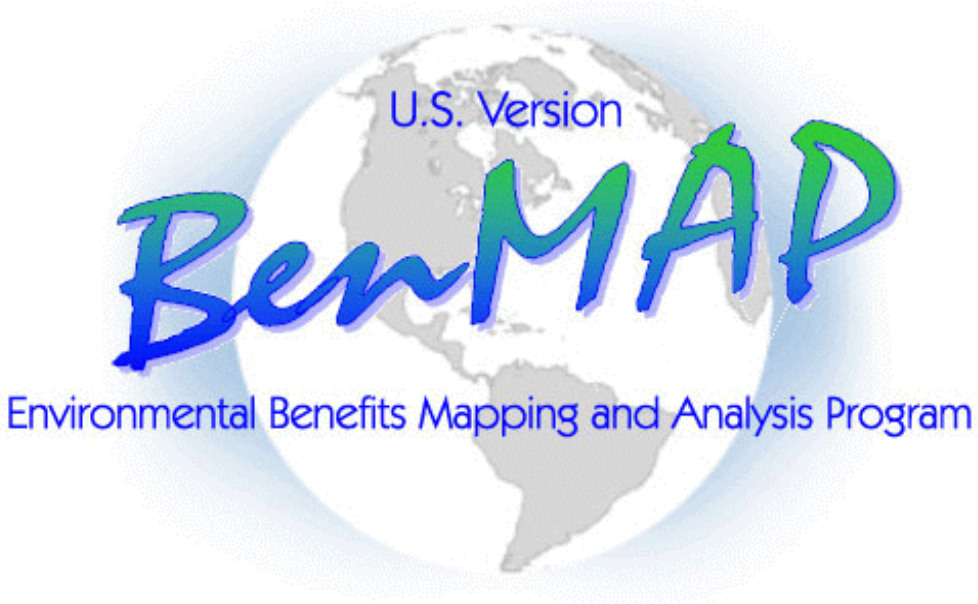
Abt Associates Inc. ■ 4800 Montgomery Lane ■
Bethesda, MD 20814 ■ www.abtassociates.com



User's Manual

November 2003

Prepared for
Office of Air Quality Planning and Standards
U.S. Environmental Protection Agency
Research Triangle Park, NC
Bryan Hubbell, Project Manager

Prepared by
Abt Associates Inc.

Welcome to BenMAP, the Environmental Benefits Mapping and Analysis Program

BenMAP is a new tool that allows you to perform customized health benefits analyses for changes in air quality, in a powerful, yet easy- to-use program. BenMAP allows you to select and customize the data inputs, select from a range of modeling options, and report the results in a variety of ways. In addition, BenMAP provides a range of mapping options. The model estimates the reduction in the incidence of adverse health effects, as well as the estimated economic value of such a reduction, and it also reports air quality and population exposure results. BenMAP can be used for a variety of purposes:

- ➢ Generation of population/community level ambient pollution exposure maps;
- ➢ Comparing benefits associated with regulatory programs;
- ➢ Estimating health impacts and costs of existing air pollution concentrations;
- ➢ Estimating health benefits of alternative ambient air quality standards;
- ➢ Performing sensitivity analyses of health or valuation functions, or of other inputs; and
- ➢ Screening analyses.

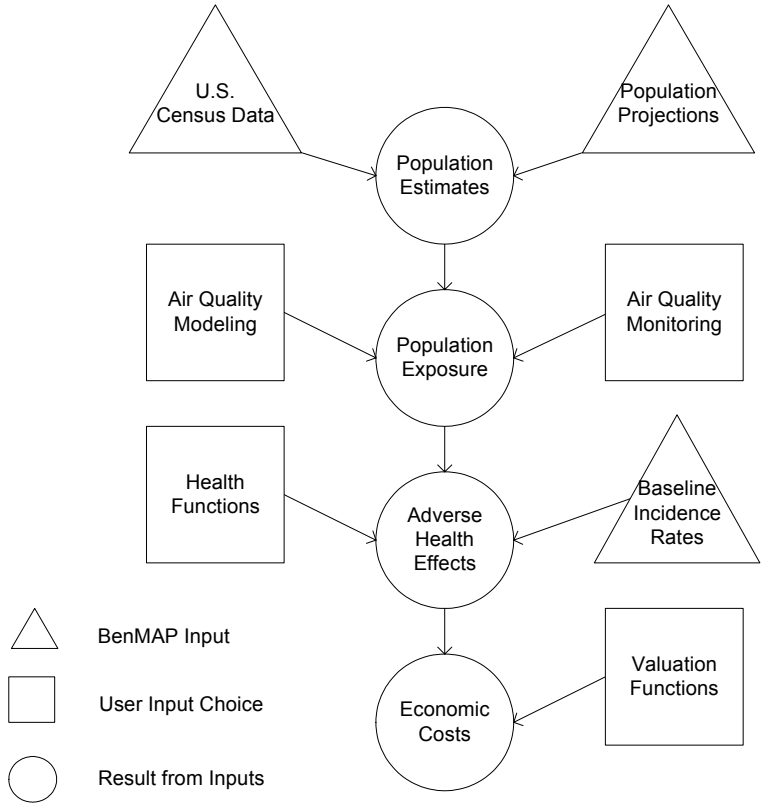
To calculate people's exposure to air pollution, BenMAP combines different sources of data, including air pollution monitoring data, modeling data, census data, and county-level population projections. And to quickly generate additional exposure scenarios, BenMAP also provides several options to directly reduce or "roll back" monitored pollution levels. Using these data and rollback capabilities, BenMAP can estimate population exposure for any particular year, and for the specified set of air pollution assumptions. Typically, you will specify two different scenarios, and BenMAP will estimate the change in population exposure between them. Using this change in population exposure, BenMAP then calculates the associated change in the incidence of adverse health effects, and its estimated economic value. For these calculations, BenMAP allows you the option of choosing from a large set of EPA Standard functions to estimate adverse health effects and to value these health effects. In addition, BenMAP gives you the flexibility to add your own health effect and valuation functions.

While BenMAP is not an air quality modeling program, it does contain a powerful set of functions to explore the impact of reducing monitored concentrations of air pollution. These functions, available under the monitor rollback option under the "Create Air Quality Grids" button, allow the user to specify specific percent or incremental reductions, or to examine a wide variety of forms of ambient air quality standards, including annual or daily standards, and incorporating such factors as ordinality, anthropogenic background, and metric form, e.g. daily average or maximum daily 8-hour average. These functions can be used without any additional input data from the user, by accessing the air quality monitoring data provided with the BenMAP program.

BenMAP also provides powerful tools for quality assurance and results presentation. These tools include a set of mapping and data query tools, and an "audit trail" feature which clearly documents all inputs, assumptions, and modeling choices used to generate a specific set of population exposure, incidence, and valuation results. This tool allows for easy QA checks, as well as comparison with other BenMAP results for purposes of replication and validation.

Exhibit 0-1 summarizes the flow of calculations in BenMAP, and the types of choices that you make regarding the modeling of population exposure, the types of health effects to model, and how to place an economic value on these health effects. This exhibit also highlights that BenMAP does not have air modeling capabilities, and instead relies on modeling and monitoring inputs. Note that the current version of BenMAP requires the use of the population data and population projections that come with the model. See Appendix B for details on the population data included in the model.

Exhibit 0-1. BenMAP Flow Diagram



Who Can Use BenMAP?

BenMAP can be used by a wide range of persons, including scientists, policy analysts, and decisionmakers. Advanced users can explore a wide range of advanced options, such as filtering of monitor data, using the map querying features, and exploring the impacts of different health and valuation functions. Less experienced users can simply apply the EPA Standard functions and reproduce the types of analyses performed in regulatory impact analyses.

Power users will also find a number of convenient features in BenMAP, including the availability of a command line version of the model that allows the user to run BenMAP in "batch mode,"

which can save a great deal of time when a large number of scenarios need to be analyzed. The command line version of BenMAP is available by contacting Bryan Hubbell at the U.S. EPA (hubbell.bryan@epa.gov).

How to Use this Manual

This manual has two main parts. Chapters 1 through 10 provide step-by-step instructions on how to use BenMAP. Appendices A through I present details on the databases and methods built into the model.

New users should start with Chapters 1 and 2, which are both very short, but provide a good overview of the model and how it works, and explain some potentially confusing terminology. You can then use Chapter 3 to get started using the model. The tutorial in Chapter 3 will show you how to define two simple scenarios, calculate the change in health effects between them, assign economic valuations, and create reports and maps. Once you have gone through this simple tutorial, you can go on to try more advanced functions. Use the rest of the manual to answer any specific questions you may have, or to walk you step-by-step through the various model functions. Chapters 4 through 7 cover each of the four main buttons, Chapter 8 covers mapping, and Chapters 9 and 10 explain the Data and Tools menus.

Each of the first ten chapters is introduced by a short section which describes what you can find within the chapter, and provides an outline of the chapter's contents. This is a good place to go if the Table of Contents does not provide enough detail for you to find the section you need. The end of most chapters has a series of "Frequently Asked Questions," which may also be helpful in answering specific questions. In sections that provide instructions on navigating the model, the following conventions are observed: menu items, buttons, and tab and selection box labels are in bold type; prompts and messages are enclosed in quotation marks; and drop-down menu items, options to click or check, and items that need to be filled in or selected by the user are italicized. Throughout the first ten chapters you will also see boxes that say ➥ TIP. These boxes present common mistakes and important things to remember when working with BenMAP. Common terms are defined in Chapter 1, but you can also find definitions in boxes along the right margins inside other chapters.

Computer Requirements

BenMAP requires a computer with:

➢Windows 2000 or greater.
➢1 GB RAM or greater (or 512 MB RAM if no 12 x12 air quality grids, e.g. CAMx or REMSAD need to be created). In general, more RAM is preferred, although BenMAP can be run with the 512 minimum if certain choices within the model are constrained.
➢1 GHZ processor or greater recommended.

BenMAP does not model air quality, so to estimate population exposure to air pollution you need to supply modeling data, monitoring data, or both. BenMAP comes supplied with a number of

years of monitor data, as well as some air quality modeling files that enable you to get started right away. Chapter 3 provides an example for you to understand how these might be used.

Installing BenMAP

➢Install **Interbase** by running ibinstall.exe.
- At the "Select Installation Type" prompt, select *Interbase Client/Server*.
- When prompted, restart the machine.

➢Run Setup.exe. This will install BenMAP 2003 Beta 2.0.

Uninstalling BenMAP

➢Go to Control Panel / Add/Remove Programs.
➢Select BenMAP 2003 Beta 2.0, and click Change/Remove.
➢When prompted, select *Yes* to completely remove BenMAP 2003 Beta 2.0 and all its components.
➢Select *Interbase 6.0*, and click **Change/Remove**.
➢Select *Automatic* at the "Select Uninstall Method" prompt and click **Next**.
➢Click **Finish**.

Contacts for Comments and Questions

For comments and questions, please contact Bryan Hubbell at the U.S. Environmental Protection Agency.
Address: C339-01, USEPA Mailroom, Research Triangle Park, NC 27711
Email: hubbell.bryan@epa.gov
Telephone: 919-541-0621.

Sources for More Information

For files that you can use in BenMAP:

➢U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards (OAQPS), BenMAP website. Available at:
http://www.epa.gov/ttn/ecas/analguid.html

For more information on conducting benefit analysis, see the following documents:

➢U.S. Environmental Protection Agency, National Center for Environmental Economics (NCEE), Guidelines for Preparing Economic Analyses. Available at:
http://yosemite.epa.gov/ee/epa/eed.nsf/pages/guidelines

➢U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards (OAQPS), Economic Analysis Resource Document. Available at: http://www.epa.gov/ttn/ecas/analguid.pdf

➢U.S. Environmental Protection Agency, Office of Air and Radiation, Costs and Benefits of the Clean Air Act. Available at: http://www.epa.gov/oar/sect812/

➢U.S. Environmental Protection Agency, Office of Air and Radiation, Draft Regulatory Impact Analysis: Control of Emissions from Nonroad Diesel Engines, Chapter 9. EPA420-R-03-008, April 2003. Available at: http://www.epa.gov/nonroad/r03008.pdf

Table of Contents

Exhibit List

1. Terminology and File Types

The first section of this chapter explains common terms used in this user's manual and in the model, and references, where possible, other sections in this manual to find more detailed information. Section 1.2 describes in detail the necessary format for externally-generated model and monitor data files that can be read into BenMAP and used to generate the air quality grid files.

1.1 Common Terms

➢**Aggregation.** The summing of grid cell level results to the county, state and national levels.

➢**APV Configuration (Aggregation, Pooling, and Valuation).** APV Configurations store the aggregation levels, pooling options, and valuation methods used in the analysis. In particular, you may specify the aggregation level for incidence estimates, how to pool incidence estimates, the valuation estimates to use, the aggregation level for the valuation estimates, and how to pool the valuation estimates. APV Configurations are stored in files with an ".apv" file extension. The results derived from an APV Configuration have an ".apvr" file extension. APV files are typically stored in the *Configurations* folder, and APV Results files are typically stored in the *Configuration Results* folder.

➢**Air Quality Grid.** An air quality grid contains air pollution data. BenMAP uses one air quality grid for the baseline scenario and a second for the control scenario, in order to estimate the change in the number of adverse health effects between the two scenarios. Air Quality Grids are stored in files with an ".aqg" file extension. AQG files are typically stored in the *Air Quality Grids* folder.

➢**Air Quality Metric.** One of the measures typically used for air pollution. Particulate matter is typically measured on a daily basis by air quality monitors, and has three metrics: daily average, annual average, and the median daily average over the course of a year. Since ozone is measured hourly by monitors, the metrics include: highest hour over the course of a day, as well as a series of averages for specified parts of the day: five-hour (10am - 3pm), eight-hour (9am - 5pm), 12-hour (8am - 8pm), and 24-hour.

➢**Binning.** The process of summarizing data by sorting the data values from low to high, dividing the sorted data into a pre-determined number of groups, and then choosing a representative value for each group, by either averaging the values in each group, or picking the mid-point of each group.

➢**CAMx (Comprehensive Air Quality Model with Extensions).** An air quality model used to measure ozone levels. It has grid cells with dimensions of approximately 12 kilometers by 12 kilometers.

➢**CMAQ (Community Multi-Scale Air Quality).** A state-of-the-art air quality model able to model ambient particulate levels, as well as other pollutants, including ozone. The grid-size of CMAQ is approximately 36 kilometers by 36 kilometers.

➢**Configuration.** The Configuration stores the C-R functions and model options used to estimate adverse health effects. Configurations are stored in files with a ".cfg" file extension. CFG files are typically stored in the *Configurations* folder. The results derived from a Configuration have a ".cfgr" file extension. CFGR files are typically stored in the *Configuration Results* folder.

➢**C-R (Concentration-Response) Function.** A C-R function calculates the change in adverse health effects associated with a change in exposure to air population. A typical C-R function has inputs specifying the air quality metric and pollutant, the age, race and ethnicity of the population affected, and the incidence rate of the adverse health effect.

➢**Cost of Illness (COI).** The cost of illness includes the direct medical costs and lost earnings associated with illness. These estimates generally understate the true value of reductions in risk of a health effect, as they include just the direct expenditures related to treatment and lost earnings, but not the value of avoided pain and suffering from the health effect.

➢**Endpoint.** An endpoint is a subset of an endpoint group, and represents a more specific class of adverse health effects. For example, within the endpoint group *Mortality*, there are the endpoints *Mortality, Long Term, All Cause* and *Mortality, Long Term, Cardiopulmonary*. In cases where an endpoint group has only a single endpoint, they share the same name.

➢**Endpoint Group.** A endpoint group represents a broad class of adverse health effects, such as premature mortality, chronic bronchitis, and hospital admissions. BenMAP only allows pooling of adverse health effects to occur within a given endpoint group, as it generally does not make sense to sum together the number of cases of disparate health effects, such as premature mortality and chronic bronchitis.

➢**Fixed Effects Pooling.** Fixed effects pooling is used to combine two or more distributions (represented by Latin Hypercube points) into a single new distribution. Fixed effects pooling assumes that there is a single true underlying relationship between these component distributions, and that differences among estimated parameters are the result of sampling error. Weights for the pooling are generated via inverse variance weighting, thus giving more weight to the input distributions with lower variance and less weight to the input distributions with higher variance.

➢**Grid cell.** One of the many geographic, or spatial components within an air quality model, or grid. For example, the REMSAD model is composed of grid cells that are approximately 36 kilometers by 36 kilometers.

➢ **Incidence Rate.** The incidence rate is the average number of adverse health effects per person per unit of time, typically a day or a year. Appendix E discusses the data sources for the incidence rates in BenMAP. Note, to avoid potentially small numbers and to ease comparison of different rates, Appendix E reports all of the incidence rates as the number of cases per 100 individuals per year.

➢ **Interpolation.** The process of estimating the air quality level in an unmonitored area by using one or more nearby air quality monitors. BenMAP uses three types of interpolation procedures: one is to simply choose the closest monitor, another is to use a technique called Voronoi Neighbor Averaging, and the third uses a statistical technique called Kriging. These interpolation methods are discussed in more detail in Appendix C.

➢ **Kriging.** A particular interpolation method for spatial data. With kriging, BenMAP calculates a weighted average of the data of neighboring monitors within a user defined distance based on the covariance structure derived from these neighboring monitors.

➢**Latin Hypercube.** A series of points generated by using specified percentiles in a given distribution, such as that of a C-R coefficient. It is a short-cut method designed to represent a distribution, while at the same time saving on computation time. For example, when using 20 Latin Hypercube points, BenMAP would use the 2.5th, 7.5th, 12.5th, ..., and 97.5th points from the distribution. The Latin Hypercube points are used when combining the results of different C-R functions (discussed in Chapter 6), and in presenting confidence intervals for the incidence estimates (discussed in Chapter 7).

➢**Modeling.** Estimating air pollution levels through the use of air quality models. The EPA website discusses a wide range of air quality models: http://www.epa.gov/ebtpages/airairquaairqualitymodels.html and at http://www.epa.gov/ttn/scram/.

➢**Monitoring.** Actual measurements of air pollution levels. Appendix A discusses the air quality monitoring data used in BenMAP. The U.S. Environmental Protection Agency has monitoring data, as well as other information related to monitoring, available through its Air Quality System (AQS): http://www.epa.gov/air/data/aqsdb.html.

➢**Particulate Matter.** Includes $PM_{2.5}$ (particles less than 2.5 microns in aerodynamic diameter), PM_{10} (particles less than 10 microns in aerodynamic diameter), and PMC (particles between 2.5 and 10 microns in aerodynamic diameter).

➢**Pooling.** The combining of different sets of data. BenMAP has several pooling methods, including fixed effects, fixed/random effects, and subjective weighting. Appendix I discusses the pooling approaches available in BenMAP.

➢**Point Mode.** When defining the configuration, you may choose to either estimate adverse health effects in point mode or using a Latin Hypercube. The point mode simply means that BenMAP will use the mean value of the coefficient in the C-R function.

➢**Population Exposure versus Personal Exposure.** Population (or ambient) exposure refers to the average air pollution level measured in a grid cell. In contrast, personal exposure keeps track over the course of a day the exposure individuals encounter in different micro-environments, such as the freeway, outdoors and indoors. BenMAP only keeps track of population exposure.

➢**Prevalence.** The prevalence specifies the percentage of individuals with a given adverse health effect.

➢**Random Effects Pooling.** Random effects pooling is an alternative to the fixed effects model (see Fixed Effects Pooling, above), and allows the possibility that the estimated parameter from different studies may in fact be estimates of *different* parameters, rather than just different estimates of a single underlying parameter.

➢**Relative Risk.** Relative risk typically is used as a measure of the change in risk of an adverse health effect associated with an increase in air pollution levels. More specifically, it is the ratio of the risk of illness with higher pollution to the risk of illness with a lower pollution level, where the "risk" is defined as the probability that an individual will become ill.

➢**REMSAD (Regulatory Model System for Aerosols and Deposition).** An air quality model able to calculate particulate matter levels, as well as other pollutants, including ozone. It has two types of grid cells, one with dimensions of approximately 12 kilometers by 12 kilometers (REMSAD12), and the second with dimensions of approximately 36 kilometers by 36 kilometers (REMSAD36).

➢**Subjective Weighting.** Subjective weights let you specify the weights that you want to use when combining two or more distributions of results. The weights should sum to one. If not, BenMAP normalizes the weights so that they do.

➢**UAM-V (Urban Airshed Monitoring - Variable grid).** An air quality model typically used to measure ozone levels. It has grid cells with dimensions of approximately 12 kilometers by 12 kilometers.

➢**VNA (Voronoi Neighbor Averaging).** An algorithm used by BenMAP to interpolate air quality monitoring data to an unmonitored location. BenMAP first identifies the set of monitors that best "surround" the center of the population grid cell, and then takes an inverse-distance weighted average of the monitoring values. This is discussed in detail in Appendix C.

➢**WTP (Willingness to Pay).** The willingness of individuals to pay for a good, such as a reduction in the risk of illness. In general, economists tend to view an individual's WTP for a improvement in environmental quality as the appropriate measure of the value of a risk reduction. An individual's willingness-to-accept (WTA) compensation for not receiving an improvement is also a valid measure. However, WTP is generally considered to be a more readily available and conservative measure of benefits.

1.2 File Types

To calculate population exposure to air pollution, BenMAP depends on externally-generated model and monitor data files. The model files can come from four models: REMSAD, UAM-V, CAMx, and CMAQ. Chapter 4 provides additional details on how these model files are used.

You may use $PM_{2.5}$, PM_{10}, and ozone monitor data from a library in BenMAP, or use your own monitor data. This is discussed in greater detail in Chapter 4, and Appendix A provides additional details on the data in the monitor library installed with BenMAP.

To provide additional flexibility, BenMAP has a number of file types that you can use to store the settings used in a BenMAP run, the results of a run, as well as maps. Exhibit 1-1 presents the names of the different file types, their functions, and their default folder locations.

Exhibit 1-1. File Types Generated by BenMAP

File Extension	Description	Default Folder Location
*.aqg	Air quality grid.	Air Quality Grids
*.adj	Adjustment factor file generated from air quality modeling data. The adjustment factor file can be used in the creation of air quality grids by adjusting air quality monitoring data.	Model Data
*.cfg	Configuration specifying the C-R functions used to generate incidence estimates.	Configurations
*.cfgr	Configuration results, containing incidence results at the grid cell level.	Configuration Results
*.apv	Aggregation, Pooling, and Valuation configuration specifying the aggregation levels, pooling options, and valuation methods used to generate aggregated incidence estimates, pooled incidence estimates, valuation estimates, aggregated valuation estimates, and pooled valuation estimates.	Configurations
*.apvr	Aggregation, Pooling, and Valuation configuration results, containing incidence results at the grid cell level, aggregated incidence results, valuation results, aggregated valuation results, and pooled valuation results.	Configuration Results
*.shp	Shape files generated by BenMAP's mapping capabilities. These files can be viewed within BenMAP or within shape file viewers, such as ArcView.	Maps
*.csv	Reports are exported as *.csv files, which may be viewed in a text editor, or easily viewed in programs such as Excel.	Reports

CHAPTER 2

Overview of BenMAP Components

In this chapter...

