never achieved and investigations of biologic mechanisms invariably lead to deeper levels of inquiry.

In the environmental setting, elimination of lead in gasoline provides a relevant precedent for accountability assessment. Soon after the 1970 US Clean Air Act was implemented, steps were quickly taken to shift the country's vehicle fleet to lead-free gasoline. Lead levels in ambient air rapidly declined, and population blood lead levels subsequently dropped (Cohen et al 1990; also see Chapters 3 and 4). These consequences of the change in regulation, strengthened by substantial clinical and laboratory studies that validated the link between blood lead levels and neuropsychologic development, established the penultimate links in the chain of accountability and convinced all but the most extreme skeptics that the regulation conferred direct health benefits.

Carbon monoxide (CO) regulation also has documented beneficial consequences. CO emissions and ambient concentrations have declined substantially in the United States since the 1970s with addition of CO-emission controls to motor vehicles. The resulting public health gains, in terms of reduced ambient concentrations and doses, were estimated using carboxyhemoglobin as a biomarker (EPA 1991, 1992). As with lead, a reasonably certain understanding of the health effects of CO and its biological basis helps in demonstrating that health benefits have occurred.

Lead and CO are useful examples because specific features of exposure to those agents have limited the debate about benefits of their regulation. Both are readily measured in ambient air, have specific biomarkers that serve as dose indicators, and have well-documented adverse health effects and mechanisms of action. In contrast, regulation of particulate matter (PM) and other components of ambient air is fraught with difficulties. Air pollution is a complex mixture, its biomarkers are not completely specific to individual components in air, health effects can be both short-term and deferred, and whether specific components in air induce short-term and long-term health effects is unclear.

Several quasi-experimental, observational studies have evaluated health effects of air pollution. These include studies in areas where new regulations had a major impact on emissions (Dublin [Clancy et al 2002]; Hong Kong [Hedley et al 2002]), studies in which point-source industrial emissions decreased dramatically (Pope 1989), and studies of people moving from areas of high or low exposure to areas of opposite exposure. For example, Avol and colleagues (2001) assessed lung function growth in children who moved from areas with high PM concentrations in southern California to areas with low concentrations, and vice versa. Although observed relations between air pollution levels and lung function could have been affected by confounding or selection bias (relocation being related to health), such biases were found to be unimportant. All of these studies used public health indicators measured before and after an intervention; because of their similarities to randomized interventions, they are relatively resistant to the types of biases that plague other observational studies.

Policy-makers and the scientific community need to work together to identify prospective opportunities for accountability assessment. Such quasi-experiments will be most informative if data collection on exposures and outcomes is standardized, either in an ongoing manner or by rapid implementation after an appropriate regulation.

APPROACHES TO ASSESSING ACCOUNTABILITY

JUST BETTER SCIENCE?

Some might argue that assessing accountability depends only on better science—that scientific methods, in particular those of epidemiology and toxicology, could be enhanced to generate better evidence and provide a platform for decision making. Although improved science is,

without question, central to both applications, new scientific methods alone would be insufficient to integrate evidence across the chain of accountability. As Figure 5.1 indicates, accountability assessment may incorporate evidence over a long, complex chain of relations. Therefore, evaluation of data requires integration of directly and indirectly related information, to which statistical synthesis is central. Statistical synthesis should describe relations among the links of the chain, produce the statistical relations needed for evaluation, and identify gaps in data or research. Developing a comprehensive synthesis is challenging, requiring a team approach among epidemiologists, statisticians, atmospheric scientists, behavioral scientists, and regulators.

INTERVENTIONS THAT REDUCE AIR POLLUTION

Much of the evidence of health effects of air pollution comes from observational studies that relate changes in health indicators to changes in exposure to air pollution in space and time. Models based on results of such studies are used to estimate the expected benefits of current and future air pollution regulations. Direct study of the effect of interventions that reduce air pollution may be a more definitive approach to determining whether air pollution regulations actually result in health benefits. Compared with the usual observational studies, studies of interventions can disrupt links between confounding factors and exposures that may be unavoidable and cause bias in many observational studies of environmental factors and health. Studies of interventions, as quasi-experiments, have the possibility of involving clinical trials in which interventions can be controlled by design. This possibility may strengthen inferences about causality, making findings from intervention studies more definitive than observational studies in which effects of confounding factors are generally controlled only in analysis.

An intervention can be broadly viewed as any intentional (planned) or unintentional (unplanned) change in air pollutant concentrations or exposures that results from some out-of-the-ordinary action (most often a human action rather than some natural process). Intentional interventions, the most obvious type, are typically due to regulations or policies; unintentional interventions usually occur indirectly as a result of human action not intended to affect air pollution concentrations or exposures. Natural experiments are one type of unintentional intervention: for example, a labor strike that closes down an industrial source of air pollutants. Intentional and unintentional interventions can be exploited to gain insight into consequences of air pollution regulations for public health.

REGULATORY AND POLICY INTERVENTIONS

Intentional regulatory interventions can aim to reduce air pollutant emissions, which should result in reducing concentrations and, ultimately, exposures. Alternatively, an intervention might aim to reduce population exposures to air pollution without attempting to reduce emissions and ambient concentrations: for example, by encouraging changes in behavior, such as advisories that instruct susceptible individuals to stay indoors on high-pollution days.

Interventions Aimed at Reducing Ambient Concentrations

Several study designs could be used to assess the impact on public health of intentional interventions aimed at reducing ambient pollutant concentrations. A cohort design involves monitoring a sample (cohort) of individuals for some time for an outcome of interest. To assess accountability, the most obvious study of this type would be one in which only part of the cohort is affected by the intervention (the exposed group) whereas the unaffected part comprises the control group (or the unexposed group). Comparability of the exposed and unexposed groups with respect to risk factors that determine health outcomes is crucial to the success of a cohort study. When air pollution regulations are applied to large geographic regions (US states, for example), it may be difficult to identify settings in which one could successfully compare outcomes in jurisdictions with and without a given regulatory intervention and have reasonable confidence that the groups are otherwise comparable.

The period of follow up required for a cohort study depends critically on how quickly air pollution regulations are implemented, the extent to which they actually affect ambient concentrations, and the amount of time required for a change in exposure to cause a change in risk of the health outcome (ie, the induction time). Estimation of the health impact must take into account and be consistent with the time course of changes in pollutant emissions, concentrations, and exposures and the induction time for health effects related to air pollution. Will changes in emissions and their effects on ambient concentrations be immediate or gradual? For airborne particles, for example, reductions following the 1997 US National Ambient Air Quality Standard (NAAQS) for $PM_{2.5}$ (PM less than 2.5 μm in aerodynamic diameter) are likely to be gradual and modest due to the relatively low current $PM_{2.5}$ concentrations. For health effects (such as lung cancer) that may require relatively long periods of time to elapse until cases produced by exposure to air pollution can be detected, the period of follow up might be prohibitively long (perhaps

more than 20 years). For some cardiovascular outcomes, however, a two- or three-year follow up may be sufficient.

Given the difficulty of identifying an appropriate comparison group for a cohort study, the same people can be followed before and after an intervention: a pre-post design. Most intervention studies are likely to be this type, comparing outcome occurrence rates before and after an intervention. A principal concern during pre-post cohort studies is that factors unrelated to the intervention might be simultaneously changing in the cohort, including, of course, age (which is strongly tied to risk for most health outcomes). Separating the effects of age and other time-varying risk factors from those of the intervention might prove difficult. This pre-post design is probably most appropriate when risk for outcomes does not vary markedly with age or when the effect of pollution exposure on disease risk is relatively rapid. In general, information should be collected on time-varying risk factors in addition to air pollution if possible.

Caution is needed in interpreting results of a pre-post analysis in the absence of an appropriate control group for which outcomes can be compared before and after. For example, we may observe a stronger association between a regulation and air pollution levels if unusually high levels prompted regulation but were themselves partially the result of natural statistical variation, a phenomenon known to statisticians as *regression to the mean*.

A cohort study of a population that is enrolled and followed over time (eg, the Six Cities Study [Dockery et al 1993]) may not always be feasible. Serial cross-sectional studies may be a realistic alternative, depending on the outcome of interest. Cross-sectional studies would be most appropriate for outcomes that might be expected to have short induction times, such as development of respiratory symptoms or lung function (measured using standard methods). A similar argument could be made for tracking levels of some types of biomarkers. For cohort studies, care is needed to ensure that populations at each cross-sectional survey period are comparable. In addition, secular changes in outcomes not associated with changes in air pollutant concentrations must be identified and controlled; a relatively short time scale helps to prevent serious complicating effects of such changes. When the induction times for changes in air pollutant concentrations and particular health outcomes are long and the expected magnitude of changes is small, it is harder for such cohort studies to accurately detect effects. Such difficulties complicate interpretation of serial cross-sectional studies such as that of Lipfert and Morris (2002), who evaluated mortality from chronic diseases in the United States in relation to declining air pollution from 1960 to 1997.

Time-series studies have played a prominent role in assessing effects of short-term changes in ambient pollutant concentrations. Whereas the usual time-series study would not be well-suited to a pre-post design, serial time-series studies (ie, studies of a series of data sets) might be. Estimates of effect from a time-series study before an intervention could be compared to those after the intervention. A change in the estimate of the effect of air pollution on a health outcome can be interpreted as representing an effect of the intervention. However, this interpretation might be too simplistic, depending on the dose-response relation. If a dose-response relation is linear, a reduction in exposure (or dose) would be followed by a reduction in disease burden proportional to the change in exposure, without a change in the effect estimate. If a dose-response relation is nonlinear, interpreting changes in effect estimates over time is potentially more difficult. For PM and mortality, for example, estimates of effect from time-series studies tend to be larger if pollutant concentrations are lower (Samet et al 2000b; Vedal et al 2003). This possible nonlinearity of effect (ie, dependency on baseline level) must be considered when estimating the health impact of regulations.

Interventions Aimed at Reducing Current Levels of Exposure

Interventions targeted at reducing exposure rather than reducing ambient concentrations can also be studied. Examples of such interventions include public health advisories prompted by monitored or predicted pollutant concentrations above defined levels that target susceptible groups of the general population. Such interventions might more readily allow identification of appropriate control group(s), including the possibility of assigning the intervention at random to different geographic areas.