➢ Get an overview of the functions available with each of BenMAP's four main buttons and two menu items.

➢ Learn about the different options for each function.

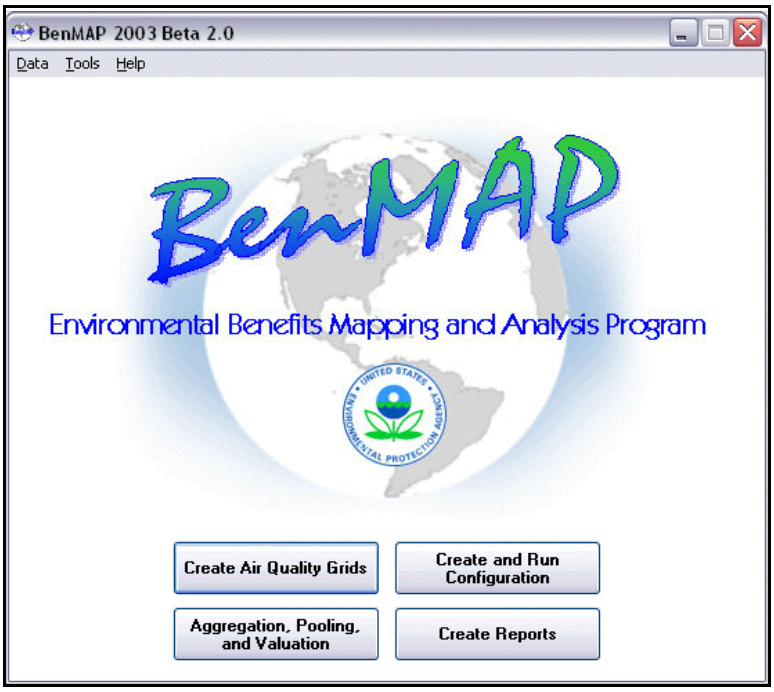
Chapter Overview

2. Overview of BenMAP Components

You can access most of BenMAP's functions by using the four large buttons on the opening screen. These buttons allow you to perform all of the actions needed to estimate the health benefits of a change in air quality. The two drop down menus at the top of the screen, **Data** and **Tools**, are for less frequently used functions, such as editing concentration-response (C-R) functions and mapping. Section 2.1 describes the main buttons and their functions, and Section 2.2 describes the Data and Tools menus. All of these topics are covered in greater detail in subsequent chapters of this manual.

2.1 Main Buttons

The four buttons take you through the steps of an analysis. The first button allows you to create air quality grids which contain estimates of population-level exposure to air pollution. The second button lets you choose the air quality grids for a particular analysis, and then to choose the C-R functions to estimate the incidence of adverse health effects. The third button gives you different options for combining the health effects estimates and placing an economic value on them. Using the fourth button, you can generate several different kinds of reports.



2.1.1 Create Air Quality Grids

BenMAP is not an air quality model, nor can it generate air quality data independently. Instead it relies on the air quality inputs given to it. To estimate population exposure to air pollution, BenMAP combines population data with an Air Quality Grid, which it generates using some combination of air quality modeling and/or monitoring data. The **Create Air Quality Grids** button allows you to put your air quality data into the format that is used by BenMAP. You must complete this step before you can perform an analysis in BenMAP.

Grid types

Air quality grids contain air pollution exposure estimates for a patchwork or "grid" that typically covers the continental United States. The grids used by BenMAP are not arbitrarily chosen, but instead either exactly match the grid or cell pattern used in common air quality models (REMSAD, CMAQ, CAMx, and UAM-V), or match political units, such as Counties.

Population data

BenMAP uses specially designed population files matched to each air quality grid specification, and then it estimates the air pollution exposure for this population using the air quality data (modeled or monitored) which you provide. Population files are provided for a limited number of air quality grid types. In the current version of the model, you must use the population data provided by BenMAP.

Modeling and Monitoring Data

To generate air quality grids, you can use air quality modeling data and air quality monitoring data in three different ways, as discussed below. However, once generated, the air quality grids all have the same structure, and have the same ".aqg" extension that BenMAP uses to designate these file types. The grids contain estimates of air pollution levels for a specific period of time, depending on the pollutant. For ozone, BenMAP uses a five-month summer ozone season (May 1 through September 30), and for particulate matter, BenMAP uses a year-long period.

➢**Model Direct.** The *Model Direct* grid creation option simply takes raw model data and converts it into a file that BenMAP recognizes as an air quality grid. With this approach, you specify a model filename, a pollutant, and a grid type (model type). Currently BenMAP can generate REMSAD (36km or 12km) and CMAQ grids for $PM_{2.5}$, PM_{10}, and PMC, and CAMx and UAM-V grids for ozone.

➢**Monitor Direct.** *Monitor Direct* grid creation uses ozone, $PM_{2.5}$, PMC, or PM_{10} air pollution monitoring data to estimate air pollution levels in the grid type you specify – either REMSAD, UAM-V / CAMx, CMAQ , or County. This may be done using one of the interpolation procedures – closest monitor, Voronoi Neighbor Averaging (VNA), or kriging. With closest monitor, BenMAP simply uses the data of the monitor closest to the grid cell's centroid. With VNA, BenMAP first identifies the set of monitors that "surround" each grid cell, and then calculates an inverse-distance weighted average of these neighboring monitors. With Kriging, BenMAP identifies all monitors within a user defined distance around the grid cell. The data from these monitors is then used to calculate a weighted average for the grid cell. The weights

are dependent on the covariance structure of the interpolated monitors. A functional form of the covariance structure must be supplied by the user.

➢**Monitor and Model Relative**. This approach allows you to combine information from both monitor and modeling data files. The basic idea is that BenMAP uses modeling data to scale monitoring data, to compensate for incomplete monitoring data coverage. Like with the *Monitor Direct* approach, you choose an interpolation method (closest monitor, VNA, or kriging), but in addition, you then choose a scaling approach, either temporal, spatial, or both. This approach is described in more detail in the next chapter, as well as in Appendix C.

➢**Monitor Rollback**. This approach allows you to reduce, or rollback, monitor data using three methods: percentage rollback, incremental rollback, and rollback to a standard. Percentage rollback reduces all monitor observations by the same percentage. Incremental rollback reduces all observations by the same increment. Rollback to a standard reduces monitor observations so that they just meet a specified standard. This approach is described in more detail in Chapter 4, as well as in Appendix A.

2.1.2 Create and Run Configuration

Using the two air quality grids as inputs, you can generate the change in adverse health effects associated with the change in air quality between them. There are several steps in this process.

➢ **Step 1.** Specify the baseline and control air quality grids that you created using the **Create Air Quality Grids** button.

➢ **Step 2.** Specify whether BenMAP should make a "point" estimate, or a set of "Latin Hypercube" points. The point estimate is the change in incidence, generated using the average coefficient value in the C-R function. The Latin Hypercube points are a series of points generated by using specified percentiles in the distribution of the C-R coefficient – these points represent the distribution of incidence values. (The Latin Hypercube points are later used when combining the results of different C-R functions, and in presenting confidence intervals for the incidence estimates.)

➢ **Step 3.** Specify the threshold, or a lowest value for air quality data. Any observations which fall below this threshold will be replaced with the threshold value in all calculations.

➢ **Step 4.** Choose the C-R functions that will be used in the estimation, and hit the **Go!** button to start estimating the change in incidence.

BenMAP can store configuration choices in a user-named file with a ".cfg" extension, and can store incidence estimates in a user-named file with a ".cfgr" extension. As needed, you can access both files for later use.

2.1.3 Specify Aggregation, Pooling, and Valuation

With this button, you can specify an aggregation level for previously calculated incidence estimates, pool these aggregated incidence estimates, place an economic value on these pooled and aggregated incidence estimates, aggregate these economic values, and finally pool these aggregated economic values. There are several steps in this process.

➢ **Step 1.** Choose a set of incidence estimates with which to work. This is a configuration results file made with the **Create and Run Configuration** button, which is stored by BenMAP with a ".cfgr" extension.

➢ **Step 2.** Choose the desired pooling and aggregation options for the incidence results.

➢ **Step 3.** Choose the economic valuation options to apply to the pooled and aggregated incidence results.

➢ **Step 4.** Choose the desired pooling and aggregation options for the economic valuations, and hit the **Go!** button to start generating results.

BenMAP can store APV Configuration choices in a user-named file with an ".apv" extension, and can store APV Configuration results in a user-named file with an ".apvr" extension. As needed, you can access both files for later use.

2.1.4 Generate Reports

BenMAP generates several types of reports - you can access these by clicking on the **Create Reports** button.

➢ *Incidence and Valuation Results* use an **Aggregation, Pooling, and Valuation Results** file (with the ".apvr" extension) to create reports for incidence, aggregated incidence, pooled incidence, valuation, aggregated valuation, or pooled valuation results. These reports are comma separated values (CSV) files (*.csv) which can be read into various spreadsheet and database programs, such as Microsoft Excel.

➢ *Raw Incidence Results* use a **Configuration Results** file (with the ".cfgr" extension) to create reports for incidence results. These reports are CSV files.

➢ *Audit Trail Reports* provides a summary of the assumptions underlying each of five types of files generated by BenMAP: **Air Quality Grid** (".aqg"), **Configuration** (".cfg"), **Configuration Results** (".cfgr"), **Aggregation, Pooling, and Valuation** (".apv"), and **Aggregation, Pooling, and Valuation Results** (".apvr"). These reports can be viewed within BenMAP in an expandable tree structure, or can be exported to tab-delimited text files.

2.2 Menus

There are three menu options, found at the top of the initial screen: **Data, Tools**, and **Help**. The **Data** and **Tools** menus provide access to tasks that are occasionally needed to perform a standard analysis, or to better understand the results of an analysis. The **Help** menu provides access to information about BenMAP - the version, contact information, and a suggested citation.

2.2.1 Data

The **Data** menu allows you to view the EPA Standard C-R and valuation functions available for use in BenMAP, as well as to add your own C-R and valuation functions. BenMAP does not allow you to edit or add to the set of EPA Standard functions. However, you can copy these functions, modify them, save them, and use them in your analyses. Additionally, you can create new functions from scratch.

2.2.2 Tools

The **Tools** menu allows you to choose: *Mapping / GIS, Adjustment Factor Creator, Neighbor File Generator, CAMx / UAM-V Model File Generator*, and *Shapefile Creator*. These are specialized tools that are not always needed for an analysis.

➢The *Mapping / GIS* tool allows the user to generate a wide variety of maps, including maps of monitor data, adjustment factors, air quality grids, population data, county incidence rates, and both incidence and valuation results. In addition, you can export the maps you have generated and view them in a shapefile viewer, such as Arcview. Note that BenMAP also has context specific mapping capabilities in various parts of the application. Chapter 8 provides additional details on BenMAP's mapping capabilities.

➢*Adjustment Factor Creator*. With this option you can generate the adjustment factor files needed for the *Monitor and Model Relative* air quality grid creation. To use the *Adjustment Factor Creator,* you simply specify a model file, along with the grid type and pollutant. Note that you can access the *Adjustment Factor Creator* from the *Monitor and Model Relative Settings* screen directly (see Section 4.3 for details). BenMAP also displays the *Adjustment Factor Creator* as a separate tool in case you only want to generate adjustment factor files, without generating air quality grids, or conducting other parts of a typical analysis.

➢*Neighbor File Creator*. You can load an air quality grid (a *Monitor Direct* grid or a *Monitor and Model Relative* grid), and generate a tab-delimited file containing information about the "neighbors" (or contributing monitors) of each grid cell and their associated weights. See the discussion on interpolation in Section 4.2 for more details on how BenMAP generates these air quality grids and determines the neighbors of each grid cell.

➢ *CAMx / UAM-V Model File Creator*. CAMx and UAM-V data comes in a series of individual files (one per day) for the Eastern United States and a separate series for the Western states. The *CAMx / UAM-V Model File Creator* allows you to select all of the Eastern domain and Western domain files, and produce a single model file that can be used in other parts of BenMAP requiring model file input, such as the *Adjustment Factor Creator*.

CHAPTER 3

BenMAP Quick Start Tutorial

In this chapter...

➢ Work through the steps of a simple health benefits analysis, from creating air quality grids to generating reports and maps.

➢ See each step clearly explained and illustrated.

➢ Become familiarized with each of BenMAP's four buttons.

Chapter Overview

3. BenMAP Quick Start Tutorial

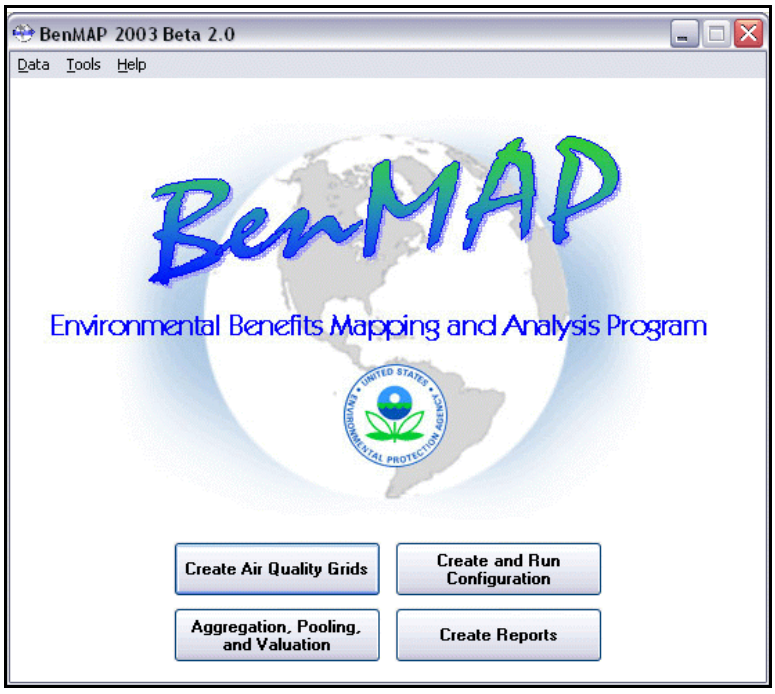
The best way to understand what BenMAP does is to start BenMAP and work through the following simple tutorial. The tutorial is based on a hypothetical scenario where ambient $PM_{2.5}$ concentrations are reduced by 10 percent in 2020. The steps in this analysis are as follows:

Step 1. Start BenMAP
Step 2. Create an Air Quality Grid for the Baseline Scenario
Step 3. Create an Air Quality Grid for the Control Scenario
Step 4. Specify Configuration Settings
Step 5. Select Concentration-Response Functions
Step 6. Specify Aggregation, Pooling and Valuation
Step 7. Generate Reports
Step 8. View Your Reports
Step 9. Map Your Results

Each step is explained in detail below.

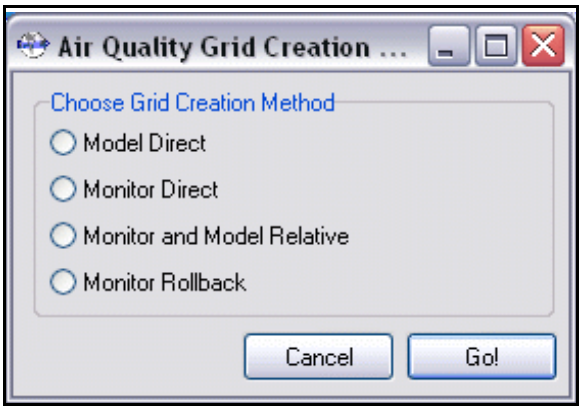
Step 1. Start BenMAP.

Double-click on the **BenMAP** icon on your desktop, and the following screen will appear:



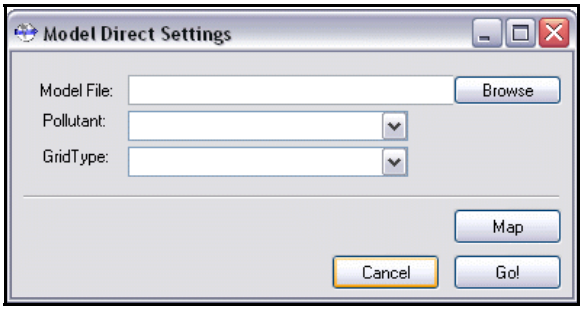
Step 2. Create an Air Quality Grid for the Baseline Scenario

Click on the **Create Air Quality Grids** button to begin inputting the air quality data needed by BenMAP. This will open up the window where you will input the air quality data. In general, you need two air quality grids to conduct a benefits analysis, one for a baseline scenario and one for the policy you are evaluating (the control scenario). We will start by entering the information about the baseline scenario. For the baseline scenario, we will use some of the modeling data that is provided with BenMAP.



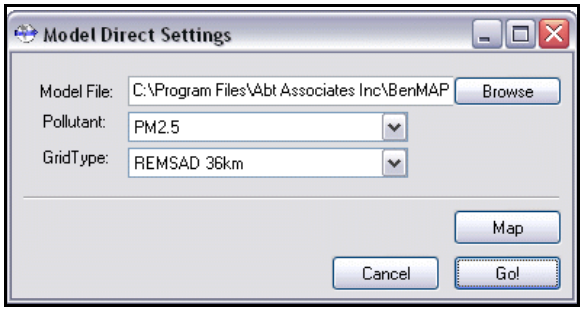
Select *Model Direct* from the list and click on **Go!** This will take you to the **Model Direct Settings** screen where you will enter the information about the baseline air quality modeling results you want to use.

To enter a **Model File** location, you can either type the path name or click on **Browse**. For this example, click on **Browse**. Browse to *C:\Program Files\Abt Associates Inc\BenMAP\Model Data* and click on the file *pm25-nonroad-Base2020.dat*, then click **Open**.



In the **Pollutant** field, use the pull down menu to select *PM2.5*.

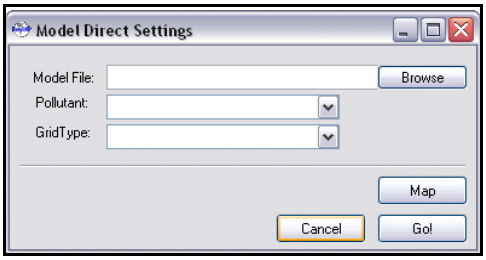
In the **Grid Type** field, use the pull down menu to select *REMSAD 36km*.



When your window looks like screen above, click Go!

BenMAP will now prompt you to save the air quality grid. Make sure you are in the *Air Quality Grids* subfolder in the *BenMAP* directory and then save the file as: *PM25 2020 REMSAD Direct example base.aqg* (you do not have to enter the ".aqg" extension). BenMAP will now create an air quality grid that you can use in your benefits analysis. When the progress bar is complete, BenMAP will return to the main screen.

Step 3. Create an Air Quality Grid for the Control Scenario

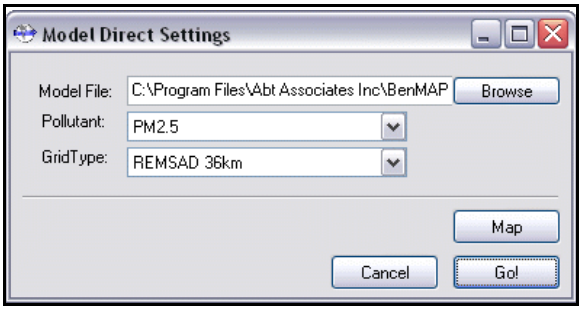


Click on the **Air Quality Grids** button on the main screen. Select *Model Direct* from the Air Quality Grid creation list and click **Go!**. This will take you back to the **Model Direct Settings** screen where you will enter the information about the air quality modeling data to be used for the control scenario.

To enter the **Model File** location, click **Browse** and browse to *C:\Program Files\Abt Associates Inc\BenMAP\Model Data* and click on the file *pm25-nonroad-2020TenPercentReduction.dat*, then click **Open**.

In the **Pollutant** field, use the pull down menu to select *PM2.5*.

In the **Grid Type** field, use the pull down menu to select *REMSAD 36km.*



.

When your window looks like the screen above, click **Go!**

BenMAP will now prompt you to save the air quality grid. Make sure you are in the *Air Quality Grids* subfolder in the *BenMAP* directory and then save the file as: *PM25 2020 REMSAD Direct example control.aqg* (you do not have to enter the ".aqg" extension). BenMAP will now create an air quality grid that you can use in your benefits analysis. When the progress bar is complete, BenMAP will return to the main BenMAP screen.

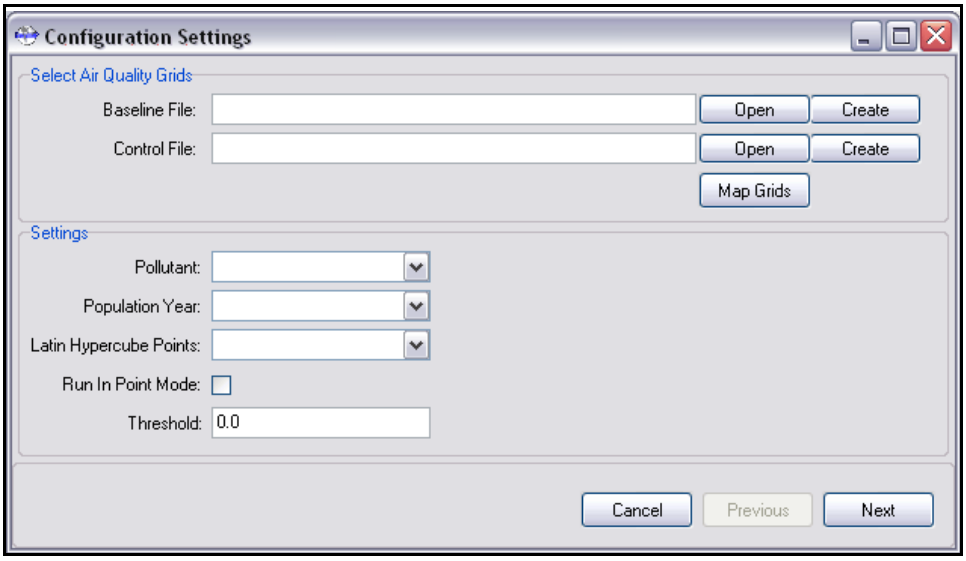
Step 4. Specify Configuration Settings

On the main BenMAP screen, click on the **Create and Run Configuration** button. In the following box, select *Create New Configuration* and click **Go!**

This will bring up the **Configuration Settings** form, where you will enter the basic information about your analysis before selecting the health effects you wish to estimate.

In the **Baseline File** field, you can either enter the path for your baseline air quality grid, or click **Open**. For this example, click **Open** and browse to the *Air Quality Grids* folder. Select *PM25 2020 REMSAD Direct example base* and click **Open**.

Next, click on **Open** next to the **Control File** field and select *PM25 2020 REMSAD Direct example control* and click **Open**.



This specifies that you want to conduct a benefits analysis of the difference between the baseline and control scenarios for which we created air quality grids in steps 3 and 4.

In the **Settings** section of this window, there are several fields which set the overall scope of the analysis.

In the **Pollutant** field, use the drop down menu to select *PM2.5*.

> This tells BenMAP that the pollutant you are analyzing is PM 2.5. Presently BenMAP only analyzes one pollutant at a time.

In the **Population Year** field, enter *2020* or select *2020* from the drop down menu.

This tells BenMAP that you want your analysis to use 2020 projected populations when calculating health impacts.

In the **Latin Hypercube Points** field, enter *10* or select *10* from the drop down menu.

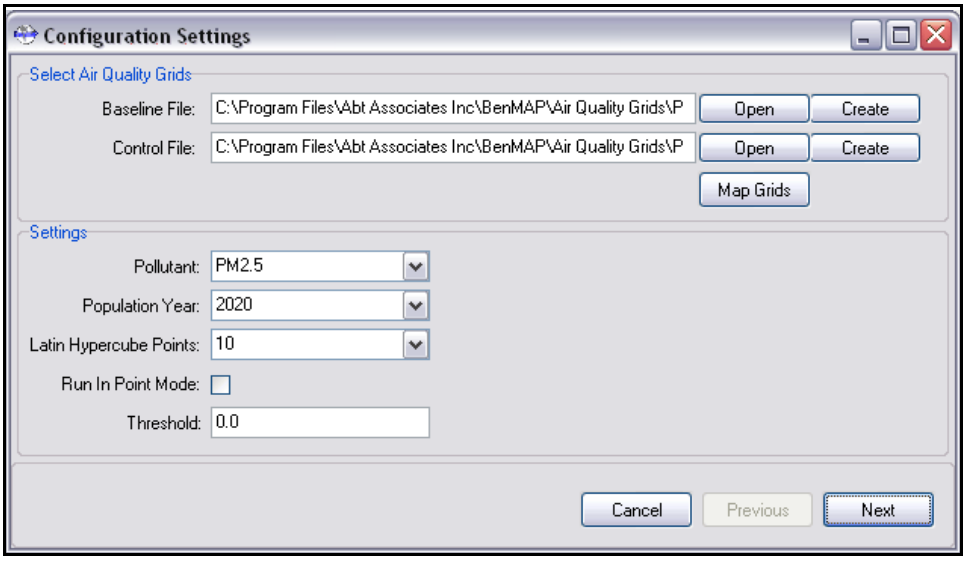
This tells BenMAP that you want to estimate the percentiles of the distribution of health endpoint incidence using Latin Hypercube Sampling with 10 percentiles of the distribution, representing the 5th, 15th, 25th, and so on up to the 95th percentile.

Leave the **Run in Point Mode** box unchecked.

You can only choose **Latin Hypercube Points** or **Run in Point Mode**. Since we are using the Latin Hypercube approach in this example, you must leave this box unchecked.

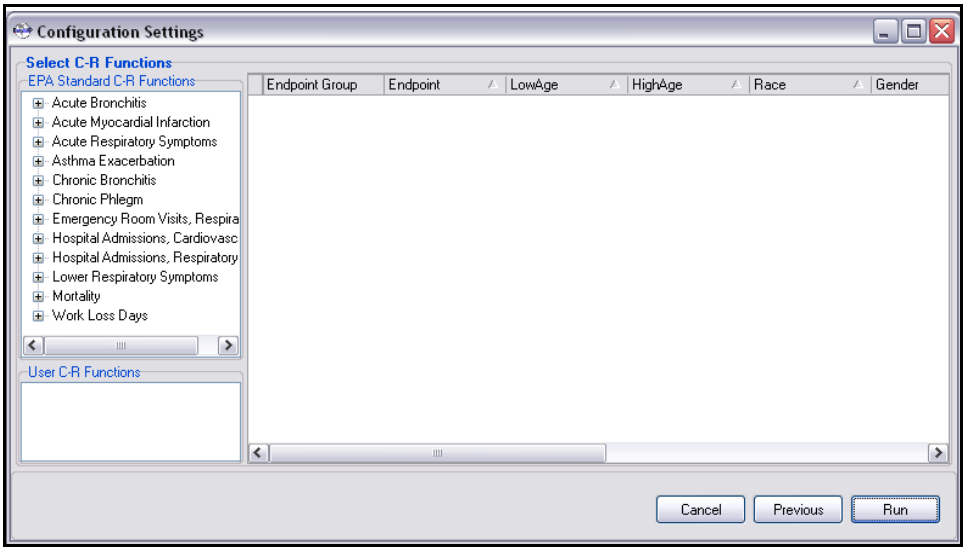
Leave the **Threshold** field at 0.0.

This tells BenMAP that you want to estimate benefits associated with all changes in PM2.5, regardless of where those changes occur along the range of PM2.5 concentrations. Selecting a non-zero threshold means that you would only want to calculate benefits for changes occurring above the threshold.



When your window looks like the above, click Next.

This will bring up the next page of the **Configuration Settings** form: the **Select C-R Functions** screen.



(Note: This screen can be resized if you are having trouble seeing all of the information. Individual columns can also be resized. Just click on the border of a column and drag to increase or decrease its width.)

Step 5. Select Concentration-Response Functions

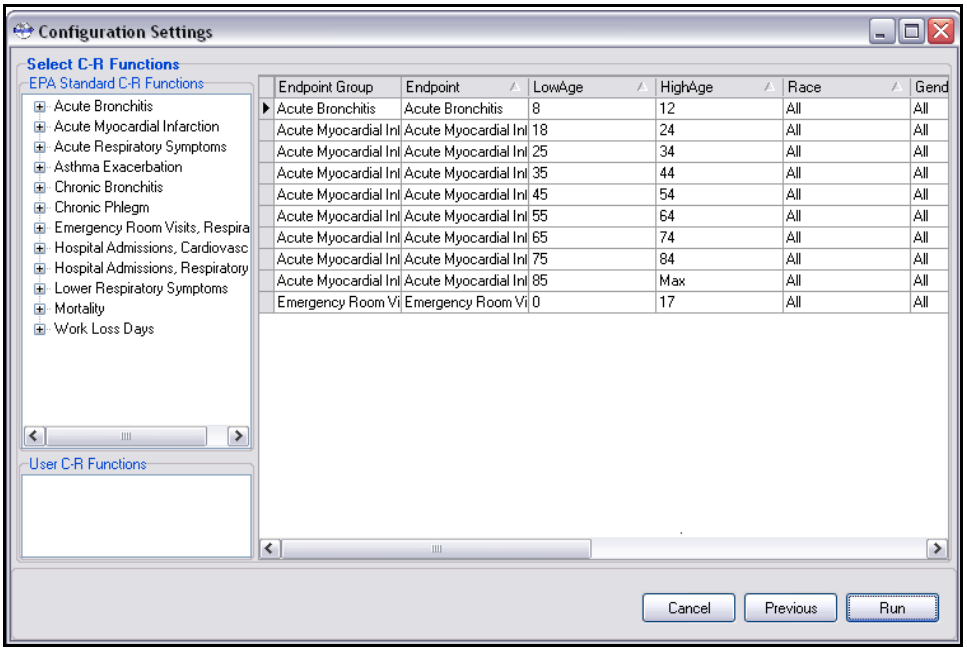
In this screen, you can select C-R functions to use in your analysis. For this example, we are going to estimate the change in incidence of three health endpoints associated with PM2.5: acute bronchitis, acute myocardial infarctions (heart attacks), and emergency room visits for asthma. To select a C-R function, you must drag it from the left-hand side of the screen to the list on the right. You can drag groups of concentration-response functions over, or drill down and drag over individual C-R functions.

For acute bronchitis, drill down until you see a group titled *Dockery et al*. Open this heading, and drag the C-R function titled *Dockery et al., 1996 | 8-12* into the right hand panel of the window. You should see a new row starting with the **Endpoint Group** *Acute Bronchitis.*

For acute myocardial infarctions (AMI), drag the entire group titled *Acute Myocardial Infarction* to the right hand panel (do not drill down). This will include the full set of age-specific C-R functions for AMI. There will be an "extra" C-R function in the list of AMI C-R functions in the right hand panel, covering an all ages version of the C-R function. Find this function by looking for the row with *18* in the **Low Age** column and *Max* in the **High Age** column. Click on this row and press the delete key to remove the function.

For asthma emergency room visits, drill down the in the group title *Emergency Room Visits, Respiratory* until you see individual C-R functions. Drag the C-R function titled *Norris, et al. 1999 | NO2, SO2* into the right hand panel of the window.

You should now have ten C-R functions listed in the right hand panel: one acute bronchitis function, eight AMI functions, and one ER visit function.



When your window looks like the above, click on **Run**.

BenMAP will then prompt you to save your file. Click **Save**. Browse to the *Configurations* subfolder within the *BenMAP* directory and save the file as:
PM25 example config.cfg (you do not need to include the ".cfg" extension).

When you have saved the configuration file, click **OK** to run the configuration.

BenMAP will prompt you to "Save Configuration Results to File". Browse to the *Configuration Results* subfolder within the BenMAP directory and save the file as: *PM25 2020 REMSAD Direct example.cfgr* (you do not need to include the ".cfgr" extension)

The first time you are prompted to save your configuration is for the options and C-R functions you have chosen. The second time you are prompted to save your results- the grid-cell level changes in incidence.

Once you have entered the filename, BenMAP will begin calculating the change in incidence for the set of C-R functions you have selected. The run may take a few minutes to finish; a progress bar will let

you know how it is proceeding. When BenMAP is finished running your configuration, it will return to the main BenMAP screen.

Step 6. Specify Aggregation, Pooling and Valuation

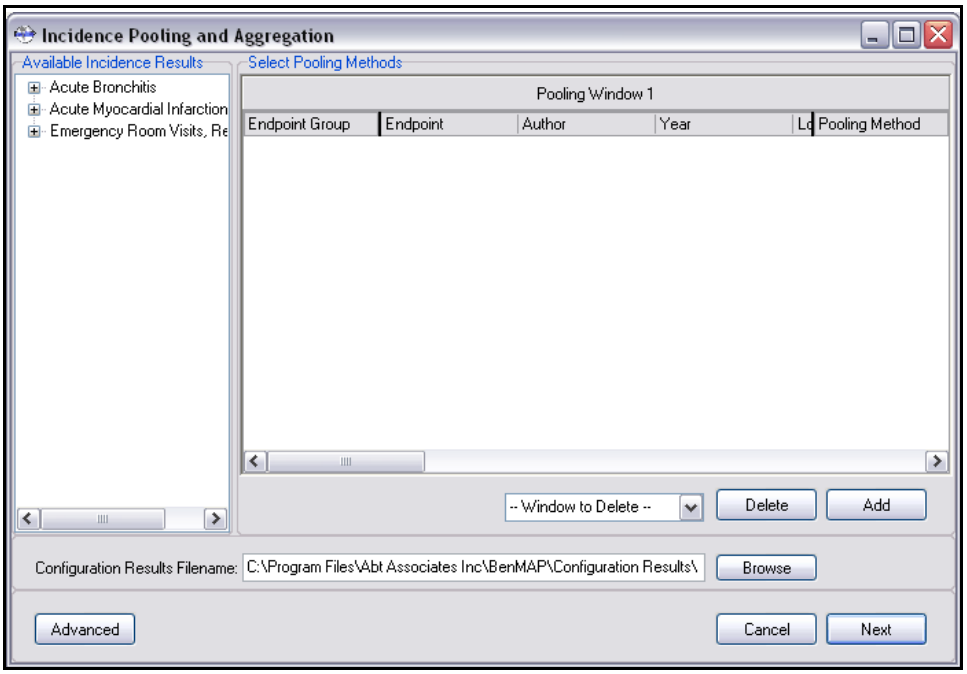
This step allows you to take the incidence results that BenMAP just produced and place an economic valuation on them, Although not covered in this tutorial, this is also where you can select the geographic level of aggregation and combine individual incidence results into pooling groups.

From the main screen, click on the **Aggregation, Pooling and Valuation** button. This will bring up a menu screen with two choices: *Create New Configuration for Aggregation, Pooling and Valuation*, or *Open Existing Configuration for Aggregation, Pooling and Valuation (*.apv file)*.

Select *Create New Configuration for Aggregation, Pooling and Valuation* and click on **Go!**

BenMAP will prompt you to open a **Configuration Results File**. Browse to the *Configuration Results* subfolder and select *PM25 2020 REMSAD Direct example.cfgr*. Then click on **Open**.

BenMAP will then open the **Incidence Pooling and Aggregation** window with the results from running your configuration. You should see a window that looks like the following:



Click on each of the results groups (acute bronchitis, acute myocardial infarction, and emergency room visits) and drag them to the right panel.

For this example, we are not pooling any of the incidence results (although we will pool valuations in the next window), so just click on **Next** at the bottom of the window.

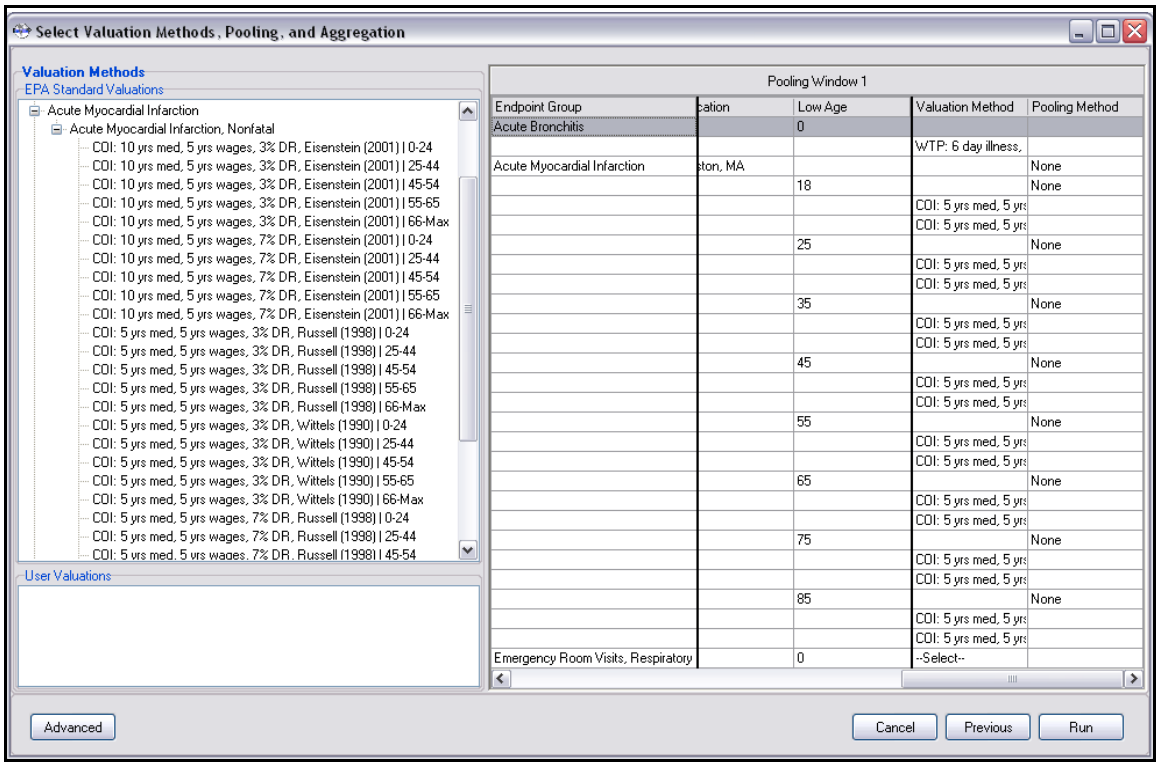
This will take you to the **Select Valuation Methods, Pooling, and Aggregation** window.

A) Select a value for acute bronchitis
To select a valuation method for acute bronchitis, drill down the *Acute Bronchitis* valuation group until you see individual valuation methods. Click on the *WTP: 6 day illness, CV studies | 0-17* method and drag it onto the *Acute Bronchitis* endpoint group in the right hand panel. You should see the method appear under acute bronchitis in the **Valuation Method** column in the right hand panel.

B) Select values for acute myocardial infarctions (heart attacks)
To select valuation methods for acute myocardial infarctions, drill down the AMI valuation group until you see a (long) list of individual valuation methods. You might find it easier to expand the column width of the **Valuation Methods** column (drag the right hand edge of the column to the right to make it wider). We will be working with the valuation estimates from two studies, *Wittels* and *Russell.* For each of the studies, there are a number of age specific valuations. There are also two different discount rates (the discount rate is the rate at which future medical costs are discounted to the present). Drag the age-specific valuation estimates from *Wittels.* for the 3 percent discount rate (COI, 5 yrs med, 5 yrs wages, 3% DR, Wittels (1990) | *age*) to each matching age-specific line in the right hand panel (**Pooling Window 1**). You may have to scroll over in the right hand panel to see the **Low Age** column.

Note that you will need to drag some age-specific valuation estimates to multiple lines in the pooling window, since there is not a perfect match between the available age-specific valuation estimates and the age groups for which the incidence of heart attacks was estimated. For example, you will have to drag the valuation estimate for the *25 to 44* age group to both the *25 to 35* age group and the *35 to 45* age group in the pooling window.

Now repeat this process using the Russell 3 percent discount rate valuation estimates. When you are finished, you should have two valuation estimates for each AMI age group, and your pooling window should look like the one below.



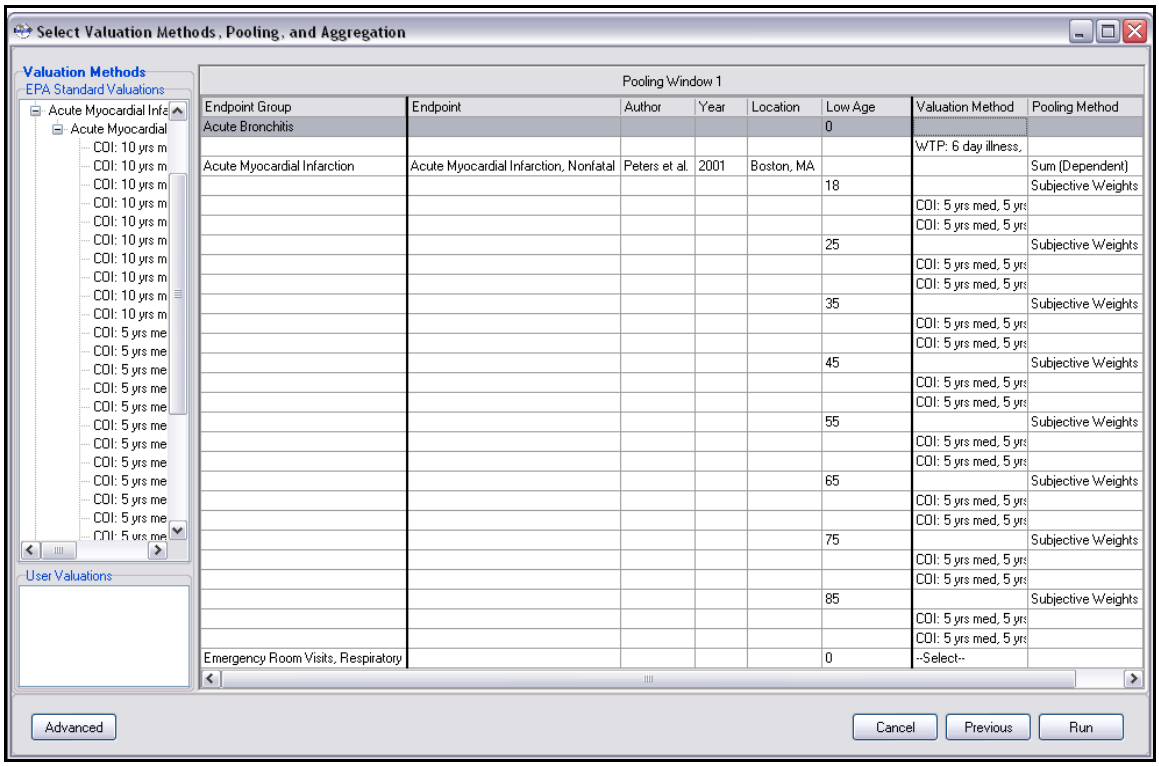
Now you can pool the valuation results for heart attacks in each age group using the unit values from both Wittels and Rusell. In order to do so, you must select a pooling method.

BenMAP lets you select from several different pooling methods. For this example, you will be using subjective weights. In other applications, you may wish to use fixed or random effects weights (see Chapter 6 for more information on pooling methods).

To set the pooling method for each age group result, click on the **Pooling Method** field in the row ABOVE each pair of valuation methods (where it says *None*) and use the drop down menu to select *Subjective Weights*. You must repeat this for EACH age group in order for pooling to take place over all age groups.

In addition to pooling the results over the two valuation methods, we also need to aggregate the results into a total estimate across age groups. In order to do so, in the row with **Endpoint Group** (**Endpoint Group** = *Acute Myocardial Infarction*) click in the **Pooling Method** field and select *Sum (Dependent)* from the drop down menu.