Randomized Interventions

Randomized allocation of reductions in air pollution concentrations has not, to our knowledge, been carried out. Such studies could be conducted by staging the implementation of a regulation in a randomized and informative fashion. This research approach will probably not be taken in the United States, given the national nature of air pollution regulation for criteria pollutants as well as feasibility and ethical constraints. Interventions might be undertaken to attempt to reduce exposures of susceptible individuals by modifying time-activity patterns or home ventilation or filtration. Whereas such studies have been implemented for indoor air pollutants, we are not aware of any such studies that have been carried out for outdoor air pollution. A randomized design has been used, however, to investigate whether an intervention can mitigate air pollution effects.

For example, vitamin C supplementation has been studied as a way of reducing effects of exposure to oxidant air pollution (Grievink et al 1999; Romieu et al 2002).

NONREGULATORY INTERVENTIONS

Events that inadvertently produce marked changes in pollutant emissions, concentrations, or exposures may closely mimic changes that might be produced by intentional interventions. Natural experiments (quasi-experimental studies) and studies of people who change their place of residence (migration studies) can be viewed as unintentional interventions.

Natural Experiments

We use the term *natural experiment* to refer to circumstances under which events not in the control of investigators lead to almost randomized changes in exposures of populations. The term *experiment* is, in a sense, inappropriate in that the researchers are not responsible for implementing the change in exposure; rather, they are opportunistically carrying out research to measure its consequences. The intervention in a natural experiment may occur in a manner that to a large degree breaks the link between change in exposure and potential confounding factors. Importantly, to exploit a natural experiment, data collection systems on exposure, outcomes and demographics must be in place or readily deployable. Such surveillance systems are central to the success of many accountability assessments.

Findings from a successful natural experiment are generally more credible than those from other types of observational studies in that little or no adjustment for selection and confounding is necessary. For this reason, natural experiments should be afforded weight when considering the causality of air pollution associations. Although these found opportunities should be exploited, however, they will never stand alone as accountability assessments. For example, the findings of studies based on the Utah mill closure (Pope 1989, Chapter 4 and this chapter) may be internally valid in that they provide accurate estimates of the effects that occurred under those specific conditions. But the observed reduction of PM concentrations may not be externally valid or generalizable. The estimates these data provide may have limited relevance to the magnitude of concentration reductions that would result from intentional regulatory efforts to improve air quality.

The often-cited example of a natural experiment in the field of air pollution health effects is the labor strike at a steel mill in the Utah Valley that resulted in closure of the mill: PM concentrations dropped sharply while the mill was closed (Pope 1989; see Chapter 4). Even in such cases, however, one still needs to ensure that other factors that could confound the observed effect did not occur in concert with the randomized change in exposure. For example, confounding will occur if unemployed people do not seek medical care as readily as when they are employed and a measure of medical-care usage (hospitalizations) is the outcome of interest.

Migration Studies

People may change their exposure to air pollution by changing their place of residence. If one can assume that the probability of a person moving to a particular location is unrelated to his or her susceptibility to air pollution, studies in which health indicators are tracked before and after the move can be informative. This assumption may have been tenable in early studies of lung cancer among migrant populations. These studies suggest that individuals who emigrated from countries with high air pollution levels to countries with lower levels develop lung cancer at rates higher than lifetime residents of the new country but lower than lifetime residents of their country of origin. These studies could not, however, account for differences in cigarette smoking and occupational risk factors (see Reid et al 1966).

More recently, Avol and colleagues (2001) studied 110 adolescents who had moved from 10 California communities that were participating in a 10-year follow-up study of air pollution and respiratory health in children. They observed that children who moved to areas with higher levels of PM less than 10 μm in aerodynamic diameter (PM_{10}) experienced, on average, decreased growth of lung function (measured as the forced expiratory volume in one second [FEV_1]). Those who moved to areas with lower levels of air pollution experienced, on average, relatively increased growth of lung function. In these communities, a relation between susceptibility and propensity to move to a particular area may be tenable because of the widespread awareness of the air pollution problem in southern California. However, these patterns were particularly pronounced when at least 3 years had elapsed between the move and the measurement of lung function.

MODEL-BASED PREDICTION BY COMBINING INFORMATION

Central to evaluating health effects of regulations is delineation of counterfactual environments (eg, circumstances of exposure that would result in the absence of regulation). Generally, these will be produced from models, with the most straightforward being an exposure-response (or concentration-response) model to compare observed

health effects to those predicted for different (generally higher) exposures or concentrations. The observed and expected values must be compared relative to an error term that incorporates appropriate uncertainties. Failure to identify important sources of uncertainty and quantify their impact (a common situation) produces comparisons that invite overinterpretation.

Effective accountability assessments will usually require combining information from a variety of sources. Quantitative synthesis (widely referred to as *metaanalysis*) adds value by formalizing the process, documenting assumptions, and incorporating the insights of diverse scientific disciplines. Information to be combined comes from a variety of research settings (experimental and observational): both human and animal studies from field, clinical, and laboratory settings. All relevant and valid evidence should be examined so that bias is not introduced through choice of evidence. Potential biases of individual studies should be identified and, if possible, findings of studies with offsetting biases should be synthesized. Generally, using quality assessments to down-weight particular studies is inappropriate when pooling their results. Preferably the analysis characterizes the contributions from different studies and takes advantage of possible triangulation of findings from studies having different and possibly opposing strengths and weaknesses.

Research synthesis will also be central in interpreting findings of accountability assessments. Evaluation of regulations will require combining evidence over all relevant databases, similar or not. Syntheses need to include all relevant human information while also incorporating data from other lines of investigation, such as human exposures, which provide insights into mechanistic phenomena and time frames of biological processes. Ideally, including such information would be done formally by an appropriate multidisciplinary team. Important issues to consider in conducting syntheses include measurement error, lining up of studies (ie, comparison of studies in a uniform fashion or on a uniform scale[s]), and ensuring that all sources of uncertainty are tracked and reported. Combination always involves both formal and informal assessments. Bayesian statistical methods are one useful approach for documenting assumptions and ensuring that appropriate uncertainties are identified and quantified (see sidebar, p 82). Although Bayesian, hierarchical modeling may be effective in this regard, it is not a panacea.

ANALYTIC ISSUES

Estimation, confidence intervals, and other statistical summaries that provide quantitative assessments are necessary for accountability. In evaluating the health effects of air pollution regulation, however, conventional hypothesis testing and generation of *P* values is largely irrelevant. We know a priori that exposure and health risk are related on the basis of the evidence that motivated the regulation.

There is no so-called statistical free lunch, which is to say that even elegant statistical models may be incorrectly specified, data may be flawed, and ultimately, results of an analysis may be no better than information already available. Space-age methods will not rescue stone-age data. However, sophisticated statistical analysis can make the best use of available information and indicate where information is most needed. Even when formal modeling is not sufficient for the task, possibly disparate sources of information must be linked to document assumptions and provide a platform for sensitivity analyses.

Uncertainty

Even when studies address similar endpoints in similar study populations, findings and conclusions will often be heterogeneous. Some heterogeneity in results from statistical variation, as no study has an infinite sample size, and some heterogeneity in the conclusions may result from different approaches to interpreting the same findings. However, a considerable amount of heterogeneity has been shown to be due to other factors. These factors include heterogeneity in design (reference population, study units [eg, individuals or groups of individuals], method of assigning treatments [*treatment* used in a generic sense], type of treatment, dependent variables, extent of follow-up, method of measurement), general study quality, differential measurement error of exposures and other inputs, statistical analysis (analytic framework [basic, hierarchical, Bayes], analytic model [Gaussian, Poisson, linear, log-linear], covariate adjustments, use of propensity scores, method of computing standard errors, multiplicity of endpoints and analyses), relation of analysis to underlying truth (that between-study heterogeneity can be induced or suppressed by incorrect analysis), and reporting.

Comprehensive and quantitative treatment of these and other sources of uncertainty have been more common in risk assessment (Morgan and Henrion 1990) than in epidemiologic analysis, but this situation may be changing. The recent report by the US National Academy of Science recommended that the EPA conduct such assessments in future studies of health benefits of air quality regulations (US National Research Council 2002).

Two kinds of uncertainty—sampling and nonsampling—permeate an accountability assessment (Morgan and Henrion 1990). Although the boundary between the two is fuzzy, a distinction is possible. Generally, sampling uncertainty concerns inherent variability that comes from

not having infinitely large samples. Even this narrow form of uncertainty can be substantial. For example, in a basic model, the standard deviation of a sample mean (sometimes termed the *standard error of the mean*) is proportional to $1\sqrt{n}$, where n is sample size. Therefore, to reduce the standard error by 50% (and thereby reduce the width of a confidence interval by 50%), sample size must be increased by a factor of four. In hierarchical models, which are increasingly applied in air pollution epidemiology, these relations are even more complicated. But for a given model specification, sampling errors are relatively easy to accommodate using standard statistical approaches; they may be a source of only minor uncertainty.

Nonsampling errors are of equal or greater importance. These include uncertainty in model specification (eg, baseline model, covariates and how they are included in the model, and structure of measurement error), prior distributions (for a Bayesian approach), and loss functions that either implicitly or explicitly guide decision making.

The leverage of these uncertainties is far greater when interpreting observational studies because, unlike randomized studies, confounding and other selection effects must be correctly modeled to produce a valid causal inference. However, uncertainties have high leverage in any study in which biomedical theory is insufficient to pin down a structural model and evidence is insufficient to make definitive conclusions. In a Bayesian context, if data are insufficiently informative, prior distributions and loss functions can dominate the decision process. That is, the new evidence will have little weight in changing interpretation of the full body of evidence. Because strong evidence is unlikely to come from individual accountability assessment studies, synthesis will be necessary (but probably not sufficient) to move widely discrepant prior beliefs to a common ground.

Uncertainties from nonsampling errors can be incorporated into sampling models (Morgan and Henrion 1990). For example, in a polynomial regression model for risk in relation to exposure, one can use standard model-selection techniques to select or put a prior distribution on the polynomial degree, produce a posterior distribution for it and other parameters in the model, and either use the posterior distribution to select a model or average over the posterior distribution to produce a Bayesian model averaged polynomial. One benefit of this approach is that it reports greater uncertainty in the regression equation than does a simple model. Analysts still must select a prior distribution, however; this choice remains a source of (nonsampling) uncertainty in the analysis.