Your screen should look like the following:

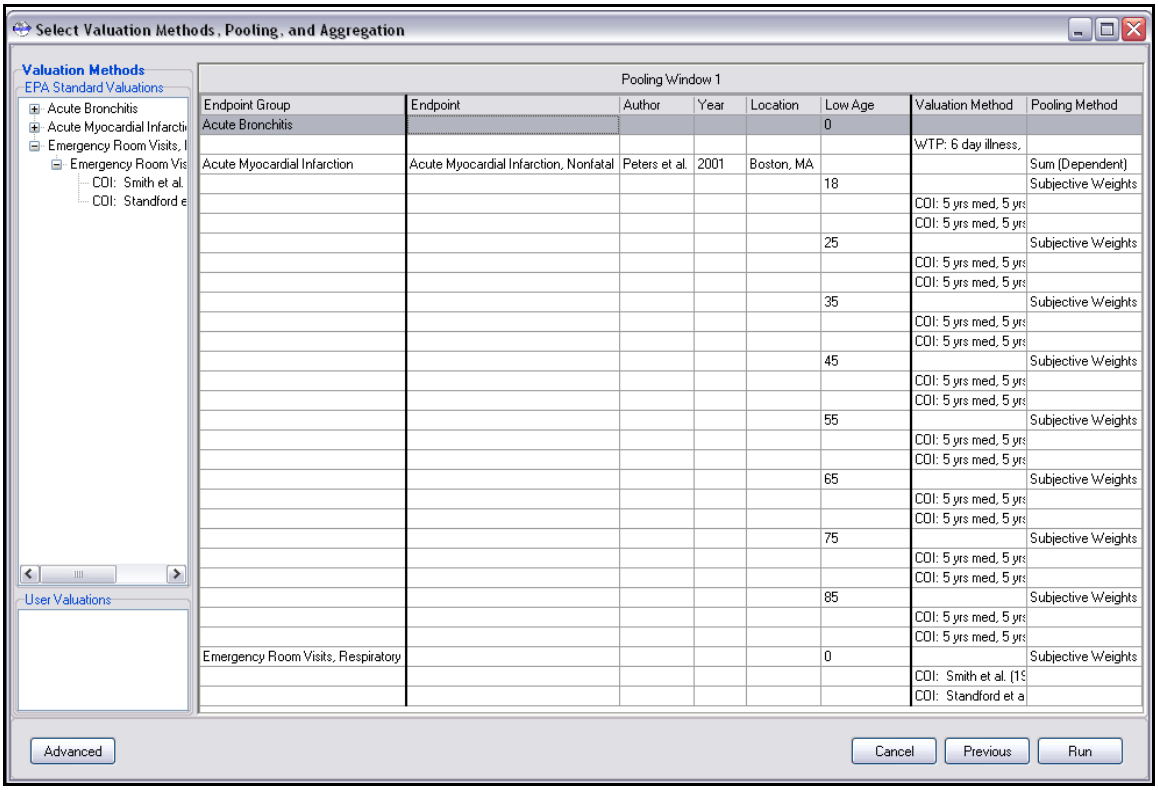


This pooling configuration for acute myocardial infarctions will assign a starting set of equal weights to each valuation method for the set of eight age groups, and then create an overall estimate of acute myocardial infarctions by summing the age-specific pooled estimates, treating the distributions for each age group as dependent (i.e. a draw from the 5th percentile of the 45 to 54 age group will be added to the draw from the 5th percentile of the 55 to 64 age group and so on).

C) Select values for asthma emergency room visits

To select values for asthma ER visits, drill down the *Emergency Room Visits, Respiratory* heading to the *Emergency Room Visits, Asthma*, and then to the individual valuation approaches. Drag both methods (COI: Smith et al, 1997 | 0-Max and COI: Standford et al, 1999 | 0 - Max) to the *Emergency Room Visits* entry in the right hand panel.

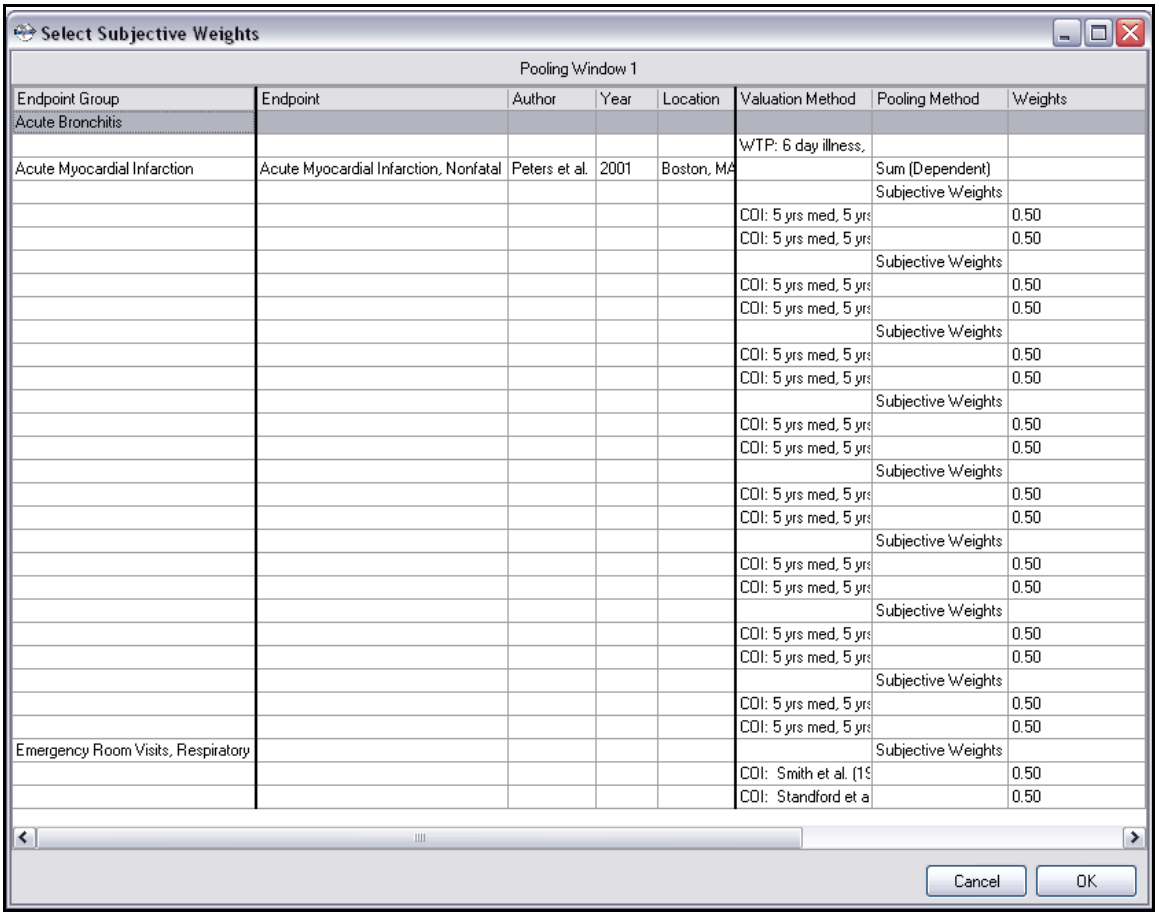
Again, this will now allow you to assign a pooling method to pool the valuation results using the two valuation methods. For this exercise, click on the pooling method in the emergency room visits row (where it says *None*), and select *Subjective Weights* from the drop down menu. Your completed screen should look like the one below.



D) Entering subjective weights

Once you have completed this step, click on **Run**. BenMAP will now bring up a window to allow you to enter subjective weights.

BenMAP assigns a default equal weight to each selected valuation method. You can change these weights by clicking in the weight cells. However, for this exercise, you should leave them at *0.5* for each study. Click on **OK** at the bottom of the screen. You should see a save dialog box. Click on **Save** to save your APV configuration. Save the file as *PM25 Direct example APV*.



Click on **OK** to start the pooling and aggregation. You will be prompted to enter a filename for the pooling and aggregation results file. Enter *PM25 Direct example APV Results* and click on **Save**. BenMAP will display a progress bar for the pooling and aggregation. When the progress bar disappears, you will be returned to the main BenMAP screen.

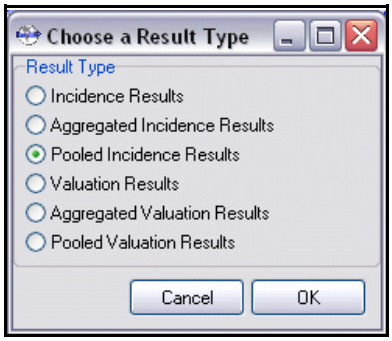
Step 7. Generate Reports

You may view your results within BenMAP, either in the preview window in the **Create Reports** button or through the mapping functions. Alternatively, you may export the results to a comma separated values file (*.csv), or a shape file (*.shp) which can be viewed in a GIS program such as ESRI's ArcView product.

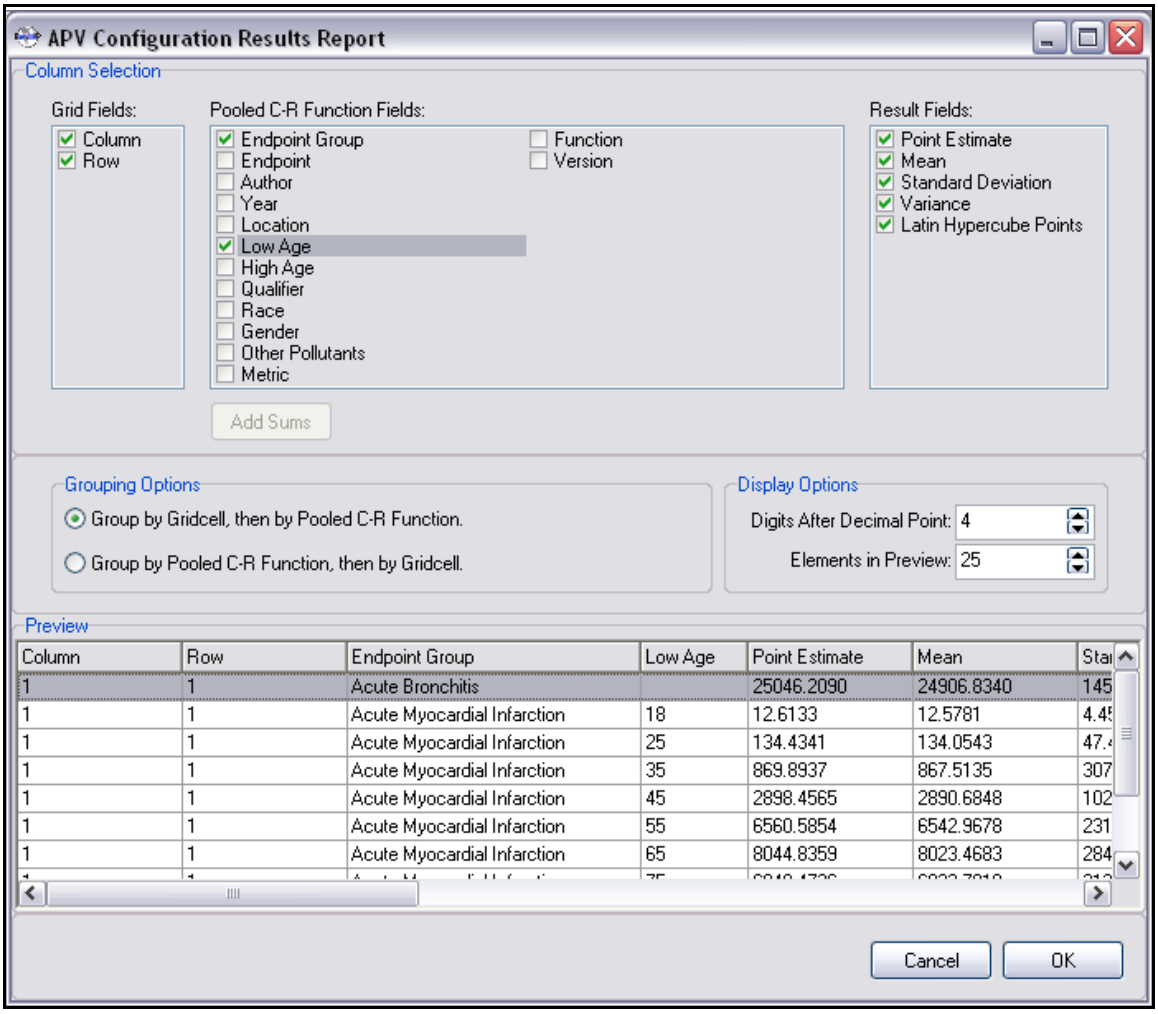
A) Generate a **Pooled Incidence Results** report.

A **Pooled Incidence Results** report contains the incidence results you previously generated, using the aggregation level and pooling that you specified in the **Incidence Pooling and Aggregation** window. Previously, in Step 6, we did not specify any pooling of incidence results (although valuations were pooled), so in this case the **Pooled Incidence Results** report will look just like the **Aggregated Incidence Results** report. If some incidence results had been pooled, the two reports would be different.

Click on the **Create Reports** button from the main BenMAP screen. This will bring up the **Select Report Type** window. Select *Incidence and Valuation Results: Raw, Aggregated and Pooled.* Click **OK**. This will bring up the **Select an APV Configuration Results File** screen. Select the *PM25 Direct example APV Results.apvr* file, and click **Open**.

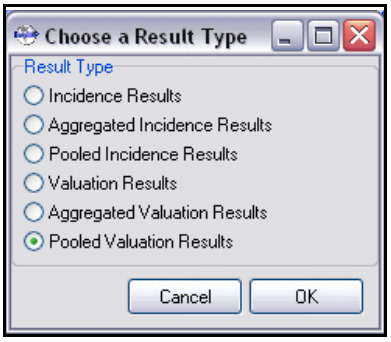


In the **Choose a Result Type** window, choose *Pooled Incidence Results*. Then click **OK**. This will bring up the **Results Grid Report Form**, where you can customize your report display and select the fields you want to see in the report. In the **Pooled C-R Function Fields** box, check off **Endpoint Group** and **Low Age.** Then click **OK**. In the **Save Results to File** window type in the file name and browse to the location where you want to store the exported file. The **Reports** subfolder is a good location to keep exported reports. Type in the name, *PM25 Direct example APV Incidence results* in the box and click **Save**. You can now open the report in another application, such as a spreadsheet or database program.



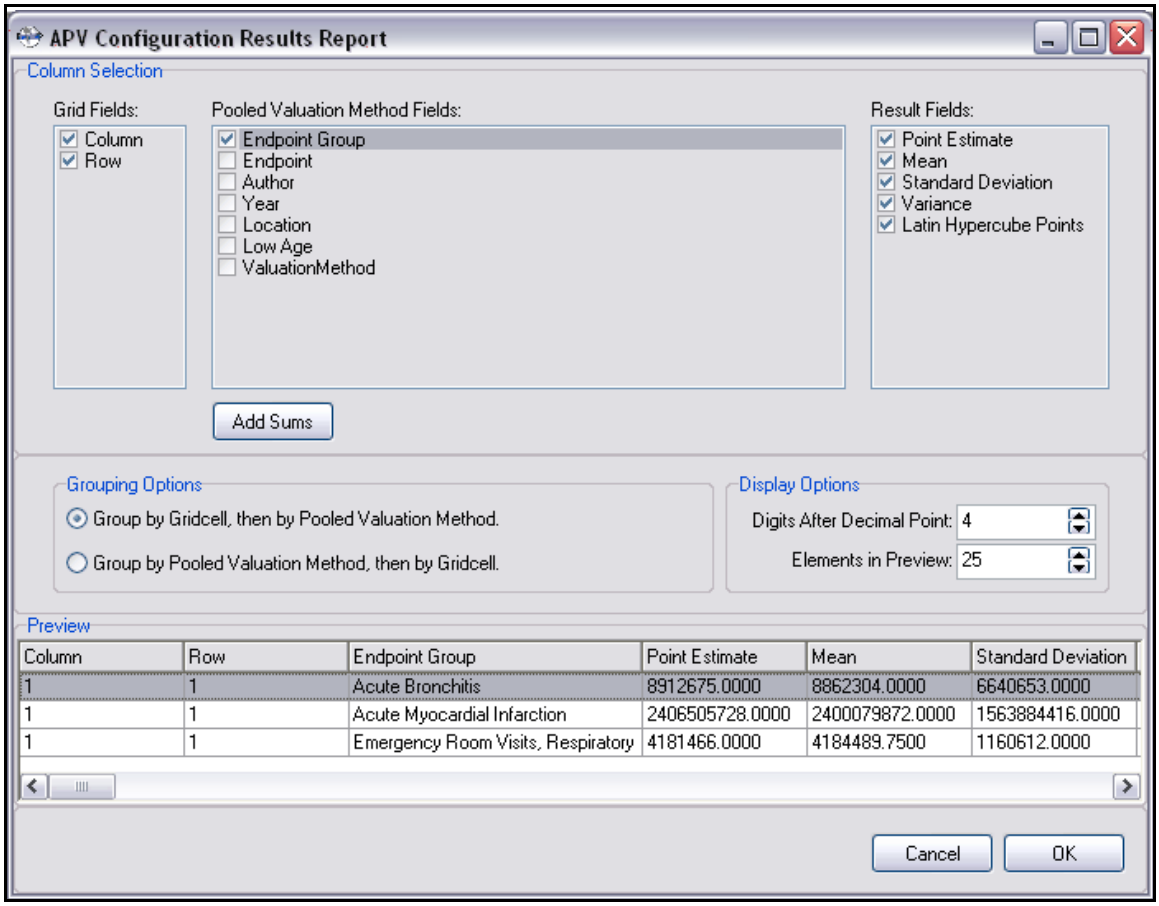
B) Generate a **Pooled Valuation Results** report.

This report is similar to the **Pooled Incidence Results** report, and uses the valuation pooling you previously specified in the **Valuation Pooling and Aggregation** window. Click on the **Create Reports** button from the main BenMAP screen. This will bring up the **Select Report Type** window. Select *Incidence and Valuation Results: Raw, Aggregated and Pooled.* Click **OK**. This will bring up the **Select an APV Configuration Results File** screen. Select the *PM25 Direct example APV Results.apvr* file, and click **Open**.



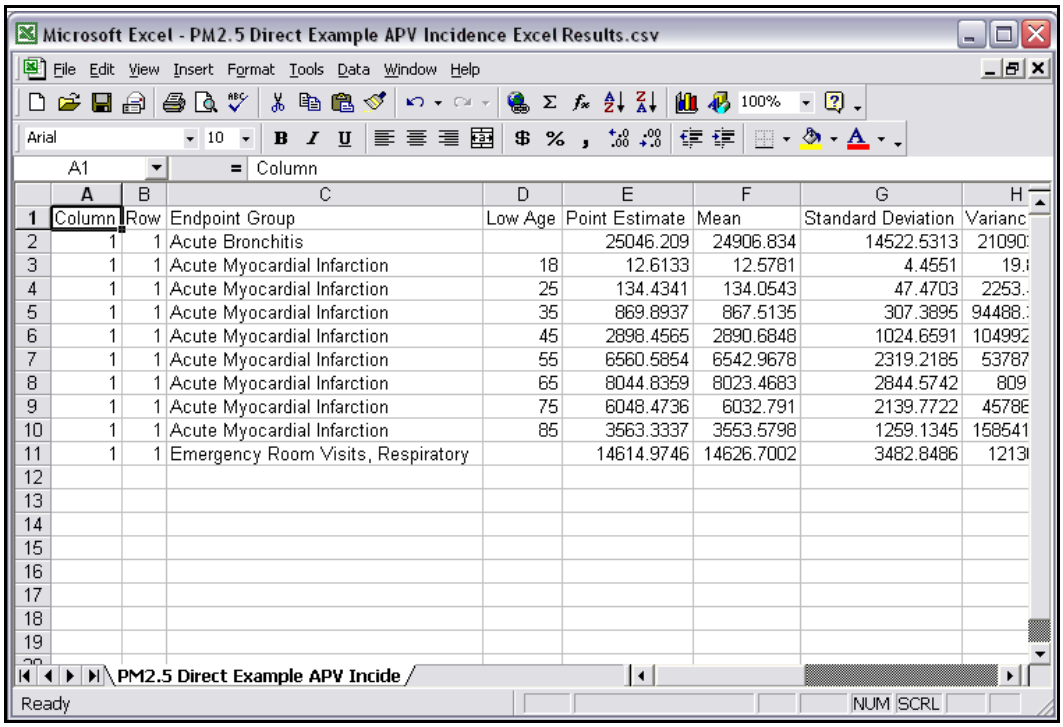
In the **Choose a Result Type** window, choose *Pooled Valuation Results*. Then click **OK**. This will bring up the **Results Grid Report Form**, where you can customize your report display and select the fields you want to see in the

report. In the **Pooled C-R Function Fields** box, check off **Endpoint Group**. Then click **OK**. In the **Save Results to File** window type in the file name and browse to the location where you want to store the exported file. Type in the name, *PM25 Direct example APV valuation results* in the box and click **Save**. You can now open the report in another application, such as a spreadsheet or database program.



Step 8. View Your Reports

BenMAP generates comma separated values files (*.csv) that can be read by various spreadsheet and database applications, such as Microsoft Excel.