Many inherent uncertainties in data cannot be resolved by analytic models. Some types of uncertainty are quite difficult to embed in a sampling model. For example, using information on rodent bioassays in risk assessments for humans involves decisions that are primarily in the nonsampling domain.

Sensitivity Analysis

In all studies, but especially in observational studies, results can be sensitive to choice of statistical model. Modeling choices include variables used to adjust for confounding, specifications of these variables in the model, and model form (eg, linear or log-linear). Therefore, sensitivity analysis is needed to describe the degree of sensitivity of results to model form and specification. Sensitivity analyses will generally require multivariate modeling, not only to accommodate the possibly increased leverage of several inputs changing together but also to build in possible attenuation of such leverage. For example, in a research synthesis, a study may have two biases: one tending to inflate an estimated effect, the other tending to attenuate it. If only one of these biases is active in a particular analysis, the study may have high leverage. However, if both biases are of the same magnitude, large or small, then the net effect will be small and only a bivariate sensitivity analysis that builds in this positive correlation will produce the true, relative insensitivity.

Approaches to sensitivity analysis run the gamut from relatively simple to statistically complex. Sensitivity of results to underlying assumptions and model choices can be described by reporting results for various scenarios, called a *scenario approach*. Data plots and other graphic methods are particularly useful in this regard; for example, Dominici and colleagues employed a sensitivity analysis of degrees of freedom used to adjust for time trends in mortality rates in the US National Morbidity, Mortality, and Air Pollution Study [NMMAPS]) (Dominici et al 2003; HEI 2003). Bayesian model averaging offers a more formal approach, in which a prior distribution is put on model choices and the posterior expected model is reported along with its uncertainty. Both approaches are important and effective; each has strengths and weaknesses. The scenario approach is relatively easy for nonstatisticians to understand but generally gives no sense of the relative viability (posterior weights) of candidate scenarios. Bayesian model averaging does consider important uncertainties in alternative models but is less transparent to nonexperts. Analysts might consider an approach with the benefits of both, in which scenario-specific results are compared graphically with estimates from Bayesian model averaging (and their uncertainty) and with posterior weights for several scenario-weight specifications specified a priori (eg, equal weighting, low-dimension models weighted more).

BAYESIAN METHODS

The Bayesian approach to statistical design and analysis is both a philosophy and a technology (see Carlin and Louis 1996; Wilson 2001; Armitage et al 2002). The philosophy operates on the premises that personal opinion pervades all of science, that such opinions should be made explicit before analyzing the next information component, that unknowns are random variables with probability distributions (called the *prior distribution*), and that a formal system should be used to update prior opinions in the context of additional information. Such updating is accomplished by the Bayesian formalism, which requires a prior distribution for all unknown variables and a sampling distribution (data likelihood) that describes the probability distribution of observed data, conditional on values of unknowns. Once data related to a particular association have been observed, the updated distribution of unknowns (termed the *posterior distribution*) can be computed using Bayes Theorem. This distribution can be used to make inferences and to serve as the prior distribution in analyzing the next component of information. This multistage, Bayesian formalism does not depend on the so-called personal probability interpretation; it provides an analytic strategy in situations wherein the prior distribution is based on empirical evidence, is induced by random selection of experimental units, or is simply a formal component of a statistical model.

The Bayesian formalism is effective in integrating information, stabilizing estimates, and tracking all relevant uncertainties. For example, in the NMMAPS 90-city analysis (Samet et al 2000a,b), city-specific estimated slopes of mortality risk versus PM concentration are modeled by using a two-stage process: the true city-specific slopes come from a normal prior distribution and, conditional on a true city-specific slope, the direct, city-specific slope estimate is normally distributed with mean equal to that slope and variance equal to that of the direct estimate. The mean and variance of the prior distribution are estimated (empirical Bayes); the posterior mean of the city-specific slope lies between the direct estimate and the estimated prior mean; the posterior variance is smaller than that for the direct estimate (shrinkage and variance reduction). Mathematical analysis, simulation, and empirical examples show that these Bayesian estimates outperform direct estimates in a broad class of applications.

Properly structured, the Bayesian approach builds on traditional modeling. For example, a standard regression model is fit, but rather than using its estimates as the predictor, predictions move partway from regression model estimates toward direct estimates; the degree of movement depends on the relative precision of the regression prediction and the direct estimate. Predictions for very stable direct estimates are close to the direct estimate, whereas predictions for imprecise direct estimates are close to the regression model estimate.

The Bayesian formalism also allows incorporation of relevant uncertainties. For example, in NMMAPS, the standard error of the prior mean (the so-called population mean) included both variability of direct estimates and between-city variability in true slopes. The formalism allows this percolation of variability in far more complicated settings, producing an honest assessment of uncertainty.

Model Uncertainty

Accountability assessment is subject to considerable uncertainty induced by sampling variability (eg, standard error of an estimate) and model uncertainty (eg, uncertainty regarding the model used to conduct a causal analysis). The boundary between sampling uncertainty and model uncertainty is fuzzy; a Bayesian analysis is effective in measuring and incorporating both. For example, in a polynomial regression one can put a prior distribution on the polynomial degree and produce a posterior distribution for it and other parameters in the model. The average across the posterior distribution produces a Bayesian model averaged polynomial. The posterior variance of this average reports greater uncertainty in the regression equation than does model selection. This approach has been used as an alternative to selecting degrees of freedom in a spline-based adjustment for potential confounding in time series studies of air pollution (F Dominici, personal communication, 2003).

Incorporating Expert Opinion

Inevitably, little or no direct empirical information is available about some components of an accountability system or about what components should be included. As a result, incorporating expert opinion and structuring complex systems and synthesis will be necessary. The Bayesian formalism is well-suited to incorporating expert opinion on possible values for parameters and systems of related components. There is a large literature concerning the elicitation of expert opinion (Morgan and Henrion 1990). Advice therein includes seeking information about quantities on which experts have information (eg, potential observables) and documenting between-expert uncertainty. Note that if little or no empirical information is available to update these a priori assessments, the posterior distribution will simply reflect prior uncertainty.

The Bayesian approach, therefore, can structure complex systems (exposures that induce short-term response, longer-term frailties, and susceptibility to high levels of air pollution) and integrate information from disparate sources (see DuMouchel and Harris 1983).

Bayesian Approaches and Models

The foregoing discussions of uncertainty and sensitivity analysis underscore that an effective accountability assessment requires a broad synthesis of available information. Such a synthesis must account for relations among data sources, build in appropriate components of variability among data sources, track and incorporate important sources of uncertainty and bias, and integrate misaligned information. Misalignment can be spatial (some information at a coarse spatial resolution, some at a fine resolution) or can relate to available covariate information (eg, information sources vary with the types of covariates collected and coding thereof). An accountability assessment must deal with measurement error in inputs and outputs. To properly inform a cost-benefit analysis, an accountability assessment must produce summaries that go beyond point estimates and standard errors, delivering full joint distributions of relevant quantities so that expected utilities and other tools for decision analysis can be computed. An accountability assessment should be founded on a platform that facilitates sensitivity analyses. Finally, by necessity, an accountability assessment could incorporate expert opinion to substitute for direct information (eg, for links in a causal chain). These opinions must be explicit and incorporate the heterogeneity of opinion. Although it is by no means a panacea, the Bayesian approach enables syntheses with all of these features (see sidebar).

The major benefits of the Bayesian approach for accountability assessments are structuring the combination of evidence and explicitly bringing in expert opinion about the state of the evidence. The NMMAPS study (Samet et al 2000b) illustrates the value of the Bayesian approach in analysis of data from multiple locations, which will most likely be a feature of future accountability studies. Explicit incorporation of expert opinion falls almost completely outside traditional modeling. At the same time it may be an important component of dealing with some sources of uncertainty in accountability analyses, such as the choice or components of models (eg, whether a low-dose curve is linear, sublinear, or supralinear) (see sidebar).

Causal Models

Accountability assessments inevitably use available information to address so-called counterfactual questions, such as the following: What would have happened to air quality and human health in the absence of regulations or under different, less stringent regulations? Such counterfactual questions form the basis for the approach taken by EPA in its Section 812 Reports (EPA 1997, 1999) to estimate health benefits of the 1970 Clean Air Act and its Amendments (see Chapter 2). Although some relevant information on counterfactuals is usually available (eg, air pollution and health effects in other countries and at other times in the United States), it is not sufficient to pin down cause and effect relations between regulations and human health. Some aspects of short-term effects of air pollution in a small number of people can be assessed experimentally with controlled exposures of volunteers; long-term effects can be assessed experimentally in animal systems. However, most assessments of exposure-response relations depend on observational information, wherein confounding is a threat to validity and causal analysis should be used to uncover causal effects.

The goal of accountability assessment is to understand the associations among changes in regulations, resultant changes in air pollution, resultant changes in personal exposures, and changes in risks for adverse health effects. People can at least partially control their own exposure through their actions, and these actions can be associated with sensitivity to exposure. If information on personal attributes that influence exposure is available, covariate adjustment can reduce bias. Success of the adjustment depends on using a correct model form, and more importantly, having measured the covariates that matter. Causal modeling provides a possibly effective approach to determining associations among personal attributes, exposures or treatments, and health effects.

To understand the association between exposures and outcomes, analysts must adjust for associations between personal attributes and behaviors that influence both exposure and outcomes. Failure to do so may yield a biased analysis, one that does not provide valid, exportable information. For example, people experiencing asthma attacks during high-ozone periods may reduce their outside activities and thereby dramatically reduce the delivered dose. In extreme cases, then, data could indicate that high ozone levels are protective.

Such behavior modification is a type of so-called selection effect, which constitutes the principal threat to validity of a naïve analysis. Examples from clinical management of disease provide illuminating examples. In the early days of human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) epidemic, clinical indications for use of azidothymidine (AZT) were a $CD4^+$ T cell (CD4) count below 200/μL blood or an AIDS-defining illness. The mean CD4 level for patients taking AZT were lower than levels for those not taking AZT, which suggested that AZT was the cause of the lower CD4 levels. Of course, this association was in fact a product of clinical management decisions and not of biology; a different clinical management policy would have produced a different association.