Microsoft Excel - PM2.5 Direct Example APV Incidence Excel Results.csv

File Edit View Insert Format Tools Data Window Help

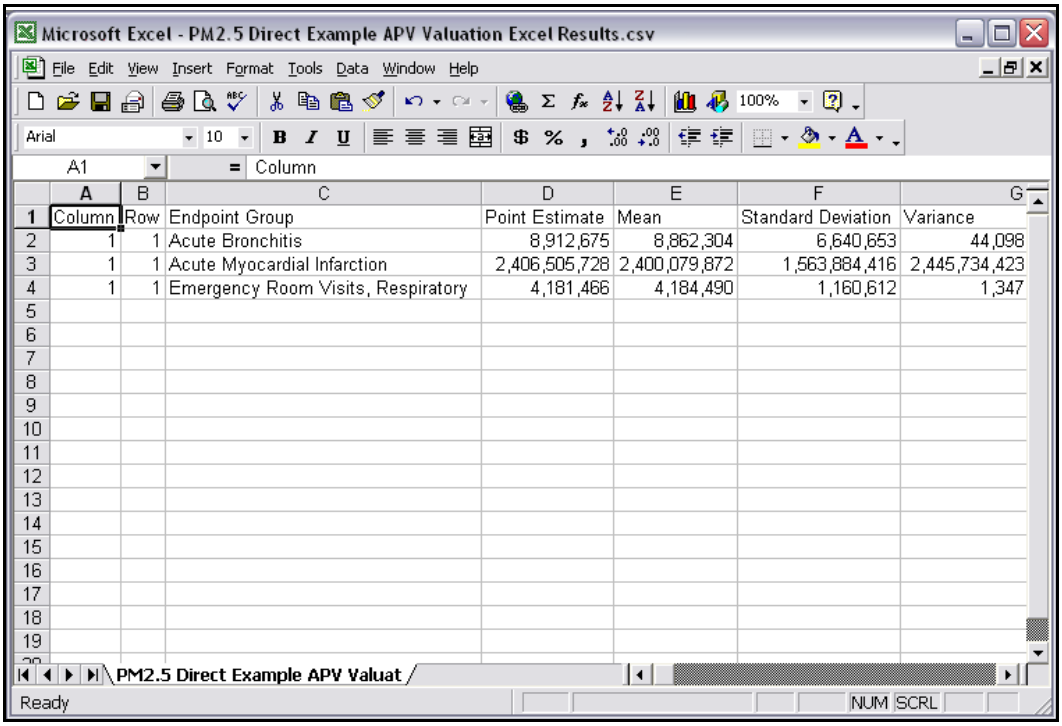
Arial 10 100%

A1 = Column

	A	B	C	D	E	F	G	H
1	Column	Row	Endpoint Group	Low Age	Point Estimate	Mean	Standard Deviation	Varianc
2	1	1	Acute Bronchitis		25046.209	24906.834	14522.5313	21090
3	1	1	Acute Myocardial Infarction	18	12.6133	12.5781	4.4551	19.
4	1	1	Acute Myocardial Infarction	25	134.4341	134.0543	47.4703	2253.
5	1	1	Acute Myocardial Infarction	35	869.8937	867.5135	307.3895	94488.
6	1	1	Acute Myocardial Infarction	45	2898.4565	2890.6848	1024.6591	104992
7	1	1	Acute Myocardial Infarction	55	6560.5854	6542.9678	2319.2185	53787
8	1	1	Acute Myocardial Infarction	65	8044.8359	8023.4683	2844.5742	809
9	1	1	Acute Myocardial Infarction	75	6048.4736	6032.791	2139.7722	45786
10	1	1	Acute Myocardial Infarction	85	3563.3337	3553.5798	1259.1345	158541
11	1	1	Emergency Room Visits, Respiratory		14614.9746	14626.7002	3482.8486	1213
12								
13								
14								
15								
16								
17								
18								
19								

PM2.5 Direct Example APV Incide

Ready NUM SCRL



Microsoft Excel - PM2.5 Direct Example APV Valuation Excel Results.csv

File Edit View Insert Format Tools Data Window Help

Arial 10 100%

A1 = Column

	A	B	C	D	E	F	G
1	Column	Row	Endpoint Group	Point Estimate	Mean	Standard Deviation	Variance
2	1	1	Acute Bronchitis	8,912,675	8,862,304	6,640,653	44,098
3	1	1	Acute Myocardial Infarction	2,406,505,728	2,400,079,872	1,563,884,416	2,445,734,423
4	1	1	Emergency Room Visits, Respiratory	4,181,466	4,184,490	1,160,612	1,347
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							

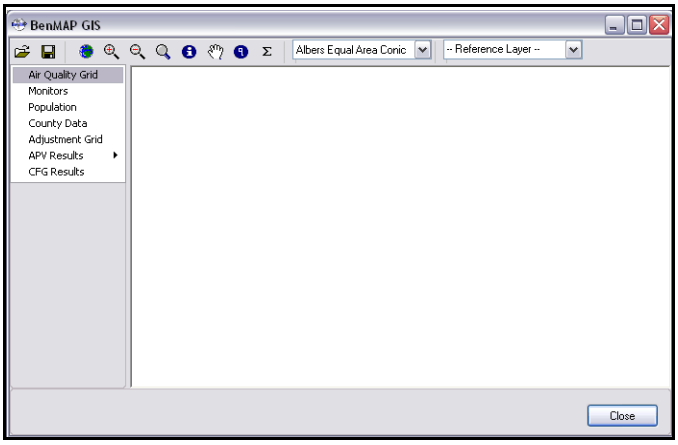
PM2.5 Direct Example APV Valuat

Ready NUM SCRL

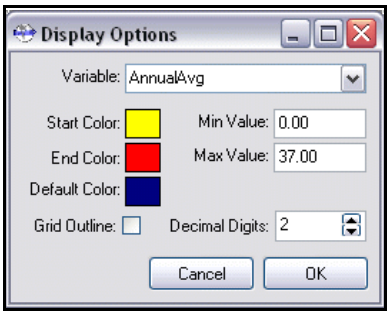
Step 9. Map Your Results

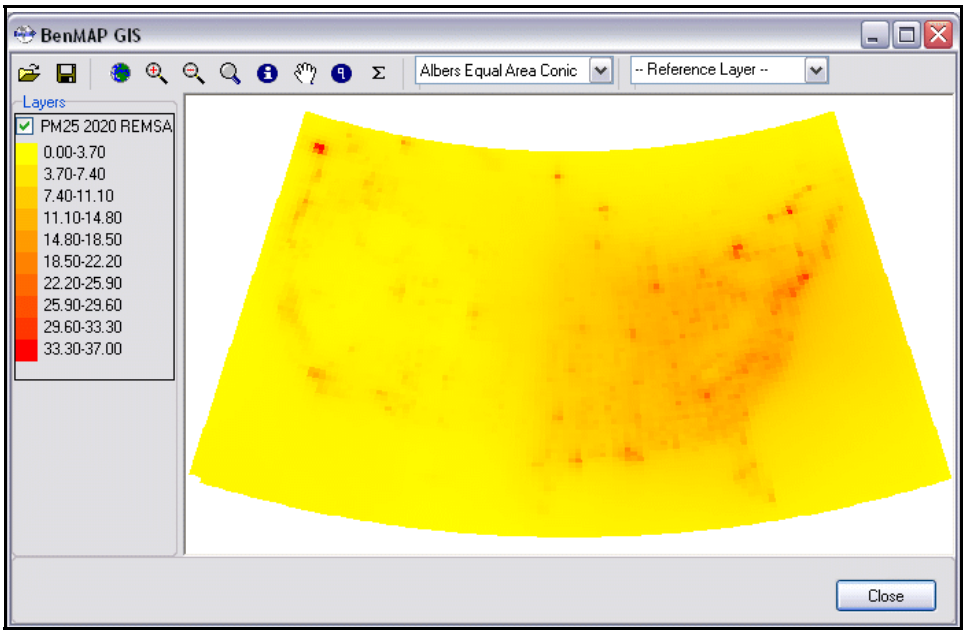
You can also map any of the results that you have generated so far. This includes the air quality grids, population data, incidence results, and valuation results. In this example, we will look at the air quality grid for the base scenario, and view our incidence results. For more information on these and other mapping functions, see Chapter 8.

To use the BenMAP mapping functionality, go to the **Tools** menu and choose *Mapping / GIS*. The BenMAP GIS window will appear, with buttons at the top for managing files and navigating the map. To see the name of each button, simply hold the cursor over it. Click on the **Open a file** button, and select *Air Quality Grid* from the drop-down menu. Browse to the file *PM25 2020 REMSAD Direct example control.aqg* file and click **Open**.

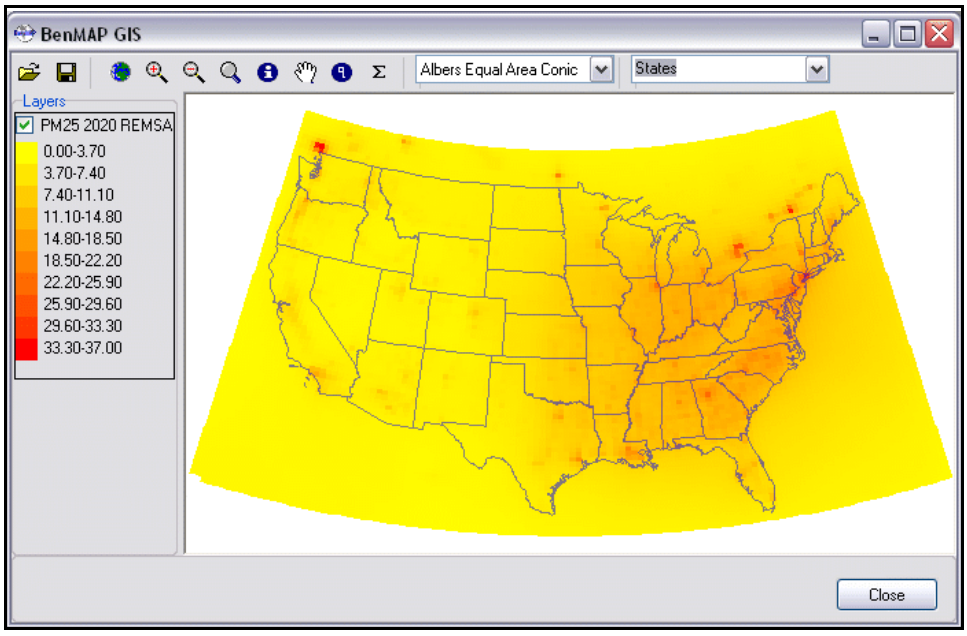


The name of the file will appear in the left-hand panel under **Layers**. Double-click on the name and a small box will appear with **Display Options** for viewing this layer. Here you select the variable contained in the layer (file) that you want to view. In the air quality grid, the variables that are available are the annual average, the annual median, and 24-hr daily average. Select *AnnualAvg* for the annual average in the **Variable**. In this box, you can also change the colors in the map display, and the maximum and minimum values displayed. Leave the defaults as they are, but uncheck the box marked **Grid Outline**. You should see the map below.

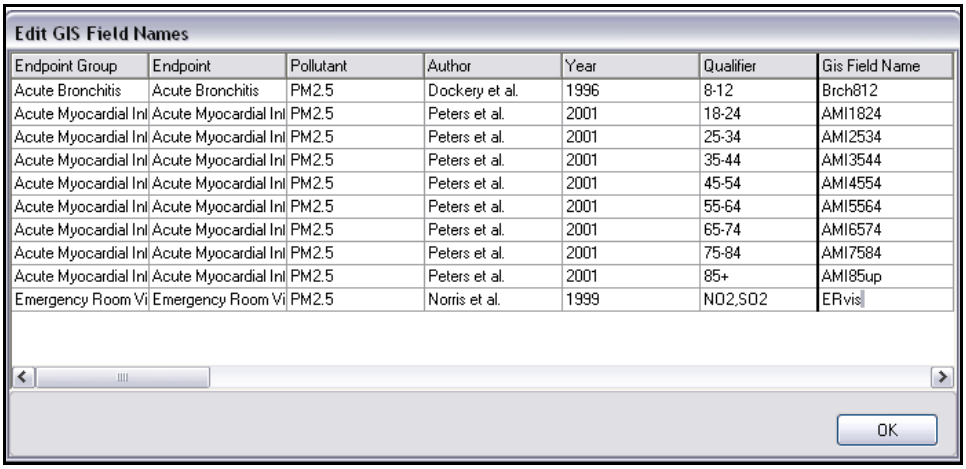




To see state outlines, select *States* from the **Reference Layer** drop-down menu at the top of the screen.



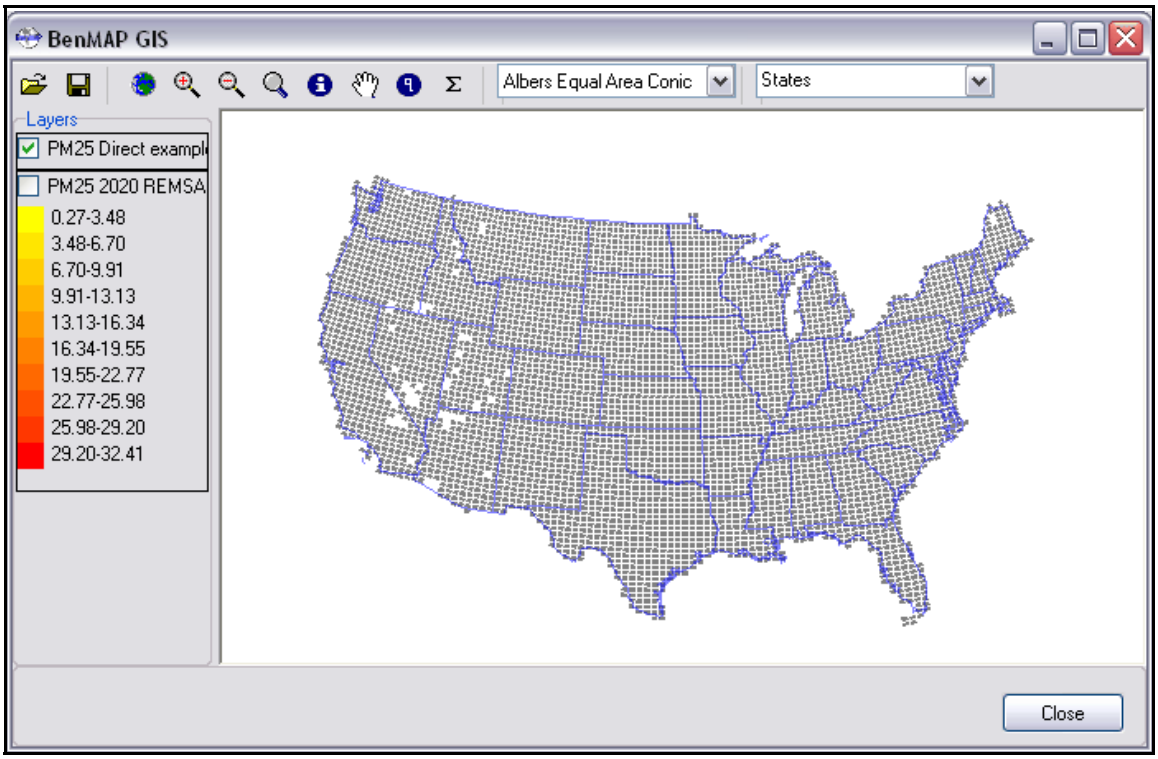
Now we can look at a geographical display of the incidence results we created for cases of bronchitis, acute myocardial infarctions, and emergency room visits. Click on the **Open a file** button at the top of the screen and select *APV Results*, then *Incidence.* In the next window, select *PM25 direct example APV results.apvr*, then click **Open**. BenMAP will load your incidence results and display them in a table. Because GIS programs can typically only accommodate field names that are 10 characters or less, there is a new column at the end of the table labeled **Gis Field Name**. Here you can name your variables, as shown in the table below.



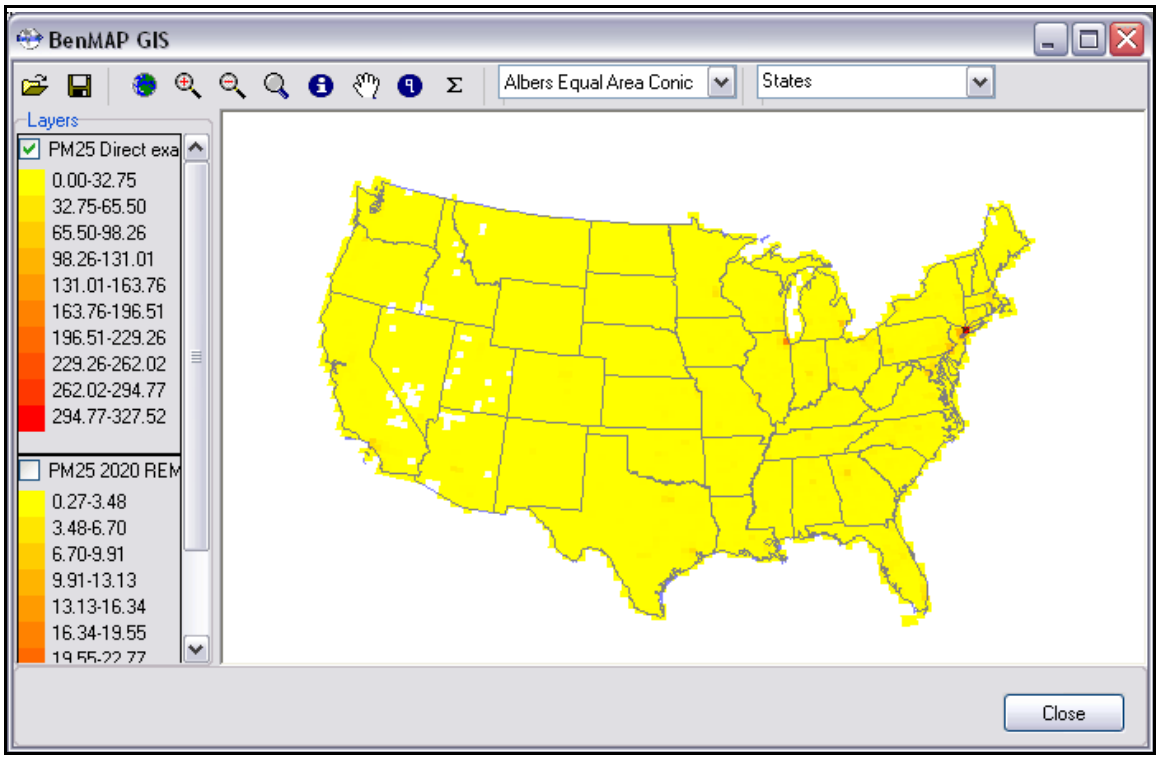
Endpoint Group	Endpoint	Pollutant	Author	Year	Qualifier	Gis Field Name
Acute Bronchitis	Acute Bronchitis	PM2.5	Dockery et al.	1996	8-12	Brch812
Acute Myocardial Inf	Acute Myocardial Inf	PM2.5	Peters et al.	2001	18-24	AMI1824
Acute Myocardial Inf	Acute Myocardial Inf	PM2.5	Peters et al.	2001	25-34	AMI2534
Acute Myocardial Inf	Acute Myocardial Inf	PM2.5	Peters et al.	2001	35-44	AMI3544
Acute Myocardial Inf	Acute Myocardial Inf	PM2.5	Peters et al.	2001	45-54	AMI4554
Acute Myocardial Inf	Acute Myocardial Inf	PM2.5	Peters et al.	2001	55-64	AMI5564
Acute Myocardial Inf	Acute Myocardial Inf	PM2.5	Peters et al.	2001	65-74	AMI6574
Acute Myocardial Inf	Acute Myocardial Inf	PM2.5	Peters et al.	2001	75-84	AMI7584
Acute Myocardial Inf	Acute Myocardial Inf	PM2.5	Peters et al.	2001	85+	AMI85up
Emergency Room Vi	Emergency Room Vi	PM2.5	Norris et al.	1999	NO2,SO2	ERvis

When you are satisfied with the variable names, click **OK.** The new layer will show in the BenMAP GIS window on top of the first layer. If the previous layer is still checked, then it will appear, but underneath the new layer. Uncheck the box next to the bottom (previous) layer to hide it. Your screen should look like the one below.

You do not have to rename your variables, but it will make it easier to identify them when selecting a variable to map. The default names (Result0, Result1, etc,) make it difficult to identify individual results.



Like the previous layer, double click on the name to bring up the **Display Options** box. Under **Variable** you will see a list of the variable names you defined in the previous step. Select *AMI7584*, uncheck the **Grid Outline** box, and click **OK**. The viewer will now display the annual increase in the number of acute myocardial infarctions for people 75 to 84, as calculated between the base and control scenarios. You can use the **Display Options** to select other variables to view or change how the values are displayed.


BenMAP GIS
Albers Equal Area Conic
States
Layers
PM25 Direct exa
0.00-32.75
32.75-65.50
65.50-98.26
98.26-131.01
131.01-163.76
163.76-196.51
196.51-229.26
229.26-262.02
262.02-294.77
294.77-327.52
PM25 2020 REM
0.27-3.48
3.48-6.70
6.70-9.91
9.91-13.13
13.13-16.34
16.34-19.55
Close

In this chapter...

➢ Create air quality grids using different methods.

➢ Find details on the file structures for data inputs.

➢ Interpolate with Closest Monitor or Voronoi Neighbor Averaging.

➢ Scale monitor data with modeling data.

➢ Learn about advanced options like monitor filtering.

CHAPTER 4

Creating Air Quality Grids

Chapter Overview

4. Creating Air Quality Grids

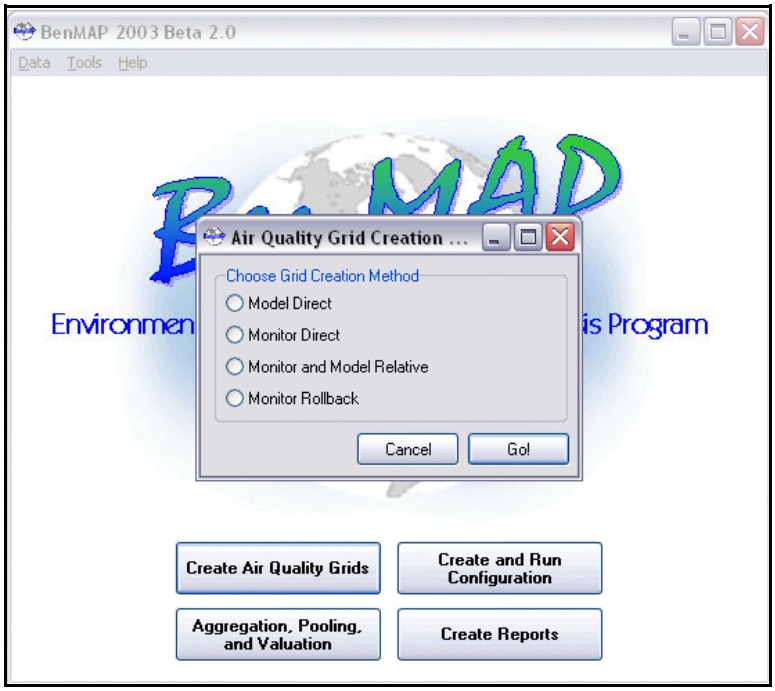
BenMAP is not an air quality model, nor can it generate air quality data independently. Instead it relies on the air quality inputs given to it. To estimate population exposure to air pollution, BenMAP uses air quality grids that it generates from input air quality data (modeling and or monitoring data).

BenMAP creates air quality grids to estimate the average exposure to ambient air pollution of people living in some specified area, or domain, such as that delineated by REMSAD, CAMx, and CMAQ models, as well as more irregular shapes, such as counties. However, BenMAP does not estimate personal exposure. Instead, the air quality grids provide the average population exposure for each grid cell that BenMAP can then use in C-R functions.

Air Quality Grids contain air pollution data. BenMAP uses one baseline and one control air quality grid and estimates the change in the number of adverse health effects between the two. **CAMx, CMAQ, REMSAD**, and **UAM-V** are all air quality models that generate data used by BenMAP to create air quality grids.

To create air quality grids, BenMAP uses a number of inputs, including modeling data, monitoring data, or both. You may enter your own modeling and monitoring data, provided that the data are in a format recognized by BenMAP. In addition, BenMAP comes supplied with some sample REMSAD modeling files, as well as a growing archive that currently has ozone, $PM_{2.5}$, and PM_{10} monitoring data from the years 1996-2002 . The current version of BenMAP creates files for four pollutants: ozone, $PM_{2.5}$, PM_{10}, and PMC (coarse particles).

Note that during the creation of air quality grids, you have access to some advanced options that allow you to use values other than the defaults typically used. For example, in using monitor data, you can specify a subset of the available data, such as a particular region of the United States or monitors that meet certain completeness criteria. We discuss the advanced options, as well as the more standard approaches, below in the appropriate sections.



To start the grid creation process, click on the **Create Air Quality Grids** button. BenMAP will then ask which of the following types of air quality data you wish to use:

➢ **Model Direct.** Choose this option if you have air quality modeling data from REMSAD, CAMx, UAM-V or CMAQ that you wish to use directly. Exhibit 4-1 below describes the input format for each grid type.

➢ **Monitor Direct.** Choose this option if you wish to import your own monitoring data (see Exhibit 4-3 for format), or you wish to use monitoring data from the BenMAP library, without additional modeling data.

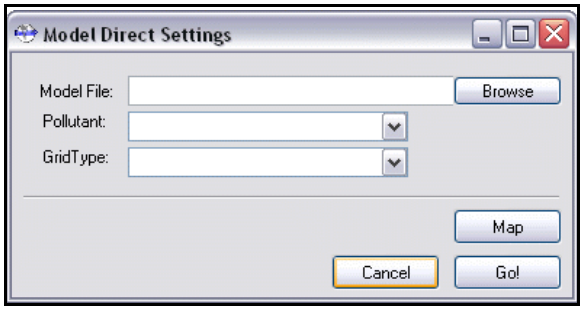
➢ **Monitor and Model Relative.** Choose this option if you wish to use a combination of monitor and model data.

➢ **Monitor Rollback.** Choose this option if you want to reduce all monitor levels by a specified amount.

Select your option and then click **Go!**. BenMAP will direct you through the necessary steps for each option.

4.1 Model Direct

After choosing the *Model Direct* option, you need to specify the location of your data file (**Model File**), the pollutant that the data is modeling (**Pollutant**), and the air quality model from which the data came (**Grid Type**).



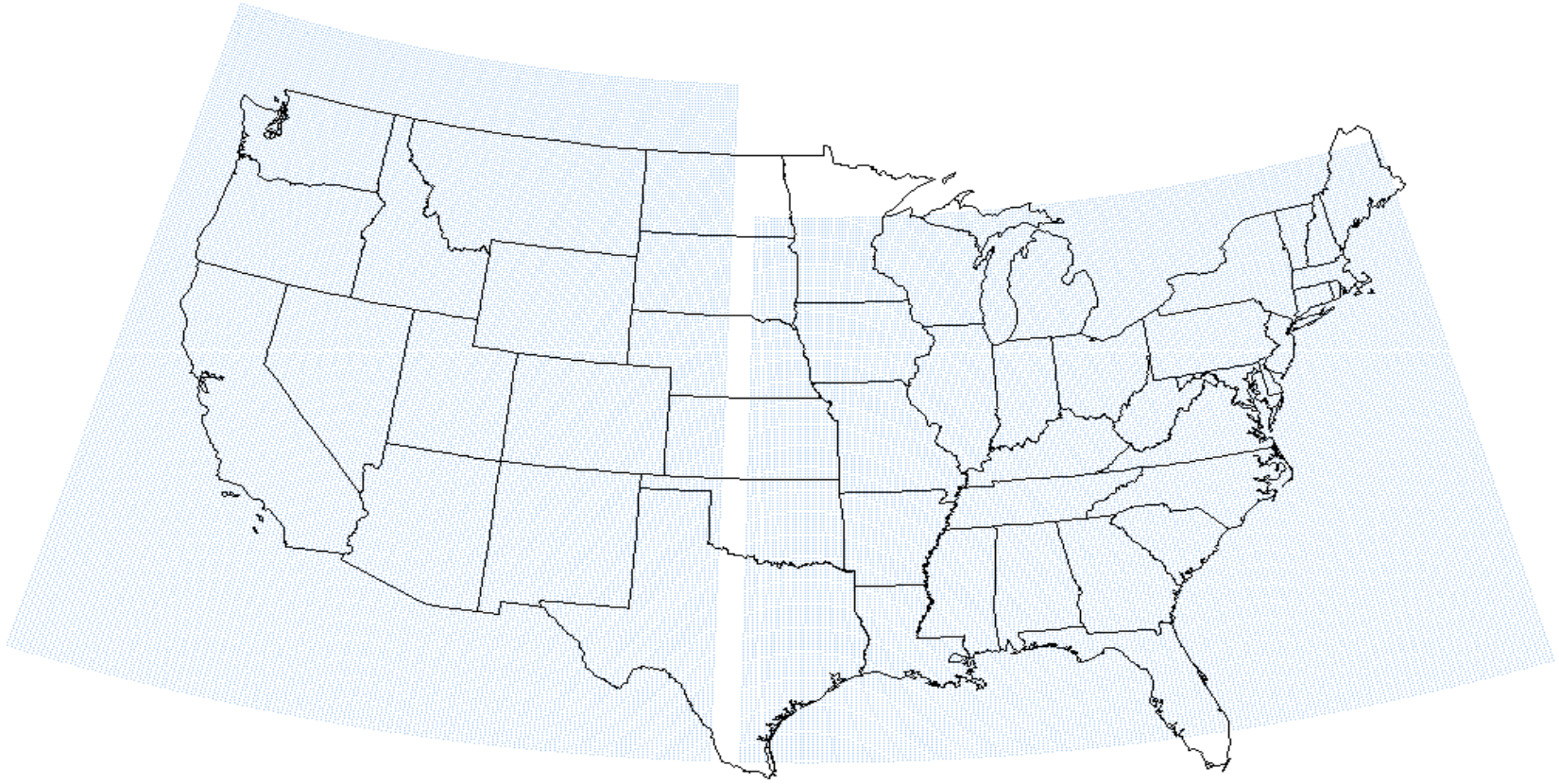
The **Model File** specifies the location of the air quality model results that you want to import. Exhibit 4-1 presents the structure that these files must have, and the pollutants that each grid type is currently designed to accept. (For more information on these models, the EPA website has detailed descriptions of a variety of air quality models: http://www.epa.gov/ttn/scram/.)

Exhibit 4-1. Air Quality Model Data Structure

Model	Pollutant [a]	Modeling Domain and Data File Description [b]
REMSAD [c]	$PM_{2.5}$, PM_{10}, and PMC	Two REMSAD Modeling Domains REMSAD12: has grid cells that are 1/6 of a degree longitude wide and 1/9 of a degree latitude high, or about 12 kilometers by 12 kilometers. The modeling domain extends from longitude -126° to -66° and latitude 24° to 52°, with a total of 90,720 grid cells that completely cover the continental United States. REMSAD36: has grid cells that are 1/2 of a degree longitude wide and 1/3 of a degree latitude high, or about 36 kilometers by 36 kilometers. The modeling domain extends from longitude -126° to -66° and latitude 24° to 52°, with a total of 10,080 grid cells that completely cover the continental U.S. Data File There is a single file for a year of data. The first line has a description of the data, and the second line has a list of the variable names: column, row, and 364 variable names for the days January 1 through December 29. (December 30-31 are missing, and a value is present for February 29th.) Each subsequent line has the actual data in the same order as the variable names. For each column-row combination there is one line of data. Variable names (on the second line) and values (on each subsequent line) are separated by whitespace.
CAMx and UAM-V	Ozone	Modeling Domain CAMx and UAM-V have grid cells that are 1/6 of a degree longitude wide and 1/9 of a degree latitude high, or about 12 kilometers by 12 kilometers. BenMAP assumes a boundary extending from longitude -127° to -67° and latitude 26° to 52°, with a total number of 84,280 grid cells that cover most of the continental United States, with the exception of the southern tips of Florida and Texas. However, modelers often divide the United States into an Eastern and Western modeling domain, with the Eastern modeling domain bounded by longitude -99° to -67° and latitude 26° to 47°, and the Western modeling domain bounded by longitude -127° to -99° and latitude 26° to 52°. At the edge of each of the modeling domains, there is a border of three grid cells with missing data, so some populated areas of the U.S. are not modeled. The actual area typically modeled in the Eastern domain extends from long. -98.5° to -67.5° and lat. 26.33° to 46.67°, and the Western domain extends from long. -126.5° to -99.5° and lat. 26.33° to 51.67° (see Exhibit 4-2). Data File The CAMx and UAM-V modeling data typically comes in multiple files, with each file representing a day of observations. A variable number of modeling days may be used for each domain. For a number of EPA analyses [e.g., \Abt Associates Inc., 2000 #2140], 30 days of modeling were used for the Eastern domain and 19 days for the Western domain. The data files must be combined with the "CAMx / UAM-V Model File Creator", accessible via the **Tools** Menu. The inputs to that tool are one or more “East” files and one or more “West” files. BenMAP expects that the Eastern domain files have data for columns numbered 1-192 and rows numbered 1-189, and the Eastern domain files have data for columns 1-168 and rows 1-234. Each file should have one line of data for each column-row combination, with each line having the column, row, and twenty four observations. Missing observations are denoted with the value -999.0000 (it is important that missing values have exactly this format). Values are separated by whitespace.
CMAQ [c]	$PM_{2.5}$, PM_{10}, and PMC	Modeling Domain The modeling domain for CMAQ covers the entire continental United States. The size of each grid cell is roughly comparable to that of REMSAD36. Data File CMAQ has the same data file structure as REMSAD.

[a] Note that the different Grid types are limited to specific pollutants. Currently, BenMAP can only input REMSAD and CMAQ model data for particulate matter, and CAMx and UAM-V are limited to ozone.

Exhibit 4-2. Eastern and Western Modeling Domains for CAMx and UAM-V



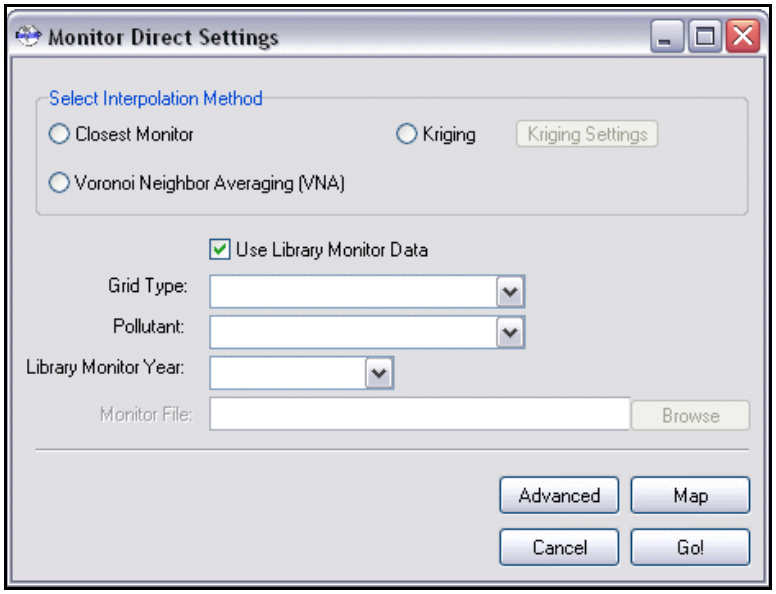
Once your file is identified, choose the **Pollutant**: either *O3 (ozone)* , $PM_{2.5}$, PM_{10}, or *PMC*. You then need to specify the **Grid Type** – either *REMSAD 36km, REMSAD 12km*, *UAM-V / CAMx*, or *CMAQ*. The current version of BenMAP is limited to certain pollutant and model combinations that you are allowed to choose. However, there are no theoretical limitations to these combinations, and over time BenMAP will accept additional combinations. Currently, BenMAP allows you to use REMSAD (36km or 12km) and CMAQ with particulate matter, and CAMx and UAM-V with ozone. Note that the CAMx and UAM-V domains are identical, so BenMAP has grouped them together.

➥ TIP: Carefully name your air quality grids so you can easily recognize them. Include the word base or control and a scenario identifier. For example, if you were analyzing a mobile source emission reduction scenario for 2020 using REMSAD for PM2.5, you might name your grids "Base 2020 mobile REMSAD PM25" and "Control 2020 mobile REMSAD PM25" or something similar.

4.2 Monitor Direct

Using the *Monitor Direct* grid creation option, you create an air quality grid directly from air pollution monitoring data, either ozone, $PM_{2.5}$, PM_{10}, or PMC. At the top of the **Monitor Direct Settings** screen, you are asked to select an interpolation method. The interpolation method is used to move from point-based monitor data to grid cell based air quality data. That is, some grid cells will have many monitors in them, some will have just one, and some will have none. BenMAP uses the interpolation methods to generate representative air quality metric values for each grid cell from monitor data for all of these cases.

BenMAP includes three interpolation methods. The *Closest Monitor* method simply uses the monitor closest to a grid cell's center as its representative value. The *Voronoi Neighbor Averaging* method takes an inverse-distance weighted average of a set of the monitors surrounding a grid cell's center as its representative value. The *Kriging* method takes the weighted average of a set of the monitors surrounding a grid cell's center as its representative value. The kriging method's weights are based on the covariance structure of the surrounding monitors. Each method is described in detail in sections 4.2.1, 4.2.2, and 4.2.3 below. For more detail, see Appendix C.



Below the interpolation options on the **Monitor Direct Settings** screen, you can choose the rest of the options to create your air quality grid. The **Use Library Monitor Data** option, and the other options at the bottom of the screen are described in section 4.2.4. The advanced monitor filtering options that you can access by clicking on the **Advanced** button at the bottom of the screen are described in section 4.5.

4.2.1 Closest Monitor for Monitor Direct

If you choose the *Closest Monitor* option, BenMAP identifies the monitor closest to each grid cell's center, and then assigns that monitor's data to the grid cell.

Closest Monitor interpolation has one advanced interpolation option, which can be modified by clicking on the **Advanced** button at the bottom of the screen and selecting the **Interpolation Options** tab:

➢ **Maximum Neighbor Distance** specifies the maximum distance that a monitor may be from a grid cell (distances are calculated using the grid cell *centroid*). Cells without any monitors within this distance will not be included in the resultant air quality grid. The default setting is infinite (i.e. no limit to the distance between a monitor and a grid cell).

4.2.2 Voronoi Neighbor Averaging (VNA) for Monitor Direct

If you choose the *VNA* option, BenMAP first identifies the set of monitors that "surround" each grid cell's center (these monitors are referred to as the grid cell's *neighbors*), and then BenMAP calculates an inverse-distance weighted average of these neighboring monitors. In this section, we provide some examples of the different ways that BenMAP calculates the average of the neighbor monitors. See Appendix C for an expanded discussion of VNA, including how the VNA algorithm actually chooses the neighbor monitors, as well as the different ways that it may be used.

VNA interpolation has three advanced interpolation options, which can be modified by clicking on the **Advanced** button at the bottom of the screen and selecting the **Interpolation Options** tab:
➢ **Maximum Neighbor Distance** specifies the maximum distance that a monitor may be from a grid cell, and still be included in the set of neighbor monitors used to calculate air pollution exposure at a particular grid cell. The default setting is infinite (i.e., no limit to the distance between a monitor and the grid cell).

➢ **Maximum Relative Distance** specifies the maximum ratio for the distance of each monitor to the distance of the closest monitor. The default setting is infinite.

➢ **Neighbor Scaling Type** specifies whether BenMAP should use inverse-distance weighting for the monitors, or inverse-distance-squared weighting of the monitors. The default setting is inverse-distance weighting.

The following examples illustrate how varying these options affects the final average concentration estimate.

Example 1: Monitor Direct VNA method
Default options

Consider the following example at an hypothetical rural grid cell, where there are relatively few monitors, and where the distance from a monitor to the grid cell can be fairly large. With VNA, BenMAP first identifies the set of "neighbor" monitors for each grid cell. The number of neighbors is usually in the range of about three to eight. In this case, assume that there are five monitors at distances of 25, 50, 100, 200, and 400 miles from the grid cell, with annual $PM_{2.5}$ levels of 8, 13, 12, 18, and 15 $\mu g/m^3$, respectively. BenMAP would calculate an inverse-distance weighted average of the monitor values as follows:

$$PM_{2.5}\ average = \frac{\frac{1}{25}\cdot 8 + \frac{1}{50}\cdot 13 + \frac{1}{100}\cdot 12 + \frac{1}{200}\cdot 18 + \frac{1}{400}\cdot 15}{\frac{1}{25}+\frac{1}{50}+\frac{1}{100}+\frac{1}{200}+\frac{1}{400}} = 10.68$$

Example 2: Monitor Direct VNA method
Maximum Neighbor Distance = 75

Using the same example that we used above, let us say you have specified a **Maximum Neighbor Distance** of 75 miles, and left unchanged the default options (infinite value) for **Maximum Relative Distance**. BenMAP would only consider the first two monitors, and would calculate an inverse-distance weighted average of the monitor values as follows:

$$PM_{2.5}\ average = \frac{\frac{1}{25}\cdot 8 + \frac{1}{50}\cdot 13}{\frac{1}{25}+\frac{1}{50}} = 9.67$$

Example 3: Monitor Direct VNA method
Maximum Relative Distance = 10

Alternatively, if you have specified that the **Maximum Neighbor Distance** is infinite, but the **Maximum Relative Distance** should have a value of, say, 10, then BenMAP would calculate the ratio of the distance for each monitor to distance of the closest monitor. In this case, the ratios would be 1 (=25/25), 2 (=50/25), 4 (=100/25), 8 (=200/25), and 16 (=400/25), and BenMAP would drop the monitor with a ratio of 16. BenMAP would then calculate an inverse-distance weighted average of the monitor values as follows:

$$PM_{2.5}\ average = \frac{\frac{1}{25}\cdot 8 + \frac{1}{50}\cdot 13 + \frac{1}{100}\cdot 12 + \frac{1}{200}\cdot 18}{\frac{1}{25}+\frac{1}{50}+\frac{1}{100}+\frac{1}{200}} = 10.53$$

Example 4: Monitor Direct VNA method
Inverse-distance squared neighbor scaling

In addition, you can specify the an inverse-distance-squared weighting of the monitors. Let us say that you have left unchanged the defaults (infinite values) for **Maximum Neighbor Distance** and **Maximum Relative Distance**, and specified that the **Neighbor Scaling Type** is inverse-distance-squared. BenMAP would then calculate an inverse-distance-squared weighted average of the monitor values as follows:

$$PM_{2.5}\ average = \frac{\frac{1}{625}\cdot 8 + \frac{1}{2{,}500}\cdot 13 + \frac{1}{10{,}000}\cdot 12 + \frac{1}{40{,}000}\cdot 18 + \frac{1}{160{,}000}\cdot 15}{\frac{1}{625} + \frac{1}{2{,}500} + \frac{1}{10{,}000} + \frac{1}{40{,}000} + \frac{1}{160{,}000}} = 9.26$$

Example 5: Monitor Direct VNA method
Maximum Neighbor Distance = 75
Maximum Relative Distance = 10
Inverse-distance squared neighbor scaling

Finally, you could specify changes to all three options: a *Maximum Neighbor Distance* of 75 miles, a *Maximum Relative Distance* of 10, and a *Neighbor Scaling Type* of inverse-distance-squared weighting. BenMAP would then calculate the following average:

$$PM_{2.5}\ average = \frac{\frac{1}{625}\cdot 8 + \frac{1}{2{,}500}\cdot 13}{\frac{1}{625} + \frac{1}{2{,}500}} = 9.00$$

4.2.3 Kriging for Monitor Direct

If you choose the *Kriging* option, BenMAP will present a set of options for you to customize. Please note that some of these options are essential for correct interpolation. See Appendix C for an expanded discussion of kriging.

You can use the default kriging configuration, or customize it by clicking on the **Kriging Settings** button. This lets you set several options.

➢ **Kriging Type** specifies whether BenMAP uses *Ordinary* kriging or *Block* kriging. Ordinary kriging interpolates data using the center point of each grid cell as a reference. Block kriging allows you to overlay the grid cell with a number of support points to which the monitor data will be interpolated. The number of support points for *Block* kriging can be customized using the **Block Kriging Grid** options. *Block* kriging allows for more accurate interpolation of a representative value for the entire grid cell area. The default setting is *Ordinary*.

➢ **Maximum Cell Distance** specifies the maximum distance (in kilometers) for the distance of each monitor to the center of each grid-cell. Monitors exceeding this distance will not be used in the interpolation. The default setting is infinite.

➢ **Covariance Options** allows you to choose the **Nugget Effect** and the **Covariance Function**. The default option is a **Nugget Effect** equal to zero, and no **Covariance Function**. **Please note**: The covariance function must be user specified and must match the empirical covariance structure of the monitor data as determined by the user. Currently BenMAP relies on the user to determine this function externally and supply it when setting paramaters. The nugget effect is the value the covariance function is to assume at a monitor to grid cell distance of 0. This value is also to be determined empirically and is used to model the typically occurring discontinuity or the covariance structure at zero distances.

4.2.4 Other Monitor Direct options

You can also input your own monitor data file, so long as it is in the format that BenMAP recognizes (defined below in Exhibit 4-3). Uncheck the **Use Library Monitor Data** box and then type a path and filename into the **Monitor File** textbox (or you can click the **Browse** button, and choose a file). Note that the file format should be a comma-delimited text file, with all of the variables, and in the same order as they are listed in Exhibit 4-3. If Exhibit 4-3 lists a variable as not "Necessary for BenMAP", it still must have a placeholder comma on each line of the file. Exhibit 4-4 provides a sample of how the data should appear.

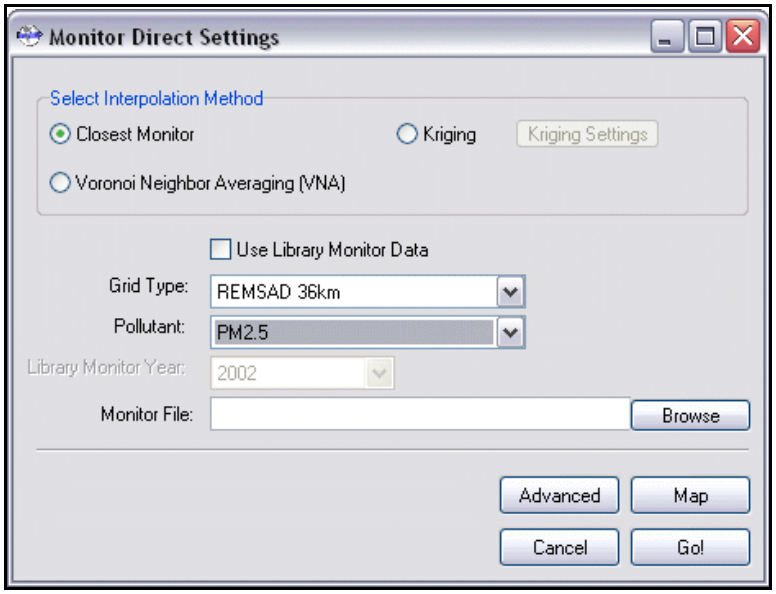


Exhibit 4-3. Monitor Data File Format

Variable [a]	Necessary for BenMAP [b]	Description [c]
year	yes	The *year* is a four-digit variable giving the year of the monitoring data.
monitor ID	yes	The *monitor ID* is a 15 character description of the monitor. It includes a state FIPS code (2 characters), county FIPS code (3 characters), site ID (4 characters), pollutant parameter (5 characters), and POC code (1 character).
latitude	yes	The *latitude* should be in decimal degrees.
longitude	yes	The *longitude* should be in decimal degrees. (Note that the longitude for the United States has a negative sign.)
land use	no	Categorization of the prevalent land use within 1/4 mile of the Monitoring Site.
method	no	The *method* identifies the approach used to collect the monitor data. For example, the Federal Reference Method for $PM_{2.5}$ includes *method* codes 116-120 and 123.
location setting	no	A description of the environmental setting within which the Site is located.
probe location	no	Identification of the location of the sampling Probe.
monitor objective	no	Identification of the reason for measuring air quality by the Monitor.
sample frequency	no	Indicates the scheduled elapsed time period between observations.
sample values	yes	Either 365 daily PM values or 8,760 hourly ozone values. Missing values are indicated with a dot (.).