Causal analysis attempts to eliminate or reduce such biases by jointly assessing the association between personal attributes, treatments (eg, exposures), and outcomes to yield results that are free of such selection effects. In the AZT example, statistical modeling of follow-up information would show that, for a similar initial CD4 level, the decline in CD4 for patients taking AZT is less steep that for those not taking AZT; this result is evidence for the appropriate causal effect. If follow-up information is not available, then no empirical evidence is available to untangle policy from biology.

A variety of approaches to causal analyses are possible, all of which attempt to adjust for selection effects and more general confounding. These approaches include use of instrumental variables (common in economics [Angrist et al 1996]) and use of counterfactuals (such as: What if this population had been exposed to a different pollution level?; or What if a regulation had not been implemented?), coupled with propensity scores and standard covariance adjustment (commonly used in biomedical research [Rosenbaum and Rubin 1983]). Marginal structural modeling (Robins et al 2000) is another popular approach.

The propensity score (first introduced to adjust sample surveys) models the association between individual attributes and treatment. If the determinants of propensity are known with reasonable certainty, then comparing outcomes

for individuals with the same propensity scores and then aggregating these comparisons allows an analysis free of selection effects. Indeed, if the propensity score is known, the scientific validity of the analysis is competitive with that of a randomized study but the propensity score is never fully known in studies that are not randomized. Developing effective propensity scores depends on having measured relevant covariates at an acceptable level of precision (ie, no unmeasured confounders). When using propensity scores, the best strategy is to model aggressively, conduct vigorous sensitivity analyses, and warn that key covariates may have been missed.

Generally, a fully convincing, fine-grained, empirically based causal analysis is impossible with available information. Even Pope's study of the closing of the Utah Valley steel mill (Pope 1989), in many ways the ideal natural experiment, is not fully convincing. For example, information on the prevalence of influenza was not available and reduced hospital admissions could have been due to a relatively mild flu season. However, hypothesized confounding by an epidemic of respiratory syncytial virus was not confirmed (Lamm et al 1991; Pope 1996). Because complete information is rarely available, analyses should be cautiously aggressive. Much modeling will go beyond the data, which in turn requires careful structuring using expert opinion coupled with aggressive sensitivity analyses.

Causal analysis, however, is no substitute for studies that can provide direct evidence about important components of a causal chain. Indeed, causal analysis can help identify the most informative, new, high-leverage information.

Ecologic Data and Disaggregation

Air pollution epidemiology frequently uses information at various levels of aggregation, from individual to national. The American Cancer Society study, for example, combined air pollution measurements at the city level and mortality and cardiovascular and respiratory disease risk factor data (such as cigarette smoking and occupation) at the individual level (Pope et al 2002). Integrating such misaligned data will require multivariate models indexed for space and time that can accommodate such data structures. Dealing with aggregation levels and misalignment is critical.

Most analysts use nonlinear models that must be applied to aggregated data with care to ensure that the feature being estimated does not depend on the degree of aggregation. For example, the exposure-response slope from a Poisson model applied to aggregate data does depend on degree of aggregation. Such dependence introduces heterogeneity between studies that used different levels of aggregation. Between-study heterogeneity should be addressed in accountability assessments, such as by using an approach that builds an individual-level model (eg, Poisson model) and then accounts for aggregation statistically. In this way, the exposure-response parameter retains its individual-level interpretation irrespective of degree of aggregation. These technical details need attention and must be made transparent when model estimates are used in accountability assessments.

Measurement Error

Some degree of measurement error, particularly in estimating exposure, is unavoidable in observational studies (Rothman and Greenland 1998b). Measurement error can produce attenuated estimates of the underlying exposure-response relation and may produce either overestimates or underestimates of true effects, depending on the nature of the error. Measurement error may also contribute to apparent between-study heterogeneity. Properly accounting for effects of measurement error is needed, both to accurately compare the results of available studies and to produce a valid assessment of the effect of air pollution regulation on public health. Hierarchical models are effective in this regard. They can be structured to include the measurement error process into the estimation procedure, automatically adjust estimates for error, and automatically report uncertainties that reflect influences of measurement error. Improved understanding of measurement error in air pollution studies is necessary. Reduced measurement error will reduce the need for statistical adjustments and their associated variance inflation (Zeger et al 2000).

Statistical Power

Statistical power refers to the ability of a study to detect (with a given level of precision) an effect of a given size. Statistical power depends on the signal to noise ratio; misspecification of either the signal or the noise can result in a false power analysis. A power analysis is informative if the sample size required to detect a small increase in rates of health outcomes is achieved. Consider the best case: a background rate of a disease is known to be zero and a study is evaluating whether a specific dose is above or below a response threshold. To detect a response probability of p with confidence of at least 0.95, the required sample size is $3/p$. For example, to detect a 1/10,000 change in rate of disease, $n = 30{,}000$. An even larger sample is required when the background rate must be estimated, when the disease rate is not completely determined by the target pollutant, or when covariates need to be adjusted to reduce confounding.

Subpopulations such as sensitive individuals, people who smoke, or people of certain ages or occupations

should be considered. In general, even studies with high precision for an overall relation or effect will not necessarily have sufficient precision for estimating these variables within subgroups. To do so would necessitate combining evidence by using regression models and Bayesian structuring or some other means.

RECOMMENDATIONS FOR FUTURE RESEARCH

Maximizing the effectiveness of accountability assessments depends on combining prospective, designed experiments and designed observational studies and taking advantage of available observational data. Information about humans and other animal and plant species will all contribute to accountability assessments. Such information will need to be linked by a conceptual model that can be used to identify high-leverage information driving the disease burden estimated to result from exposure to air pollution.

Formalizing design and evaluation of accountability assessments by use of statistical and decision models can quantify the value of new information to inform scientific and policy aspects of a regulatory decision, which would result in an accountability assessment of research rather than of regulations. Accountability assessments depend on an effective information infrastructure coupled with properly designed, implemented, and analyzed studies. Therefore, accountability of research and other information gathering that are needed to inform and evaluate regulations goes hand in hand with accountability of regulations.

Success of future research requires systematic identification of research needs and opportunities and commissioning studies to address them. This research will likely entail both adaptation (or tuning) of existing methods to suit specific needs and development of long-term surveillance of both health outcomes and potential confounders.

We are optimistic about prospects for future research. We reiterate, however, the need for realistic expectations, in light of considerable challenges involved in measuring the health impact of air quality regulations. Stakeholders' expectations are high, and as a result of past regulatory successes, current regulations pertain to ambient concentrations at which health effects are expected to be small. We can expect to be able to detect some evidence of changes in pollutant emissions, concentrations, and exposures relatively soon after regulations are implemented. Detecting the health impact, especially effects on mortality and morbidity from chronic cardiovascular and respiratory diseases, will probably be more difficult. It may only be possible over time as complete implementation of regulations is achieved, exposures are reduced, and the biological effect of those reductions occurs. Both short-term and long-term research approaches will therefore be needed.

SHORT-TERM RECOMMENDATIONS

Reviews of Available Information

A comprehensive and detailed review of studies and settings in which information is already available can provide information relevant to accountability assessment and provide guidance as to additional information that is needed. Findings from existing natural experiments, such as those from the Utah Valley study of the steel mill closure (Pope 1989), for example, need to be reviewed for validity and relevance of the findings. Focused, immediate-impact, air pollution regulations and policies have been implemented in various settings, such in Dublin (Clancy et al 2002) and Hong Kong (Hedley et al 2002). These should also be critically reviewed for relevance and potential impact. Rapid retrospective review of relevant health indicators can then be used to address the health impact of such macrolevel intervention.

Formal Analysis of Existing Studies

Studies that address critical uncertainties would exert particularly great leverage with regard to decision making. Formal analysis of existing studies to determine critical gaps, and studies that could address them, may be useful. This analysis should precede and be incorporated into the research planning process.

Requests for Applications

HEI (2002) and the California Air Resources Board (2001) have both recently issued Requests for Applications seeking studies of the health impact of air quality regulations. Because one round of requests will probably not elicit the needed research, funding agencies should adopt a long-term perspective that incorporates periodic progress assessments for such research and sharing of data among funding agencies and researchers.

Advance Opportunities

Mechanisms should be implemented for identifying in advance settings that may be impacted by changes in air pollution regulations or policies. To do so, funding agencies will probably need to contact regulators, and other governmental agencies, and affected communities. In addition, research models that facilitate cooperation in

anticipating, planning, and conducting accountability studies should be developed.

LONGER-TERM RECOMMENDATIONS

Serial Cross-Sectional Studies

Although serial cross-sectional studies might initially be considered longer-term studies, some could conceivably be completed within relatively short time frames. For example, cross-sectional survey data on many health outcomes of interest are available, some of which are routinely, serially collected (eg, US National Health and Nutrition Examination Survey [NHANES]). A comparable cross-sectional study can be repeated at a point in time after an air pollution intervention. The repeat study or studies would be timed to occur after the postulated latency period between any change in exposure and the health outcome of interest. For example, serial cross-sectional studies were performed before and after the reunification of Germany, exploiting the dramatic changes in air pollution concentrations after reunification (Heinrich et al 2002).

Randomized Studies

Randomized studies should be aimed at providing information on one or several links in the chain of accountability (Figure 5.1) in a rigorous manner. Randomizing exposure in real-life settings through randomized behavioral manipulation (eg, provision of air conditioners or indoor versus outdoor exercise regimens) is one approach to providing information on the connection between exposure to air pollution and health effects.

Cohort Studies

Ongoing cohort studies may be able to provide limited information for accountability by using a windows-of-exposure approach to gain insight into the effects of exposure in different time periods and extent of pollution exposure on an outcome, typically mortality (eg, Lipfert et al 2000; Pope et al 2002). With HEI support, continuing analyses of the American Cancer Society cohort (Pope et al 2002) and The Netherlands Dietary Cohort (Hoek et al 2002) are using decades-long residence histories to study timing of pollution exposure relative to mortality. Migration studies may also contribute information, although the need to account for factors related to the health outcome that may have influenced the decision to relocate will be challenging.