[a] Monitor data available from the EPA AQS (contact: Virginia Ambrose (ambrose.virginia@epa.gov). Each monitor and method represents a unique set of sample values, and occupies one line of data. BenMAP allows you to choose the desired methods, and then averages the data so that a monitor ID has only a single set of sample values.

[b] The year, monitor ID, latitude, longitude, and sample values are necessary for BenMAP to function. On the other hand, other variables are not strictly necessary, and may have empty values.

[c] Appendix A provides further details on the standard values for the variables, such as land use, location setting, and probe location.

Exhibit 4-4. Sample $PM_{2.5}$ File Format for User-Generated Monitor Text Files

Description	Sample Data[a]
List of variables	year, monitor ID, latitude, longitude, land use, method, location setting, probe location, monitor objective, sample frequency, sample values
Sample daily data with some missing sample values.	2002 , 010270001881011 , 33.281111, -85.802222, agricultural, 116, rural, side of building, highest concentration, 3, 15.2, . , . , . , 18.7, . , . , 12.3, . , . , 22.8, . , . , . , . , . , 10, etc.
Sample daily data with some missing sample values, as well as with missing *land use, location setting, probe location, monitor objective,* and *sample frequency* variables.	2002, 010270001881011, 33.281111, -85.802222, , 116, , , , , 3, 15.2, . , . , . , 18.7, . , . , 12.3, . , . , 22.8, . , . , . , . , . , . , 10, etc.

After choosing the source of the **Monitor Data**, you need to specify the **Pollutant** and the **Grid Type**, and sometimes the **Library Monitor Year**. The choices for **Pollutant** include: *O3 (ozone)*, *$PM_{2.5}$*, *PM_{10}*, and *PMC*. The choices for **Grid Type** include: *REMSAD 36km, REMSAD 12km, UAM-V / CAMx, CMAQ*, and *County*. A **Library Monitor Year** must be selected when using library monitors. Choices are generated automatically, and include all years for which the monitor library contains data for the selected **Pollutant**. For air quality grids created with the **Monitor Direct** option, you may use any pollutant / grid type combination.

Not all of the monitor data for a given **Pollutant** and **Library Monitor Year** is necessarily used by BenMAP in a given run. BenMAP has certain defaults that filter the monitor data, and remove any monitoring data that fails the filtering. For example, BenMAP typically drops any monitoring data with a POC code greater than four. Exhibit 4-5 presents the default filtering options used by BenMAP. See Section 4.5 below for details on how to modify the default settings using the **Advanced** button.

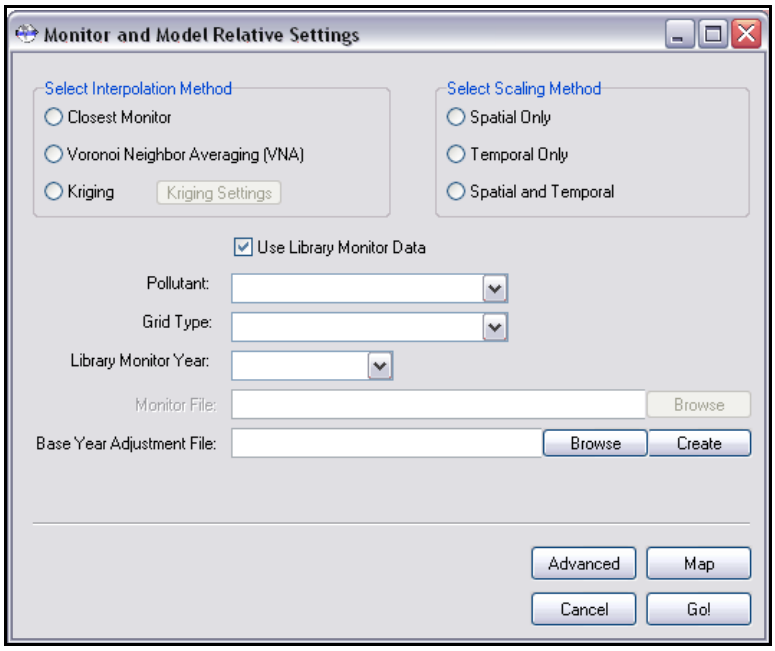
Exhibit 4-5. Default Options Used by BenMAP to Filter Air Quality Monitoring Data

Pollutant	Default Filtering Options
Ozone	Use all method codes and objectives, including missing values. Use a maximum POC code equal to four, and prefer POC code one, followed by two, three, and then four. Use all monitors in the continental U.S. A valid day has at least 9 observations between 8:00 am and 7:59 pm [*Start Hour* = 8 and *End Hour* = 19]; 50 percent of the days must be valid between May 1 and September 30.
$PM_{2.5}$	Use Federal Reference Monitors, which have method codes 116-120 and 123. Use all objectives, including missing values. Use a maximum POC code equal to four, and prefer POC code one, followed by two, three, and then four. Use all monitors in the continental U.S. Each monitor must have a minimum of 11 observations each quarter; use only local $PM_{2.5}$ data.
PM_{10}	Use all method codes and objectives, including missing values. Use a maximum POC code equal to four, and prefer POC code one, followed by two, three, and then four. Use all monitors in the continental U.S. Each monitor must have a minimum of 11 observations each quarter; use both standard and local PM_{10} data, with a preference for local; output data to local.

4.3 Monitor and Model Relative

The *Monitor and Model Relative* option lets you scale interpolated monitor values with model data. As with the *Monitor Direct* option, you can choose *Closest Monitor*, *Voronoi Neighbor Averaging (VNA)*, or *Kriging* interpolation*;* see Section 4.2 above for a discussion of those options. In addition, you can scale the monitor values with three different approaches: *Spatial Only*, *Temporal Only*, and *Spatial and Temporal*. As discussed below, these approaches let you combine the advantage of the actual monitor observations with the information provided by the models.

Monitor and Model Relative air quality grid creation is exactly the same as *Monitor Direct* air quality grid creation (see section 4.2) with the exception of scaling. The concept of scaling is to use modeling data to improve interpolation and/or forecast future air quality trends.



Rather than using raw modeling data, BenMAP uses *Adjustment Factors* created from modeling data. These are loaded as the *Base Year Adjustment File* and, in some cases, the *Future Year Adjustment File*. These files can be created using the *Adjustment Factor Creator*, accessible via the **Tools** menu on BenMAP's main screen or by clicking the **Create** buttons next to the *Adjustment File* input text boxes. The base year file should be created with modeling data which closely reflects the historical conditions of the monitor data to be used. Typically multiple future year files will be used to create multiple air quality grids for use in an analysis. That is, one future year file might represent a future projection of current trends, while another might represent the results of implementing a regulatory program.

Adjustment Factors are discussed in detail in Appendix C. Basically, however, they are representative concentrations sorted low to high per grid cell. Each scaling method uses these adjustment factors to create ratios which are used to scale the concentrations of each neighboring monitor used in the interpolation method selected.

4.3.1 Spatial Scaling

Spatial scaling involves only a *Base Year Adjustment File*, and scales the concentrations of each neighboring monitor by the ratio of the modeled concentration at the grid cell to the modeled concentration at the grid cell containing the monitor. This approach takes into account what the air quality modeling reveals about spatial heterogeneity in pollution levels. For example, if the monitors are in relatively polluted urban areas, and the grid cell is in a relatively unpolluted rural area, then the scaling will result in multiplying the monitor values with ratios less than one, and thus produce lower values at the rural grid cell than would be estimated with interpolation of the unscaled monitor data.

You can create Base Year and Future Year Adjustment Files using the Adjustment Creator under the Tools menu.

Spatial scaling, then, is useful because, while monitors provide invaluable information about historical conditions, there are only a limited number of monitors. Many areas, particularly rural

areas, do not have close monitors. Model data can provide additional information that improves the interpolated concentration estimates, and provides a more accurate picture of air quality.

4.3.2 Temporal Scaling

Temporal scaling involves both a *Base Year Adjustment File* and a *Future Year Adjustment File*, and scales the concentrations of each neighboring monitor by the ratio of the modeled concentration at the grid cell containing the monitor in the future year to the modeled concentration at the grid cell containing the monitor in the base year. This approach takes into account what the air quality modeling reveals about the changes in pollution levels over time at the monitor sites. For example, if the modeling forecasts that in the future, pollution levels will decrease, then the scaling will result in multiplying the monitor values with a ratio less than one, and thus produce lower forecasts at the grid cell than would be gotten with the unscaled monitor data.

Temporal scaling, then, is useful because monitors cannot provide any information about future conditions. Model data can provide this information, which can then be used to project future monitor concentrations.

4.3.3 Spatial and Temporal Scaling

Using both spatial and temporal scaling involves both a *Base Year Adjustment File* and a *Future Year Adjustment File*, and is simply a combination of spatial scaling and temporal scaling, except that the future year data is used for the spatial scaling. That is, it scales the concentrations of each neighboring monitor first by the ratio of the modeled concentration at the grid cell in the future year to the modeled concentration at the grid cell containing the monitor in the future year (spatial scaling), and then by the ratio of the modeled concentration at the grid cell containing the monitor in the future year to the modeled concentration at the grid cell containing the monitor in the base year (temporal scaling). Notice, however, that the two future year concentrations at the grid cell containing the monitor cancel out, allowing the ratio used to be simply the modeled concentration at the grid cell in the future year to the modeled concentration at the grid cell containing the monitor in the base year.

Using both spatial and temporal scaling gives the benefits of both approaches - it both improves the interpolated estimates of air quality, and provides a future forecast of air quality.

4.3.4 Examples

The first example presented is a *Monitor Direct* example, which will provide a foundation for the following *Monitor and Model Relative* examples. Recall that *Monitor and Model Relative* air quality grid creation is exactly the same as *Monitor Direct* grid creation with the exception of scaling. For additional examples with more detail, see Appendix C.

Example 1: Monitor Direct VNA Method
Default options

Consider the example at an hypothetical rural grid cell, from Section 4.2.2, where there are five monitors at a distance of 25, 50, 100, 200, and 400 miles from the grid cell. Further, let us say that the monitors have annual $PM_{2.5}$ levels of 8, 13, 12, 18, and 15 μg/m^3. Without any model-based scaling, BenMAP would calculate an inverse-distance weighted average of the monitor values as follows:

$$PM_{2.5}\ average = \frac{\frac{1}{25}\cdot 8 + \frac{1}{50}\cdot 13 + \frac{1}{100}\cdot 12 + \frac{1}{200}\cdot 18 + \frac{1}{400}\cdot 15}{\frac{1}{25}+\frac{1}{50}+\frac{1}{100}+\frac{1}{200}+\frac{1}{400}} = 10.68$$

Example 2: Monitor and Model Relative VNA Method
Default options
Spatial Only scaling

Assume the same monitors and monitor concentrations as above. Additionally, assume that the grid cell model value in the base-year is 6 μg/m^3, and the model values at the monitors in the base-year are: 10, 14, 11, 17, and 15 μg/m^3. (This modeling suggests that the grid cell is a less-polluted area than the area around the monitors.) Incorporating *Spatial Only* scaling, BenMAP would calculate an inverse-distance weighted average of the monitor values using the same approach as before, with the difference being that the monitor values are scaled with the modeling values:

$$PM_{2.5}\ average = \frac{\frac{1}{25}\cdot\left(8\cdot\frac{6}{10}\right) + \frac{1}{50}\cdot\left(13\cdot\frac{6}{14}\right) + \frac{1}{100}\cdot\left(12\cdot\frac{6}{11}\right) + \frac{1}{200}\cdot\left(18\cdot\frac{6}{17}\right) + \frac{1}{400}\cdot\left(15\cdot\frac{6}{15}\right)}{\frac{1}{25}+\frac{1}{50}+\frac{1}{100}+\frac{1}{200}+\frac{1}{400}} = 5.36$$

Example 3: Monitor and Model Relative Closest Monitor Method
Default options
Spatial Only scaling

Assume the same monitors and monitor concentrations as above. Additionally, assume the same model values as above. The *Closest Monitor* interpolation of these same values using *Spatial Only* scaling would be calculated as follows:

$$PM_{2.5}\ average = 8\cdot\frac{6}{10} = 4.8$$

Example 4: Monitor and Model Relative VNA Method
Default options
Temporal Only scaling

Again, assume the same monitors and monitor concentrations as above. Additionally, assume that the model values at the monitors in the future-year are: 8, 11, 9, 14, and 11 μg/m^3, and that base-

year model values at the monitors are: 10, 14, 11, 17, and 15 μg/m^3. (This modeling suggests that the future-year model values are generally lower.) Incorporating the *Temporal Only* scaling, BenMAP would calculate an inverse-distance weighted average of the monitor values as follows:

$$PM_{2.5}\ average = \frac{\frac{1}{25}\cdot\left(8\cdot\frac{8}{10}\right)+\frac{1}{50}\cdot\left(13\cdot\frac{11}{14}\right)+\frac{1}{100}\cdot\left(12\cdot\frac{9}{11}\right)+\frac{1}{200}\cdot\left(18\cdot\frac{14}{17}\right)+\frac{1}{400}\cdot\left(15\cdot\frac{11}{15}\right)}{\frac{1}{25}+\frac{1}{50}+\frac{1}{100}+\frac{1}{200}+\frac{1}{400}} = 8.52$$

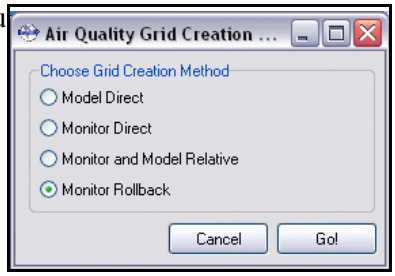
Example 5: Monitor and Model Relative VNA Method
Default options
Spatial and Temporal scaling

Again, assume the same monitors and monitor concentrations as above. Additionally, assume that the future-year model value at the grid cell is 4 μg/m^3, and that base-year model values at the monitors are: 10, 14, 11, 17, and 15 μg/m^3. (This modeling suggests that the future-year model value at the grid cell is significantly lower than the base-year model values at the monitor.) Incorporating the *Spatial and Temporal* scaling, BenMAP would calculate an inverse-distance weighted average of the monitor values as follows:

$$PM_{2.5}\ average = \frac{\frac{1}{25}\cdot\left(8\cdot\frac{4}{10}\right)+\frac{1}{50}\cdot\left(13\cdot\frac{4}{14}\right)+\frac{1}{100}\cdot\left(12\cdot\frac{4}{11}\right)+\frac{1}{200}\cdot\left(18\cdot\frac{4}{17}\right)+\frac{1}{400}\cdot\left(15\cdot\frac{4}{15}\right)}{\frac{1}{25}+\frac{1}{50}+\frac{1}{100}+\frac{1}{200}+\frac{1}{400}} = 3.58$$

4.4 Monitor Rollback

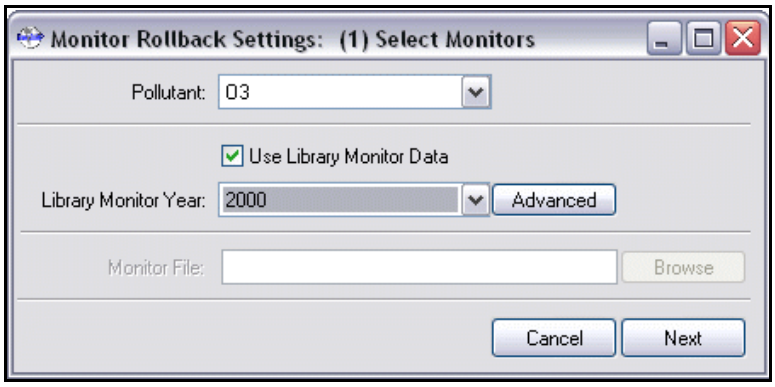
The *Monitor Rollback* option allows you to reduce, or "roll back," monitor data using three methods: percentage rollback, incremental rollback, or rollback to a standard. These approaches let you quickly test what the benefits would be from reducing historical monitor levels. Percentage rollback reduces all monitor observations by the same percentage. Incremental rollback reduces all observations by the same increment. Rollback to a standard lets you choose a standard, and then reduces monitor observations so that they just meet the standard. Note that with each of these methods you can use the same three interpolation algorithms (closest monitor, VNA, and kriging) as you can use with *Monitor Direct* and *Monitor and Model Relative*.



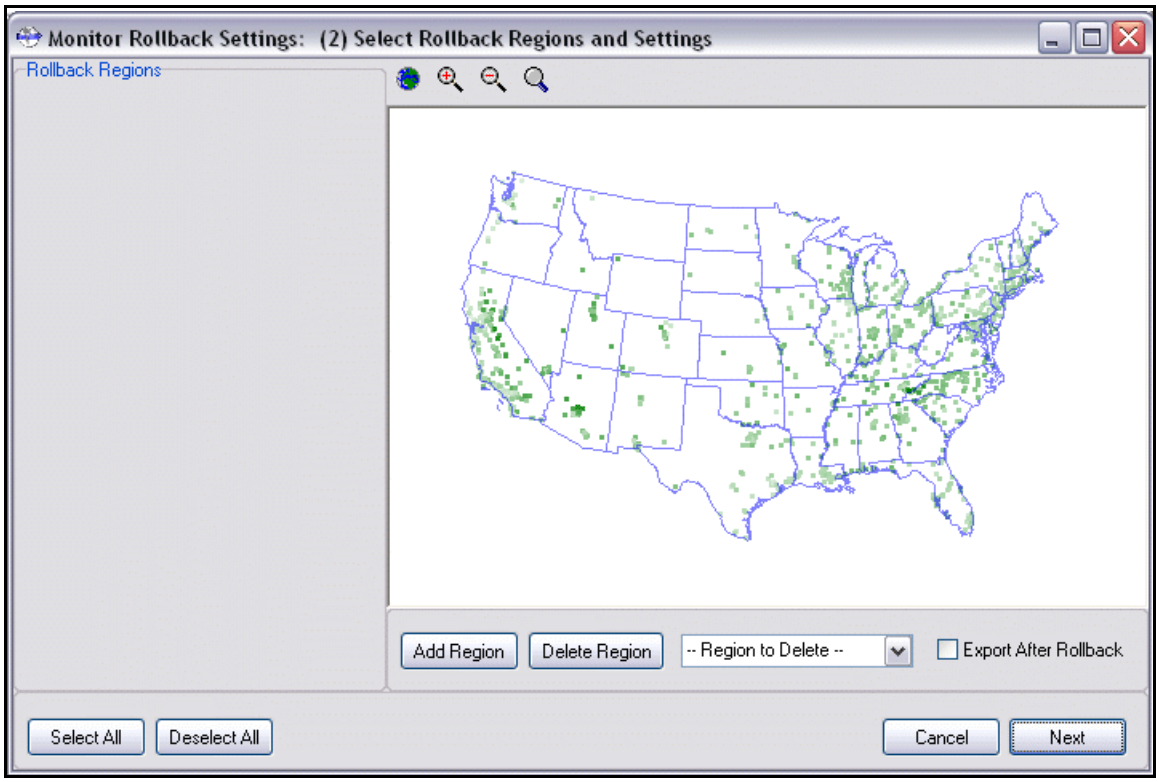
To apply a monitor rollback, first click the **Create Air Quality Grids** button. On the **Air Quality Grid Creation Method** screen, choose *Monitor Rollback*.

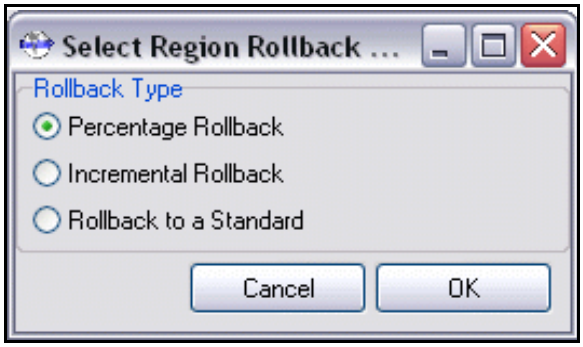
There are three steps to the *Monitor Rollback* method.

1. Monitor Rollback Settings: (1) Select Monitors. Choose the **Pollutant** and the **Year** for the monitor data. If you want to use your own data, then uncheck **Use Library Monitor Data**, and browse for the file that you want to use. The file should have the format specified in Exhibit 4-3. After clicking the **Next** button, BenMAP filters the monitor data using the default parameters for that pollutant, available under the **Advanced** button. (See Section 4.5 below for details on how to modify the default settings using the **Advanced** button.)



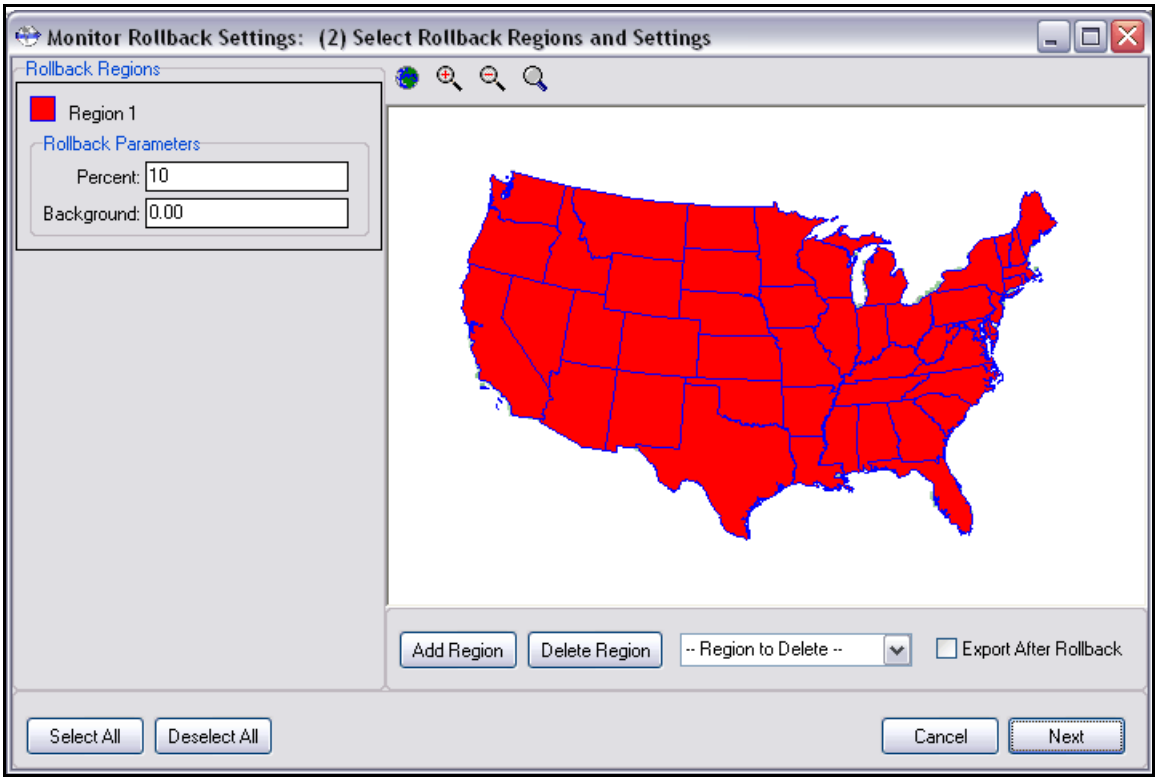
2. Monitor Rollback Settings: Select Rollback Regions and Settings. In this section, you can specify the type of the rollback method(s) that you would like to use, and you can specify the location of the monitors that you want to rollback.