Analytic Approaches

Analytic approaches utilize available data to address both macrolevel and microlevel questions. Examples of such approaches include: (1) syntheses in which one of several approaches (including triangulation) is taken to combine data across studies, (2) model-based predictions comparing predicted effects to observed effects while accounting for model uncertainty, and (3) causal analyses. When such analyses appropriately address uncertainty, their findings can provide insight into information required to make more valid accountability assessments (forms of leverage analyses).

Surveillance Systems

Many of the types of studies described above could be greatly facilitated by data collected through ongoing surveillance of the major time-varying links in the chain of accountability, or at least those components that concern exposure and health outcomes. In fact, direct assessment of the health impact of air quality regulations is probably unlikely on state, regional, or national scales without implementing such systems. Some surveillance elements already exist: the US National Centers for Health Statistics, the NHANESs, and the EPA's and California's air-monitoring networks.

A system for long-term surveillance of the health impact of air quality regulations will, however, initially require an evaluation of the adequacy of the existing surveillance elements. This evaluation would also need to consider what kinds of information are required for long-term evaluation of the health impact, including goals for efficacy and effectiveness. Fortunately, evaluations are already underway. The US Centers for Disease Control's Environmental Public Health Tracking Program has already begun to develop plans for a national network that will "(1) be standards-based; (2) allow direct electronic data reporting and linkage within and across health effect, exposure, and hazard data; and (3) interoperate with other public health systems," (US Centers for Disease Control and Prevention 2003). The US Congress is also considering legislation (the National Health Tracking Act) to establish a comprehensive system for identifying and monitoring chronic diseases and possibly correlating their causes with environmental, behavioral, socioeconomic, and demographic risk factors. Linking efforts to design air pollution accountability studies with these national efforts is critical.

The planning process for the use of surveillance approaches in accountability studies of air pollution could begin with a series of workshops to bring together academic researchers and public health and air-regulatory officials at state and federal levels.

REFERENCES

Angrist JD, Imbens GW, Rubin DB. 1996. Identification of causal effects using instrumental variables (with discussion). J Am Stat Assoc 91:444–472.

Armitage P, Berry G, Matthews JNS. 2002. Statistical Methods in Medical Research. Blackwell Science, London, England.

Avol EL, Gauderman WJ, Tan SM, London SJ, Peters JM. 2001. Respiratory effects of relocating to areas of differing air pollution levels. Am J Respir Crit Care Med 164:2067–2072.

California Air Resources Board. 2001. Request for Proposals: Health Benefits of Incremental Improvements in Air Quality. California Air Resources Board, Sacramento CA.

Carlin BP, Louis TA. 1996. Bayes and Empirical Bayes Methods for Data Analysis. Chapman and Hall/CRC, Boca Raton FL.

Centers for Disease Control and Prevention (US). 2003. Environmental public health tracking program (last updated 6/24/03). *www.cdc.gov/nceh/tracking/default.htm*. Accessed 7/21/03.

Clancy L, Goodman P, Sinclair H, Dockery DW. 2002. Effect of air pollution control on death rates in Dublin, Ireland: An intervention study. Lancet 360:1210–1214.

Cohen J, Brion G, Haines J. 1990. Review of the National Ambient Air Quality Standards for Lead: Assessment of Scientific and Technical Information. EPA 450/2-89/022. NTIS PB91-206185. US Environmental Protection Agency, Research Triangle Park NC. Available from *www.ntis.gov*.

Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG Jr, Speizer FE. 1993. An association between air pollution and mortality in six US cities. N Engl J Med 329:1753–1759.

Dominici F, McDermott A, Daniels M, Zeger SL, Samet JM. 2003. Mortality among residents of 90 cities. In: Revised Analyses of Time-Series Studies of Air Pollution and Health, pp 9–24. Special Report. Health Effects Institute, Boston MA.

DuMouchel WH, Harris JE. 1983. Bayes methods for combining the results of cancer studies in humans and other species (with discussion). J Am Stat Assoc 78:293–315.

Environmental Protection Agency (US). 1991. Air Quality Criteria for Carbon Monoxide. EPA/600/8-90/045F. NTIS PB85-103471/HSU. Office of Health and Environmental Assessment, Research Triangle Park NC. Available from *www.ntis.gov*.

Environmental Protection Agency (US). 1992. Review of the National Ambient Air Quality Standards for Carbon Monoxide: 1992 Reassessment of Scientific and Technical Information. EPA-452/R-92-004. NTIS PB93-157717. Office of Air Quality and Standards, Research Triangle Park NC. Available from *www.ntis.gov*.

Environmental Protection Agency (US). 1997. The Benefits and Costs of the Clean Air Act, 1970 to 1990. EPA-410-R-97-002. Office of Air and Radiation, Washington DC.

Environmental Protection Agency (US). 1999. The Benefits and Costs of the Clean Air Act 1990 to 2010: EPA Report to Congress. EPA-410-R-99-001. Office of Air and Radiation, Washington DC.

Gamble JF. 1998. $PM_{2.5}$ and mortality in long-term prospective cohort studies: Cause-effect or statistical associations? Environ Health Perspect 106:535–549.

Grievink L, Zijlstra AG, Ke X, Brunekreef B. 1999. Double-blind intervention trial on modulation of ozone effects on pulmonary function by antioxidant supplements. Am J Epidemiol 149:306–314.

Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. 1995. Users' guides to the medical literature: IX. A method for grading health care recommendations. JAMA 274:1800–1804.

Health Effects Institute. 2002. Request for Applications: Winter 2002 Research Agenda. Health Effects Institute, Boston MA.

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

Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. 2002. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: An intervention study. Lancet 360:1646–1652.

Heinrich J, Hoelscher B, Frye C, Meyer I, Pitz M, Cyrys J, Wjst M, Neas L, Wichmann H-E. 2002. Improved air

quality in reunified Germany and decreases in respiratory symptoms. Epidemiology 13:394–401.

Hill AB. 1965. The environment and disease: Association or causation. Proc R Soc Med 58:295–300.

Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA. 2002. Association between mortality and indicators of traffic-related air pollution in the Netherlands: A cohort study. Lancet 360:1203–1209.

Lamm SH, Hall TA, Engel A, White LS, Rueter FH. 1991. Assessment of viral and environmental factors as determinants of pediatric lower respiratory tract disease admissions in Utah County, Utah (1985–1989) (unpublished report). Consultants in Epidemiology and Occupational Health, Washington DC.

Lipfert FW, Morris SC. 2002. Temporal and spatial relations between age specific mortality and ambient air quality in the United States: Regression results for counties, 1960–97. Occup Environ Med 59:156–174.

Lipfert FW, Perry HM Jr, Miller JP, Baty JD, Wyzga RE, Carmody SE. 2000. The Washington University-EPRI veterans' cohort mortality study: Preliminary results. Inhalation Toxicol 12(Suppl 4):41–73.

Morgan MG, Henrion M. 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, Cambridge, United Kingdom.

National Research Council (US). 2002. Estimating the Public Health Benefits of Proposed Air Quality Regulations. National Academy Press, Washington DC.

Ostro B. 1993. The association of air pollution and mortality: Examining the case for inference. Arch Environ Health 48:336–342.

Pope CA III. 1989. Respiratory disease associated with community air pollution and a steel mill, Utah Valley. Am J Public Health 79:623–628.

Pope CA III. 1996. Particulate pollution and health: A review of the Utah Valley experience. J Expos Anal Environ Epidemiol 6:23–34.

Pope CA III, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287:1132–1141.

Public Health Service (US). 1964. Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service. US Department of Health, Education, and Welfare, Washington DC.

Reid DD, Cornfield J, Markush RE, Siegel D, Pederson E, Haenszel W. 1966. Studies of disease among migrants and native populations in Great Britain, Norway, and the United States: III, Prevalence of cardiorespiratory symptoms among migrants and native born in the United States. In: Epidemiological Study of Cancer and Other Chronic Diseases, Monograph 19 (Haenszel W, ed), pp 321–346. National Cancer Institute, Bethesda MD.

Reigelman RK, Hirsch RP. 1996. Studying a Study and Testing a Test: How to Read the Health Science Literature (3rd edition). Little, Brown & Co, Boston MA.

Robins JM, Hernan MA, Brumback B. 2000. Marginal structural models and causal inference in epidemiology. Epidemiology 11:550–560.

Romieu I, Sienra-Monge JJ, Ramírez-Aguilar M, Téllez-Rojo MM, Moreno-Macías H, Reyes-Ruiz NI, del Río-Navarro BE, Ruiz-Navarro MX, Hatch G, Slade R, Hernández-Avila M. 2002. Antioxidant supplementation and lung functions among children with asthma exposed to high levels of air pollutants. Am J Respir Crit Care Med 166:703–709.

Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55.

Rothman KJ, Greenland S. 1998a. Causation and causal inference. In: Modern Epidemiology, 2nd edition (Rothman KJ, Greenland S, eds). Lippincott-Raven Publishers, Philadelphia PA.

Rothman KJ, Greenland S. 1998b. Precision and validity in epidemiologic studies. In: Modern Epidemiology, 2nd edition (Rothman KJ, Greenland S, eds). Lippincott-Raven Publishers, Philadelphia PA.

Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000a. The National Morbidity, Mortality, and Air Pollution Study, Part 1: 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. 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.

Vedal S, Brauer M, White R, Petkau J. 2003. Air pollution and daily mortality in a city with low levels of pollution. Environ Health Perspect 111:45–51.

White C. 1990. Research on smoking and lung cancer: A landmark in the history of chronic disease epidemiology. Yale J Biol Med 63:29–46

Wilson JD. 2001. Advanced Methods for Dose-Response Assessment: Bayesian Approaches. Final Report. Resources for the Future, Washington DC.

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.

ABBREVIATIONS AND OTHER TERMS

AZT	azidothymidine
CD4	$CD4^+$ T cell
CO	carbon monoxide
EPA	Environmental Protection Agency (US)
NHANES	National Health and Nutrition Examination Survey (US)
NMMAPS	National Mortality, Morbidity, and Air Pollution study (US)
PM	particulate matter

Chapter 6. DIRECTIONS FOR FUTURE RESEARCH

INTRODUCTION

In preparing this Communication, the Accountability Working Group of the Health Effects Institute had several purposes, including elaborating the concept of accountability, setting out a framework for assessment of accountability, and advancing an agenda of research and methods development that would enhance accountability assessments. This final chapter offers recommendations for the research and methods agenda on the basis of the framework afforded by the chain of accountability and the methodologic and conceptual reviews of Chapters 3, 4, and 5.

The concept of accountability is fundamental to public health. Through surveillance mechanisms and other ways of watching for threats to health, problems are identified that require interventions to reduce risks, programs are implemented for this purpose, and the consequences of such programs are evaluated by continued tracking of risk indicators. These actions are expected to benefit the public health. The estimated magnitude of attributable burden of disease in each case determines the priority of the intervention. In the well-chronicled example of smoking and lung cancer, the epidemic of lung cancer was noticed by astute clinicians and also reflected in routine mortality statistics. After epidemiologic studies documented the causal link between smoking and lung cancer, an array of interventions was implemented, varying over time as smoking became better understood as an addiction that begins during adolescence and young adulthood. The burden of lung cancer caused by smoking and associated costs are periodically estimated, and mortality and incidence data are monitored nationally. These data now suggest a decline in smoking among men (US Department of Health and Human Services 1990).

As described in Chapter 2, the concept of accountability has been inherent in air pollution regulation in the United States, United Kingdom, and elsewhere, even though it was not explicitly stated until recently. Evidence used to support air quality standards for criteria or hazardous pollutants characterizes the risks of exposures experienced by the population; if the level of associated risk is not acceptable, then the level of the standard is changed (ie, an intervention is taken). The US Clean Air Act provides guidelines for evaluating the level of risk, calling for an adequate margin of safety, in particular for the criteria pollutants. Lowering the population's exposure along a dose-response curve is assumed to reduce the burden of disease caused by the pollutant of concern. For the criteria pollutants, the US Clean Air Act requires the development of state implementation plans (SIPs*) for areas not in compliance with the standards. These plans focus on source control and compliance is assessed by tracking concentrations of pollutants. Reducing a pollutant's concentration is assumed to initiate a cascade of consequences along the chain of accountability, resulting in a reduction of disease burden associated with the pollutant.

In general, regulatory agencies do not track the right-hand links of the chain of accountability (ie, those beyond ambient concentrations): changes in exposure, dose, or frequency of adverse health outcomes. Indirect assessments are made, however: Quantitative risk assessment may be used to calculate the burden of attributable disease at the time at which a standard is being considered; comparison might be made to the population exposures that would prevail after implementation of the new standard; projections might also be made of the disease burden that will be avoided if exposures decline in the future, rather than staying at current levels. The chapters of this Communication lay out the conceptual basis for such assessments and offer several examples, including cost-benefit analyses of the US Clean Air Act.

Chapter 5 distinguishes such long-term accountability assessments of complex environmental regulations from shorter-term interventions that can affect population exposures more abruptly or specifically, thereby facilitating evaluation of the consequences. Examples cited in this Communication include removal of sulfur from fuels in

This chapter is one part of Health Effects Institute Communication 11, *Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research,* which also includes five other chapters and an Executive Summary of the project.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R82811201 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 this chapter.

Hong Kong (Hedley et al 2002) and banning of coal sales in Dublin (Clancy et al 2002) and, subsequently, throughout Ireland. In these examples, changes in pollution concentrations were linked in time to interventions and changes in key health indicators that were tracked with hope of finding temporal trends consistent with the intervention.

Chapter 5 also describes so-called natural experiments, referring to changes in air pollution emissions or exposures resulting from actions that led to a clear change in population exposures. In these situations, assessing the consequences of a change in exposure is possible under circumstances that decouple the change in air pollution exposure from exposures to other factors. These natural experiments are closer to randomized controlled trials than are epidemiologic studies based on the usually less dramatic variations in exposure that occur temporally or geographically. Perhaps the best known example of a natural experiment is the series of studies based on the shutdown of the Geneva steel mill in Provo, Utah (Pope 1989). Such natural experiments may provide a higher level of evidence for inferring causality than other epidemiologic studies of air pollution do.

RESEARCH NEEDS AND OPPORTUNITIES

The following sections of this chapter set out recommendations for an agenda to advance understanding and assessment of accountability. After identifying current targets of opportunity for accountability research, specific recommendations are offered on the basis of discussions in Chapters 3, 4, and 5. This chapter concludes with several general recommendations. The recommendations are a wide ranging, extensive list of possible research opportunities. As HEI and other organizations move forward on accountability research, priorities among these opportunities need to be set.

TARGETS OF OPPORTUNITY

The continually changing regulation of air pollution in the United States, Europe, and elsewhere affords an immediate set of opportunities for accountability assessment on national, regional, and local scales.

Targets at US National and State Levels

Heavy-Duty Diesel/Low Sulfur Fuel Rule The US Environmental Protection Agency (EPA) recently promulgated regulations (EPA 2001) to reduce heavy-duty diesel-vehicle emissions via reductions beginning in 2006 and 2007 in fuel sulfur content and emission control technologies (eg, particle traps and various nitrogen oxides [NO_x] reduction technologies). As part of this rule making, benefits, defined in terms of reduced levels of $PM_{2.5}$ (particulate matter less than 2.5 μm in aerodynamic diameter) and ozone, and associated reductions of health risks were projected for 30 years after implementation. The anticipated reduction in sulfur content and vehicle emissions provides a possibly useful context in which accountability could be assessed into the future. Both costs and benefits of the regulation could be tracked over time during implementation and compared against cost and benefit trends predicted in the regulatory impact analysis. Consideration of data that may currently be lacking would allow gaps to be filled in advance of upcoming reductions.

$PM_{2.5}$ and Ozone NAAQS Implementation The SIP process is now in its initial stages for the recently promulgated US National Ambient Air Quality Standard (NAAQS) for $PM_{2.5}$ and ozone (EPA 1997). Meanwhile, extensive data on nationwide $PM_{2.5}$ concentrations are being collected from a new monitoring network, establishing baseline conditions against which future emissions reductions can be assessed. The SIP process for $PM_{2.5}$ and ozone could provide an opportunity for accountability assessments that address changes in emissions, ambient concentrations, and exposures or doses to the population. Both large-scale population surveys and targeted studies of susceptible subgroups could be useful. Additional monitoring undertaken as part of studies intended to measure the health impact of SIPs must accord with temporal and spatial scales of the SIPs themselves.

Many elements of the SIP process make it well suited to accountability analyses, although this potential has been largely untapped. These elements include modeled predictions of ambient air quality under planned emissions-control programs and extensive ambient monitoring over space and time. An historical record of prescribed emissions reductions, model-based projections of future air quality, and observed air quality now exists. A focused effort to analyze such data could increase the effectiveness of the SIP planning process. The SIP process now beginning for particulate matter should be considered an immediate opportunity for designing accountability studies. Researchers, funding agencies, and state regulatory agencies should soon begin to plan how best to take advantage of it.

EPA Air Toxics Control Plan Hazardous air pollutants, also termed *air toxics*, are pollutants that cause or may cause cancer or other serious health effects (such as reproductive ailments or birth defects) or adverse environmental and ecologic effects. EPA is required to assess risks and, if necessary, control the 188 air pollutants now classified as hazardous. Over the past seven years, EPA has

issued maximum achievable control technology (MACT) standards covering more than 80 categories of major industrial sources (such as chemical plants, oil refineries, aerospace manufacturers, and steel mills), as well as categories of smaller sources (such as dry cleaners, commercial sterilizers, secondary lead smelters, and chromium electroplating facilities). The agency has put into place important controls for fuels and vehicles; these controls are expected to reduce inventories of selected motor-vehicle air toxics by more than 75% from 1990 levels by 2020. EPA has also implemented programs that reduce indoor air toxics.

Relevant research for accountability might include longitudinal measurements of pollutant emissions and ambient concentrations and identification of health endpoints that could be tracked in the near term. This approach is most applicable for hazardous air pollutants associated with short-term responses (eg, irritants). If long-term changes are found in emissions and air concentrations, cancer rates could be tracked by using population-based registries, if carcinogens are of concern. The etiologic signal of a particular carcinogen may prove difficult to track, however, given the multiple causes of most cancers, and that among them the relative contribution of air pollution to the overall incidence may be small. Retrospectively, plant closures that might have led to major changes in population exposures to major airborne carcinogens could be identified and studied, although such studies are subject to the constraints and limitations discussed in Chapter 5.

Targets at Local Level

Relatively rapid changes in ambient concentrations may occur in a local area as a result of a major change in local-source emissions due to regulatory action (eg, closing downtown streets to traffic, installing new emission controls on a fleet of diesel trucks, converting a bus depot from diesel to natural gas, or closing down or adding controls to a large power plant). These types of interventions present opportunities for studies aimed at documenting cause-effect relations between emissions changes and changes in exposure or health. Because these interventions occur over relatively short times and small areas, assessment studies of them can be both economically and logistically feasible (although considerable challenges exist [see Chapters 3 and 5]).

Numerous opportunities exist for studies of such interventions throughout the United States and elsewhere. For example, the New York City Metropolitan Transit Authority has plans (Office of the Governor 1997) to convert bus fueling and storage depots from diesel to natural gas, thereby possibly reducing neighborhood levels of elemental carbon and other diesel-related particle components. Similarly, low-sulfur diesel fuel and particle-trap control technology are being implemented in fleets of diesel vehicles (including garbage trucks and school buses) in many locations. Recently, municipal authorities in London have proposed a series of interventions designed to reduce the impact of vehicular traffic on the general urban environment and on air pollution levels in particular (Mayor of London 2002). The most recent intervention involved the levying of a tax on any vehicle entering central London during certain times (Mayor of London 2002). Studies that document changes in ambient pollutant concentrations, personal exposures, and even health status in connection with such interventions could be quite informative. However, such studies must anticipate regulatory actions and should be planned in cooperation with the authorities implementing the intervention to the extent practicable.

FROM REGULATION TO EXPOSURE TO DOSE

Data collection directed at the left-hand links of the chain of accountability is embedded in some regulatory processes (eg, tracking of pollutant concentrations). In general, however, emissions changes are neither tracked nor are consequences of these changes assessed for exposures and doses received by the population. A few informative exceptions include carbon monoxide and lead, for which specific biomarkers (carboxyhemoglobin and levels of lead in blood, respectively) can be tracked in population samples. In general, studies are needed that track changes in these links of the chain.

Changes in Emissions in Response to Regulation

The EPA compiles emission factors and estimates total emissions data for a variety of chemical species and reports trends in these data and estimates on an annual basis (EPA 1998). Studies are needed that assess consequences of implementing emissions regulations and identify actual changes in emissions that follow regulations. Methods are needed for these purposes, with consideration of the range of sources that could be sampled and the approaches needed to monitor changes over time.

Changes in Nature of Air Pollution

The expansion of monitoring fine particles and subcomponents of PM (EPA 2002) will make possible better studies directed at identifying changes in the nature of ambient air pollution. Both intended and unintended consequences may result from air quality regulations and other actions. Development of maintained, detailed monitoring programs could enable monitoring of more of the air

pollution mixture. Accountability studies can and should be designed to take advantage of monitoring data. In particular, studies could include: (1) tracking the ambient impact of emissions changes for specific chemical species, thereby judging the validity of modeled projections and the effectiveness of corresponding control programs; (2) corroborating emission inventories; and (3) designing and validating techniques to attribute source contributions to ambient levels of particular species.

Changes in Human Exposure

The impact of regulations on population exposures should also be addressed to better understand how changes in ambient concentrations affect personal exposures, particularly for people considered at risk from particular pollutants. Efficient strategies are needed to take advantage of ongoing data collection efforts for this purpose.

National surveys afford opportunities to measure exposures to pollutants and levels of biomarkers of exposure or dose, if available. Continued research is needed to develop validated biomarkers for this purpose and apply them in appropriate population samples. Because of its large size, national scope, and representative sampling design, the US National Health and Nutrition Examination Survey (NHANES) is a model for exposure and dose assessment in the United States. Work is encouraged to develop new tools for measurement of air pollution exposure and dose for incorporation into future rounds of NHANES.

Given the importance of PM in current assessments of the health impact of air pollution, incorporating collection of $PM_{2.5}$ and chemical speciation data into future NHANES personal monitoring surveys would be especially valuable. This incorporation should strive to be as inexpensive and efficient as possible, given the constraints of large, national surveys. Use of personal monitoring would first require pilot demonstrations of new technology, including development of miniature, lightweight, battery-powered samplers capable of accumulating sufficient material for speciation analyses. Biomarkers of exposure and dose for important air toxics, such as benzene and butadiene, would also be valuable.

Regulatory interventions to improve air quality may result in changes of behavior among target populations that may in turn affect exposure and dose. Researchers designing studies of the health impact of air pollution regulations should anticipate the possibility of such changes. Measurement protocols (surveys) could be developed ex ante to characterize baseline behavior or exposure and then track changes in time-activity patterns that could affect exposure.

HEALTH OUTCOMES

A wide range of adverse health outcomes has been linked to air pollution, ranging from loss of well being and reduced comfort to increased risk of dying. Accountability assessments will be almost inevitably complicated by the nonspecificity of these outcomes and the possibility that other factors that increase risk for them are changing on a time course similar to that for the air pollutant(s) of interest. This problem can best be addressed by making available information about the other factors and using large databases that span diverse populations, so that estimated consequences of changes in air pollution exposures are less likely to be confounded. Specifically, the following solutions are recommended.

Systematic Assessment of Available Data Resources

An inventory of relevant databases would be of great value to scientists and research-funding agencies by identifying both immediate opportunities and critical gaps in data. An inventory should comprise, at a minimum, the following elements.

- Systematic description of national, state, or local public or private databases on health outcomes that might be used for studies of accountability. A catalog of data elements and their accessibility, strengths, and weaknesses would facilitate accountability assessments. Some existing databases could be explored to determine their usefulness. EPA's recent guide to the use of the NHANES database for environmental health analysis is one excellent example (EPA 2003).
- A similar list of databases of key, potential confounding factors in assessments of the health impact of air quality regulations. The list should include systems that track factors such as trends in diagnosis and treatment (eg, medication use), smoking behavior, and nutritional data at a population level over time. The Behavioral Risk Factor Surveillance System (BRFSS) (US Centers for Disease Control and Prevention [CDC] 2003a) is an example.

Assessment of Potential Biomarkers of Response

Biomarkers of response would be invaluable, particularly if they could be obtained as part of national or local health surveys. Assessment of such biomarkers would enable researchers to accurately gauge the feasibility of using available markers and the effort required to bring promising markers to the point at which they could be used for health-impact assessments. Unfortunately, few highly specific markers such as those for exposure to lead, environmental tobacco smoke, and carbon monoxide have been developed; biomarkers of response may not be possible for

most pollutants. Recent work on biomarkers of exposure for the air toxics butadiene and benzene offer some promise (Albertini et al 2003; Qu et al 2003), however.

DESIGN AND ANALYSIS

Explicit research on study designs for accountability research is needed, as discussed in Chapter 5. Both conceptual and methodologic issues need to be addressed, including the fundamental step of assuring a uniform view of the concept of accountability among researchers and regulators. Some needed activities include the following.

Focus Research Planning

- Workshops involving the many stakeholders concerned with accountability to move toward a shared understanding of the concept. This Communication would be a useful basis for beginning this needed discussion, which should lead to further elaboration of the concepts brought forward.
- Further elaboration of study designs to assess accountability, including assessing their sensitivity in detecting meaningful and anticipated changes in indicators along the chain of accountability.

Review Available Information

Syntheses of studies relevant to accountability assessment should be conducted in which different approaches are taken to combining data across studies to properly gauge the weight of the evidence. Such a comprehensive and detailed review of studies and settings in which information is already available can provide information relevant to accountability assessment and provide guidance as to additional information that is needed.

Natural experiments have provided needed and powerful information concerning the health risks of air pollution. The concept of the natural experiment should be more formally elaborated and the types of opportunities likely to be informative should be described. A mechanism for identifying possibly informative natural experiments should be set in place, as well as a process for following up on them when warranted. A formal analysis could assist in identifying the most important gaps in information, which, if filled, would exert the most leverage on both scientific knowledge and public policy decisions. This activity could usefully precede, and be incorporated into, the research planning process.

Analyze Model-Based Predictions

Model-based predictions could be analyzed to compare predicted and observed effects while accounting for model uncertainty. Formal causal modeling could also be informative. Findings from such analyses, when they appropriately address uncertainty, can provide insight into the information required to make more accurate accountability assessments, providing a form of leverage analysis.

Identify Opportunities to Develop Cooperative Research Models

Mechanisms for identifying in advance settings that may be affected by changes in air pollution regulations or policies should be implemented. Such mechanisms will likely require funding agencies to reach out to regulators and other governmental agencies and affected communities. It will also require development of research models that facilitate cooperation in anticipating, planning, and conducting accountability studies.

Implement Specific Study Designs

- **Serial cross-sectional studies.** Serial cross-sectional studies could conceivably be completed within relatively short timeframes. For example, cross-sectional survey data on many health outcomes of interest are available, some of which are routinely collected serially (eg, through NHANES). Repeating a comparable cross-sectional study after an air pollution intervention is one practical approach; the timing of the repeat study or studies would be determined by the postulated latency period between any change in exposure and the health outcome of interest. One example of serial cross-sectional studies is studies conducted before and after the reunification of Germany, which reported extensive changes in air pollution concentrations after reunification due to changes in power generation, transportation, and industrial production (Heinrich et al 2002).
- **Randomized studies.** These studies should be aimed at providing information on one or several links in the chain of accountability in a rigorous manner. Randomizing exposure in real-life settings through randomized manipulation of behavior (eg, provision of air conditioners or indoor versus outdoor exercise regimens) is an example of providing information on the link between exposure and health effect.
- **Cohort studies**. Ongoing cohort studies provide limited information for the windows of exposure applicable to the study participants (eg, HEI 2001; Pope et al 2002). Insights from cohort studies might be improved by combining evidence from multiple cohorts to widen the exposure windows that could be assessed.

GENERAL RECOMMENDATIONS

Ideally, any effort to control air pollution would be based on an understanding of the population burden of disease to be avoided and would be accompanied by a plan to evaluate the health impact along the chain of accountability. This Communication is a first step toward that goal. The Health Effects Institute is committed to developing methods for accountability assessment and to funding studies that address accountability.

CONTINUE FUNDED RESEARCH

The research proposed in this Communication and by other planning efforts is essential to making progress in collecting evidence on the basis of the chain of accountability. Even as planning proceeds, opportunities should be sought for accountability studies. As this Communication makes clear, we will learn what works and what does not from the conduct of research coupled with critical analysis of results. HEI and the California Air Resources Board issued Requests for Applications (RFAs) seeking studies of the health impact of air quality regulations (California Air Resources Board 2001; HEI 2002). Both agencies are now funding research proposed in response to those RFAs (information on the HEI studies available from *www.healtheffects.org*). Funding agencies should adopt a long-term perspective on supporting accountability research that incorporates periodic assessments of progress.

HEI continues to actively seek research through a preliminary application process (HEI 2003a), even as we consider how to best implement the recommendations of this Communication. We anticipate that the recommendations in this chapter will provide the basis for both research planning activities and additional RFAs in the near future.

DEVELOP SURVEILLANCE SYSTEMS

Carrying out some of these research recommendations would be greatly facilitated by availability of data collected through ongoing surveillance of the major time-varying links of the chain of accountability, or at least those components dealing with exposure and health outcomes. In fact, direct assessment of the health impact of air quality regulations on state-wide, regional, or national scales is probably impossible without such systems being implemented.

Existing data on air pollution and precedents for using national databases for air pollution analyses (eg, the National Morbidity, Mortality, and Air Pollution Study: Dominici et al 2003; Schwartz et al 2003), suggest that the issues of air pollution and public health provide an excellent opportunity to pilot the use of emerging surveillance systems for informing decisions about public health interventions. Some elements of such a system already exist, such as the US National Centers for Health Statistics, NHANES, and EPA's and state air monitoring networks. The US National Children's Study (The National Children's Study 2002) may provide additional resources in the future.

A system for long-term surveillance of the health impact of air quality regulations, however, will initially require evaluation of the adequacy of these existing resources in the context of a proposed design. This evaluation would also need to consider what kinds of information would be required for long-term evaluation of health impact, including goals for efficacy and effectiveness. Several recent and ongoing efforts have already made important contributions to such an effort.

- The Pew Environmental Health Commission examined a number of national health outcome databases for information on chronic diseases linked to environmental factors (Pew Charitable Trusts 2003). These databases (National Hospital Discharge Survey, National Ambulatory Medical Center Survey, and National Hospital Ambulatory Medical Care Survey) provide information on patient demographics and delivery of health services. They would need to be modified or augmented to be of use in air pollution accountability research, however, as they were not designed to describe the communities in which the illnesses occurred or the environmentally related health outcomes (Environmental Health Tracking Project Team 2000).
- A Nationwide Health Tracking Act is being considered by the US Congress to develop a comprehensive system for identifying and monitoring chronic diseases and correlating their causes with environmental, behavioral, socioeconomic, and demographic risk factors. Such a system would generate information that could be used as a basis for interventions to alleviate the sources of diseases affecting a particular community or the population as a whole. This system would have obvious potential for accountability research.
- The CDC, in their *Environmental Public Health Indicators Project* (2003b), and the Canadian Government, in their report on *National Consensus Conference on Population Health Indicators* (Canadian Institute for Health Information 1999), have both recently begun to formally evaluate a range of health indicators that could be used to track changes in health outcomes caused by environmental factors. Both efforts have specifically considered those outcome measures and listed data sources that are

related to outdoor air pollution. The CDC is supporting Centers of Excellence at several US academic institutions that are charged with developing these environmental health indicators further and supporting states and other localities in their application.

Planning a national surveillance approach for assessing the health impact of air quality regulations could begin with a series of workshops that bring together academic researchers and public health and air-regulatory officials at the state and federal levels. Participants could comprehensively review and discuss how existing resources could best be marshaled to estimate the health impact of air quality regulations.

COMBINE PUBLIC HEALTH ASSESSMENT AND COST-BENEFIT ANALYSIS

As this Communication has discussed, quantifying health benefits (eg, numbers of lives saved or hospitalizations prevented) is central to assessing regulatory interventions from a public health standpoint. As discussed in Chapter 2, however, policy decisions are frequently informed by assessments of how such interventions affect economic welfare. If such assessments are to be framed in terms of economic welfare, then an appropriate tool is cost-benefit analysis, and an appropriate measure of accountability is net benefits. In these terms, we would define a regulatory action as being accountable if it delivered positive net benefits and, further, we might rank a variety of regulatory actions to see which have delivered the greatest net benefits. Alternatively, cost-benefit analysis could be used in an ex ante fashion to examine which of a host of proposed regulatory alternatives deliver the greatest net benefits.

A monograph, complementary to this Communication, that explores the intersection of public health assessment and cost-benefit analysis seems warranted. Such an effort might best be undertaken by the economics community in cooperation with HEI and other health research organizations. The effort should bring together health scientists and economists to consider how studies of health effects can be designed to contribute more effectively to cost-benefit analysis, with particular focus on selection and measurement of health endpoints most useful for such work.

CONCLUSION

Air quality in the United States has improved considerably in recent decades, due in large part to air quality regulation. Yet evidence of continuing adverse health effects has prompted increasingly stringent air quality regulations. Demonstrating that these regulations are producing the desired health benefits will require creative and rigorous application of epidemiologic research methods and public health surveillance approaches within a conceptual framework for assessing accountability at each stage of the regulatory process. This Communication proposes such a framework and begins to identify opportunities to conduct accountability research. It also provides grounds for optimism that the efforts of HEI and others will ultimately provide the more definitive evidence on the health impact of air quality regulations sought by governments, scientists, and the public.

REFERENCES

Albertini RJ, Šrám RJ, Vacek PM, Lynch J, Nicklas JA, van Sittert NJ, Boogaard PJ, Henderson RF, Swenberg JA, Tates AD, Ward JB Jr, Wright M and others. 2003. Biomarkers in Czech Workers Exposed to 1,3-Butadiene: A Transitional Epidemiologic Study. Research Report 116. Health Effects Institute, Boston MA.

California Air Resources Board. 2001. Request for Proposals: Health Benefits of Incremental Improvements in Air Quality. CARB, Sacramento CA.

Canadian Institute for Health Information. 1999. National Consensus Conference on Population Health Indicators. Final Report. Canadian Institute for Health Information, Ottawa ON, Canada. Available from *http://secure.cihi.ca/cihiweb/en/downloads/infostand_ihisd_e_phi.pdf*.

Centers for Disease Control and Prevention (US). 2003a. Behavioral Risk Factor Surveillance System (last updated 3/14/03). *www.cdc.gov/brfss/*. Accessed 8/18/03.

Centers for Disease Control and Prevention (US). 2003b. Environmental public health indicators project (last updated 4/10/03). *www.cdc.gov/nceh/indicators/default.htm*. Accessed 8/18/03.

Clancy L, Goodman P, Sinclair H, Dockery DW. 2002. Effect of air-pollution control on death rates in Dublin, Ireland: An intervention study. Lancet 360:1210–1214.

Department of Health and Human Services (US). 1990. Smoking and Health: A National Status Report (2nd edition). Report to Congress. DHHS 87-8396 (revised 02/90). Government Printing Office, Washington DC.

Dominici F, McDermott A, Daniels M, Zeger SL, Samet JM. 2003. Mortality among residents of 90 cities. In: Revised

Analyses of Time-Series Studies of Air Pollution and Health, pp 9–24. Special Report. Health Effects Institute, Boston MA.

Environmental Health Tracking Project Team. 2000. America's Environmental Health Gap: Why the Country Needs a Nationwide Health Tracking Network. Technical Report. Johns Hopkins School of Hygiene and Public Health, Baltimore MD. Available from *http://pewenvirohealth.jhsph.edu/html/reports/pewtrackingtechnical.pdf*.

Environmental Protection Agency (US). 1997. National Ambient Air Quality Standards for Particulate Matter; Final Rule. 40 CFR, Part 50. Fed Regist 62:38651–38760.

Environmental Protection Agency (US). 1998. National Air Pollutant Emission Trends. Procedures Document 1900–1996. EPA-454/R-98-008. Office of Air Quality Planning and Standards, Research Triangle Park NC.

Environmental Protection Agency (US). 2001. Control of Air Pollution from New Motor Vehicles: Heavy-Duty Engine and Vehicle Standards and Highway Diesel Fuel Sulfur Control Requirements; Final Rule. 40 CFR, Parts 69, 80, and 86. Fed Regist 66:5001–5050.

Environmental Protection Agency (US). 2002. Monitor network and IMPROVE Protocol Network maps. Office of Air and Radiation, Washington DC. Available from *www.epa.gov/ttn/amtic/pmspec.html*.

Environmental Protection Agency (US). 2003. EPA Handbook for Use of Data from the National Health and Nutrition Examination Surveys (NHANES): A Goldmine of Data for Environmental Health Analyses. EPA/600/R-02/044. National Center for Environmental Assessment, Office of Research and Development, Washington DC.

Health Effects Institute. 2001. Airborne Particles and Health: HEI Epidemiologic Evidence. HEI Perspectives. Health Effects Institute, Cambridge MA.

Health Effects Institute. 2002. Request for Applications: Winter 2002 Research Agenda. Health Effects Institute, Boston MA.

Health Effects Institute. 2003a. Request for Applications: Winter 2003 Research Agenda. Health Effects Institute, Boston MA.

Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. 2002. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: An intervention study. Lancet 360:1646–1652.

Heinrich J, Hoelscher B, Frye C, Meyer I, Pitz M, Cyrys J, Wjst M, Neas L, Wichmann H-E. 2002. Improved air quality in reunified Germany and decreases in respiratory symptoms. Epidemiology 13:394–401.

Mayor of London. 2002. Cleaning London's Air: The Mayor's Air Quality Strategy. Greater London Authority, London, United Kingdom. Available from *www.london.gov.uk/approot/mayor/strategies/ air_quality/index.jsp*.

Office of the Governor (New York State). 1997. Governor Pataki announces clean fuel bus plan for NYC [press release]. Office of the Governor, Albany NY. Available from *www.state.ny.us/governor/press/jan29_3.html*.

Pew Charitable Trusts. 2003. Public Health Initiative: The Health and Human Services Program at the Pew Charitable Trusts. Pew Charitable Trusts, Philadelphia PA. Available from *www.pewtrusts.com/pdf/hhs_public_health_overview.pdf*.

Pope CA III. 1989. Respiratory disease associated with community air pollution and a steel mill, Utah Valley. Am J Public Health 79:623–628.

Pope CA III, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287:1132–1141.

Qu Q, Shore R, Li G, Jin X, Chen L-C, Cohen B, Melikian AA, Eastmond D, Rappaport S, Li H, Rupa D, Waidyanatha S, Yin S, Yan H, Meng M, Winnik W, Kwok ESC, Li Y, Mu R, Xu B, Zhang X, Li K. 2003. Validation and Evaluation of Biomarkers in Workers Exposed to Benzene in China. Research Report 115. Health Effects Institute, Boston MA.

Schwartz J, Zanobetti A, Bateson T. 2003. Morbidity and mortality among elderly residents of cities with daily PM measurements. In: Revised Analyses of Time-Series Studies of Air Pollution and Health, pp 25–58. Special Report. Health Effects Institute, Boston MA.

The National Children's Study. 2002. April 2002 study assembly meeting (last updated 4/16/02). *http://nationalchildrensstudy.gov/events/study_assembly/SAM_040702.cfm*. Accessed 8/18/03.

ABBREVIATIONS AND OTHER TERMS

CDC	Centers for Disease Control and Prevention (US)
EPA	Environmental Protection Agency (US)
NHANES	National Health and Nutrition Examination Survey (US)
$PM_{2.5}$	particulate matter less than 2.5 µm in aerodynamic diameter
RFA	Request for Applications
SIP	state implementation plan

HEALTH
EFFECTS
INSTITUTE

Charlestown Navy Yard
120 Second Avenue
Boston MA 02129-4533
Phone +1-617-886-9330
Fax +1-617-886-9335
www.healtheffects.org

COMMUNICATION 11
September 2003