Choosing the **Add Region** button brings up the three rollback methods: *Percentage Rollback*, *Incremental Rollback*, and *Rollback to a Standard*.

After choosing the rollback type, you then need to specify the amount of the rollback and the region to which you want to apply it. In this example, we specified a 10 percent reduction, a background of 0 ppb, and applied it to all monitors in the United States by clicking the **Select All**. So, BenMAP will reduce all observations for all monitors in the United States by 10 percent.



Note that just above the map of the United States there are four buttons, typically seen in mapping programs, that allow you to zoom in and zoom out, and to focus on the particular groups of states that interest you.

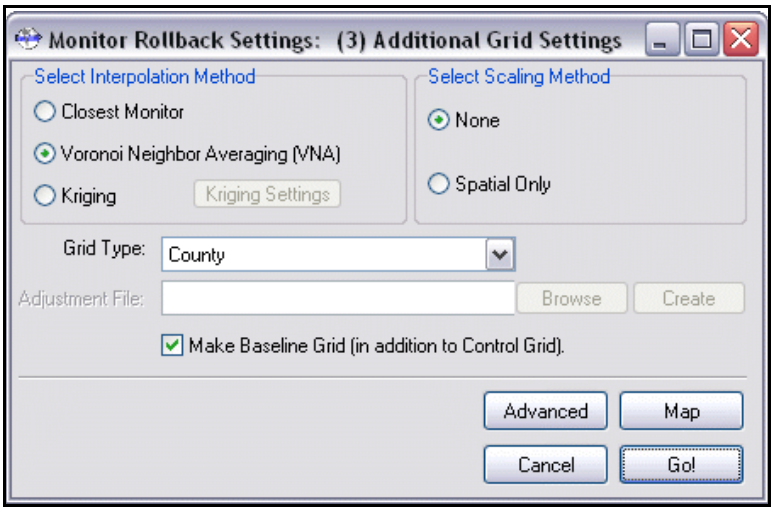
At any time you can change the states that you have selected. This particular example is quite simple, so we will use a more complicated example below. After choosing the states where you

want to rollback monitors, click on the **Next** button. BenMAP will then perform the rollback you specified on the states that you have chosen.

3. Monitor Rollback Settings: Additional Grid Settings.

The third stage is similar to the *Monitor Direct* grid creation method. As in *Monitor Direct*, you need to specify the **Interpolation Method** (*Closest Monitor, VNA,* or *Kriging*) and the *Grid Type*. You may select a *Scaling Method* (*None* or *Spatial Only*). If you choose spatial scaling, it works exactly as described in Section 4.3.1, where the modeling data is used to provide information in those areas that are unmonitored.

By checking the *Make Baseline Grid (in addition to Control Grid)* you may create a baseline grid at the same time as the control grid. The baseline grid uses the same parameters as the control grid, with the exception of the rollback. That is, the baseline uses the same monitor year and filtering, interpolation method, scaling (if any), and the same grid type.

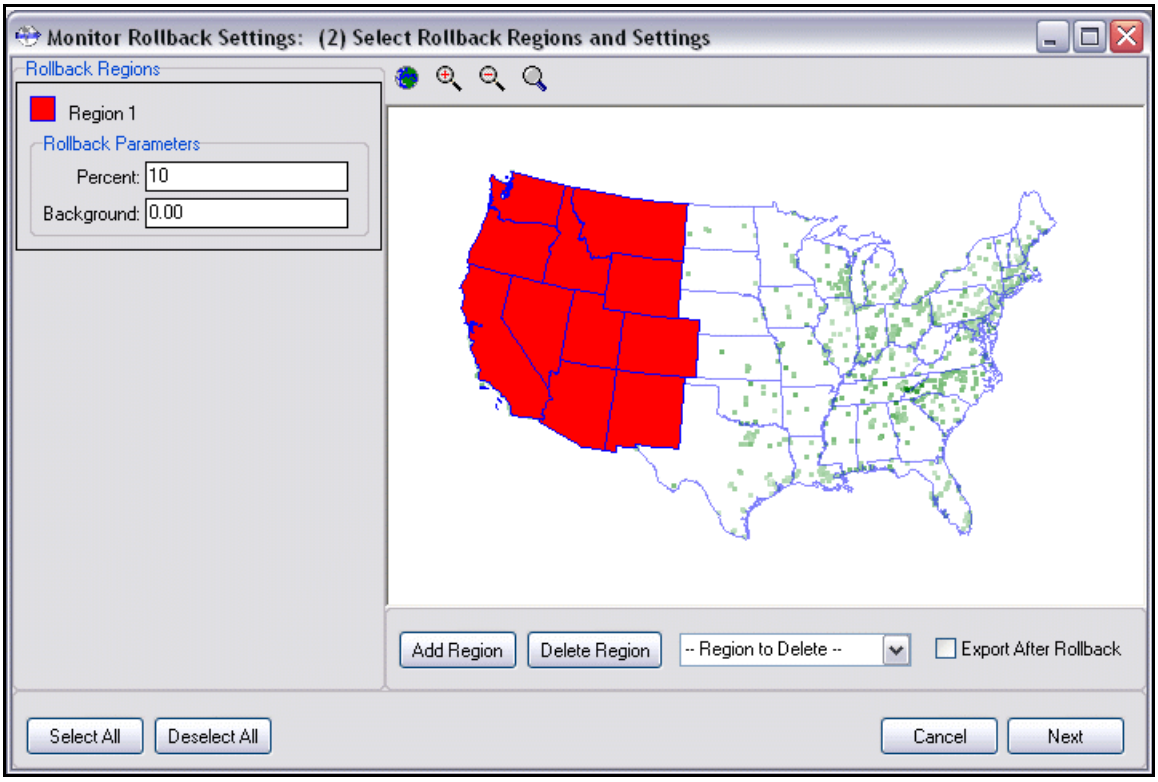


Note that there is an **Advanced** button that lets you select the *Maximum Neighbor Distance*, *Maximum Relative Neighbor Distance*, and *Neighbor Scaling Type*. (These are described further in Section 4.5.) The specific availability of advanced features depends on the interpolation method that you choose. You may also click the **Map** button to view the inputs to the rollback grids that you are creating, as well as to view the grids themselves. Chapter 8 discusses the mapping of grids in more detail.

4.4.1 Example: Combining Three Rollback Approaches in Different Regions

BenMAP allows you to have different rollback approaches in different regions. In this example, we combine the three rollback types: *Percentage Rollback*, *Incremental Rollback*, and

Rollback to a Standard. As in previous example, start by clicking on the **Create Air Quality Grids** button, and choosing *Monitor Rollback.* On the **Monitor Rollback Settings: (1) Select Monitors** screen, select your pollutant (O3 - ozone) and year (2000), and click **Next**. On the next screen, click the **Add Region** button and enter *10* for the **Percent**. In the previous example, we used the **Select All** button to include all states in the rollback region. In this example we want to create three regions instead, so just click on the 11 Western states to add them to the region. The states you have added to the region will turn red, as in the picture below.



To add states with a second type of rollback, click on the **Add Region** button, choose the rollback type, and then click on the states to include in this second region, which BenMAP denotes as "Region 2." In this example, we have chosen an *Incremental Rollback* of 5 and a background of 20, and applied it to the rest of the states West of the Mississippi River.

The map now depicts two rollback regions. We can toggle back and forth between each region by clicking on the legend on the left side of the map. Any states that have not yet been included in a region may be added to an existing region, or we may create one or more regions for these

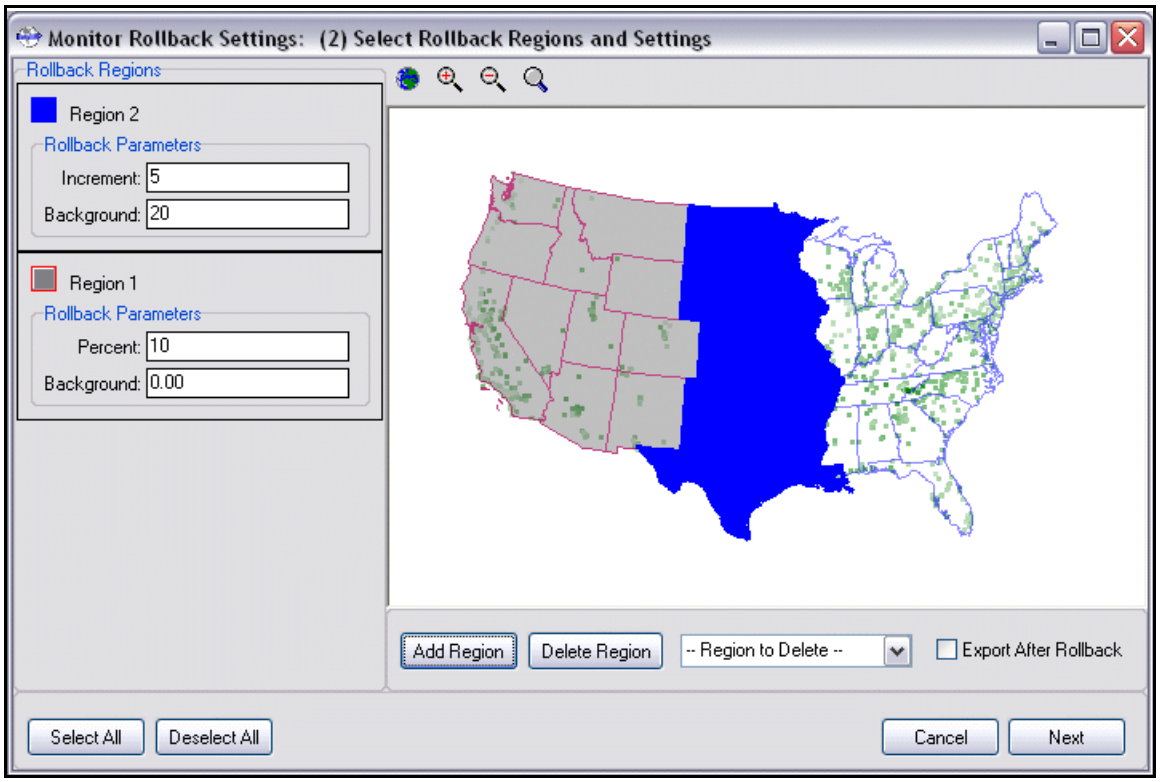
Add a state to a region by clicking on it so that it is highlighted.

Remove a state from a region by clicking on it again.

Add multiple regions using the Add Region button.

Change the displayed region by clicking on the region name in the legend in the left-hand panel.

remaining states. Note that once states have been included in a rollback region, they cannot be included in a different rollback region. In our example, the 11 westernmost states are highlighted in dark gray.



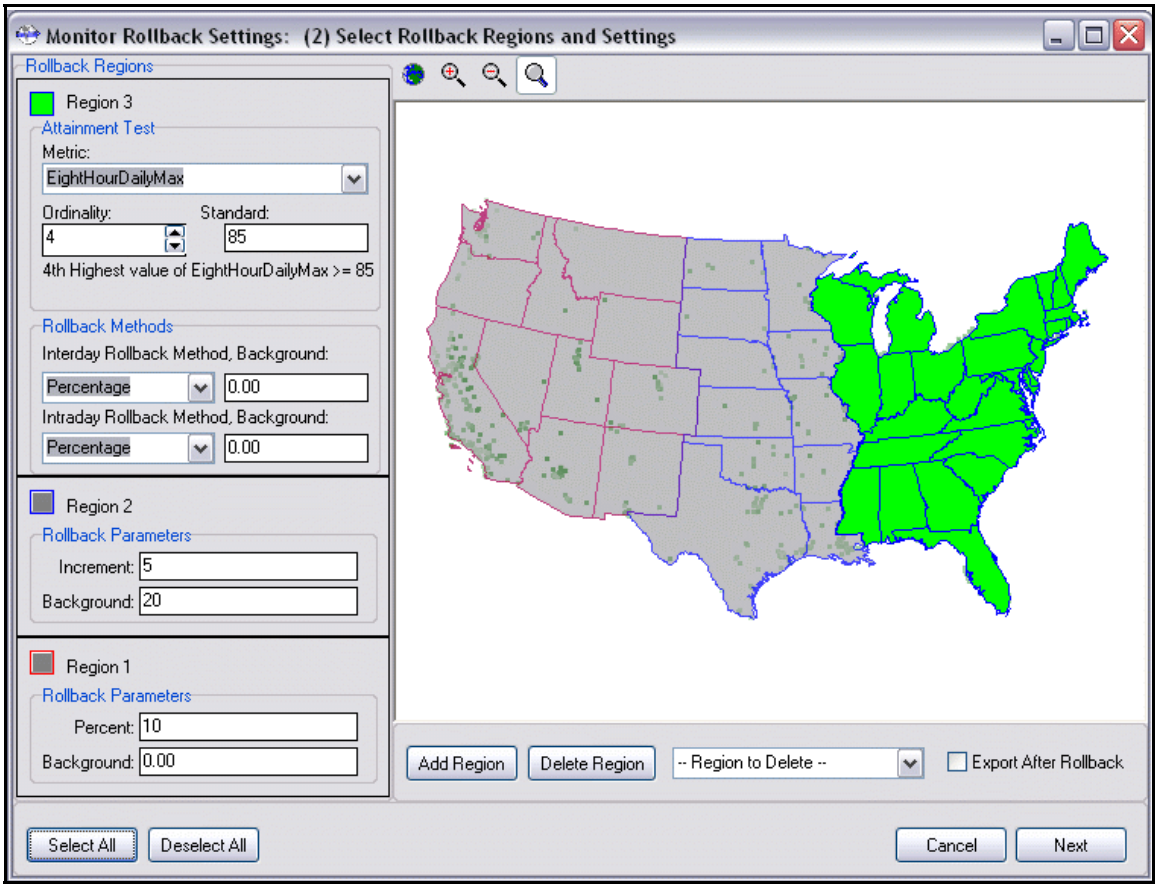
To add a third rollback type covering the states East of the Mississippi River, click again on the **Add Region** button, and then choose the rollback type. However, instead of choosing individually the Eastern states, simply click the **Select All** button. This will select all of the states that are not yet included in a region, and these remaining states will now become Region 3.

Once you have added states to a region, the Select All button will only add the remaining states to the current region. If you want to add all U.S. states, you must first delete the previous region.

In this third region, we have chosen a *Rollback to a Standard*, which involves two groups of parameters - those associated with the **Attainment Test**, which determines whether a monitor is in attainment (meets the standard), and those associated with the **Rollback Methods**, which are used to bring out-of-attainment monitors into attainment.

The **Attainment Test** parameters are **Metric**, **Ordinality**, and **Standard**. A monitor is considered in attainment if the n^{th} highest value of the metric specified by **Metric** is at or below

the value specified by **Standard**, where *n* is the value specified by **Ordinality**. For example, if **Metric** is *EightHourDailyMaximum*, **Ordinality** is four, and **Standard** is eighty-five, a monitor will be considered in attainment if the fourth highest value of the eight-hour daily maximum is at or below eighty-five ppb. In this step BenMAP calculates the metric to be used to determine whether a monitor's values must be rolled back and, if so, how much (e.g., if **Metric** is *EightHourDailyMaximum*, BenMAP calculates the 8-hour daily maximum for each day at each monitor).



The **Rollback Method** parameters are **Interday Rollback Method**, **Interday Background Level**, **Intraday Rollback Method**, and **Intraday Background Level**. These four parameters determine the rollback procedures used to simulate out-of-attainment monitors coming into attainment. The **Interday Rollback Method** and **Background Level** are used to generate target values for the metric specified by the **Attainment Test**. The **Intraday Rollback Method** and **Background Level** are used to adjust hourly observations to meet the target metric values generated in the previous step.

BenMAP provides several types of **Interday Rollback Methods** (*Percentage, Incremental, Quadratic, and Peak* Shaving) and several types of **Intraday Rollback Methods** (*Percentage, Incremental, and* Quadratic). The methods involved for each can be somewhat complicated, so we have included a section in Appendix A that goes through several examples.

4.5 Advanced Monitor options

If the default filtering methods are not desired for a particular analysis, you can use the **Advanced** option, to choose different filtering methods. This option differs by pollutant, so we discuss the **Advanced** option for each pollutant individually below. Access these functions by clicking on the **Advanced** button at the bottom of the **Monitor Direct Settings** screen, the **Monitor and Model Relative Settings** screen, or the **Monitor Rollback Settings: Select Monitors screen**. The **Advanced** option works the same in each case.

There are six filtering options common to all pollutants, and several which are specific to individual pollutants. Once all the options are specified, click the **Go!** button to have BenMAP start the monitor filtering. After filtering is completed, you can export the data (click the **Export** button), map the data (click the **Map** button), or simply click **OK** to confirm the monitor filtering and proceed with the analysis. If you wish to modify your filtering options, simply do so and click the **Go!** button again.

Note that the exported monitor data is in the proper format to be read into BenMAP again in the future (see Section 4.2.4).

Common Filtering Options

The six filtering options common to all pollutants include:

1. Inclusion of individual monitors by ID. If no monitor IDs are entered, all monitors which meet the rest of the selection criteria will be included.

2. Exclusion of individual monitors by ID. No monitors are excluded by default.

3. Geographic filtering. Enter comma-separated state abbreviations to include monitors by state - for example "MD,VA" would only include monitors in Maryland and Virginia. Additionally, enter minimum and maximum latitude and longitude values for monitor inclusion. The default values include all monitors in the continental United States.

4. POC code filtering and preferences. No monitors will be included which have POC codes greater than the specified **Maximum POC**. Additionally, for those monitor IDs which are associated with multiple POC codes, the **POC Preference Order** will be used to determine which set of sample values to use. The default is to use POC codes less than or equal to four, and to prefer lower POC codes - one, followed by two, three, and then four.

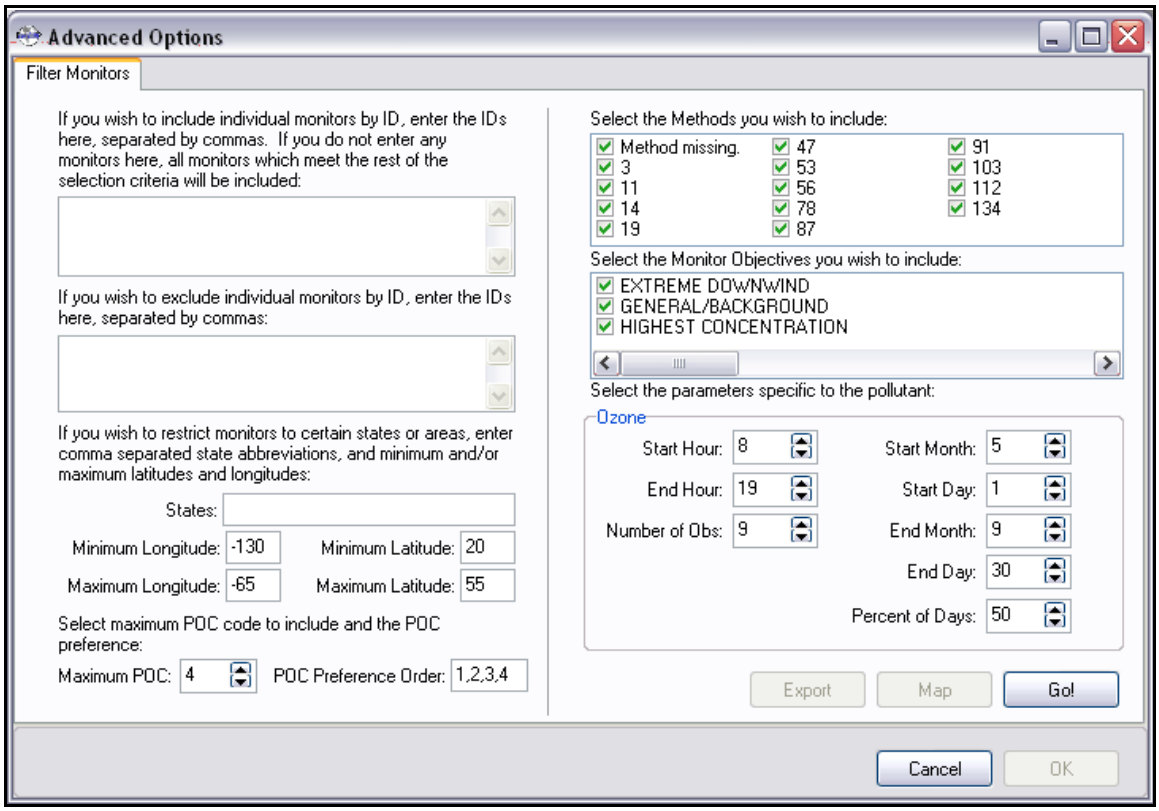
5. Inclusion of method codes. Note that in some cases the method code is missing, and you can exclude these if you want. The default is to use all method codes (including missing method codes) for O3 and PM10, and to use the Federal Reference Monitors for PM2.5 (method codes 116-120 and 123).

6. Inclusion of monitor objectives. Note that in some cases the objective is missing, and you can exclude these if you want. The default is to use all objectives, including missing values.

Advanced Monitor Option for Ozone

There is only one ozone specific filtering option:

1. Identify the completeness criteria for valid monitors. This involves specifying what constitutes a valid day of monitor observations, and the minimum number of valid days over some specified period of time. To specify a valid day, you specify the *Start Hour* and the *End Hour* and the minimum *Number of obs.* that must be available between the *Start Hour* and the *End Hour*. You then specify the beginning of the period of interest with the *Start Month* and *Start Day*, and the end of the period of interest with the *End Month* and *End Day*. Finally, you specify the *Percent of Days* that need to valid in the period of interest. (The **default** option: A valid day has 9 observations between 8:00 am and 7:59 pm [*Start Hour* = 8 and *End Hour* = 19]; 50 percent of the days must be valid between May 1 and September 30.)

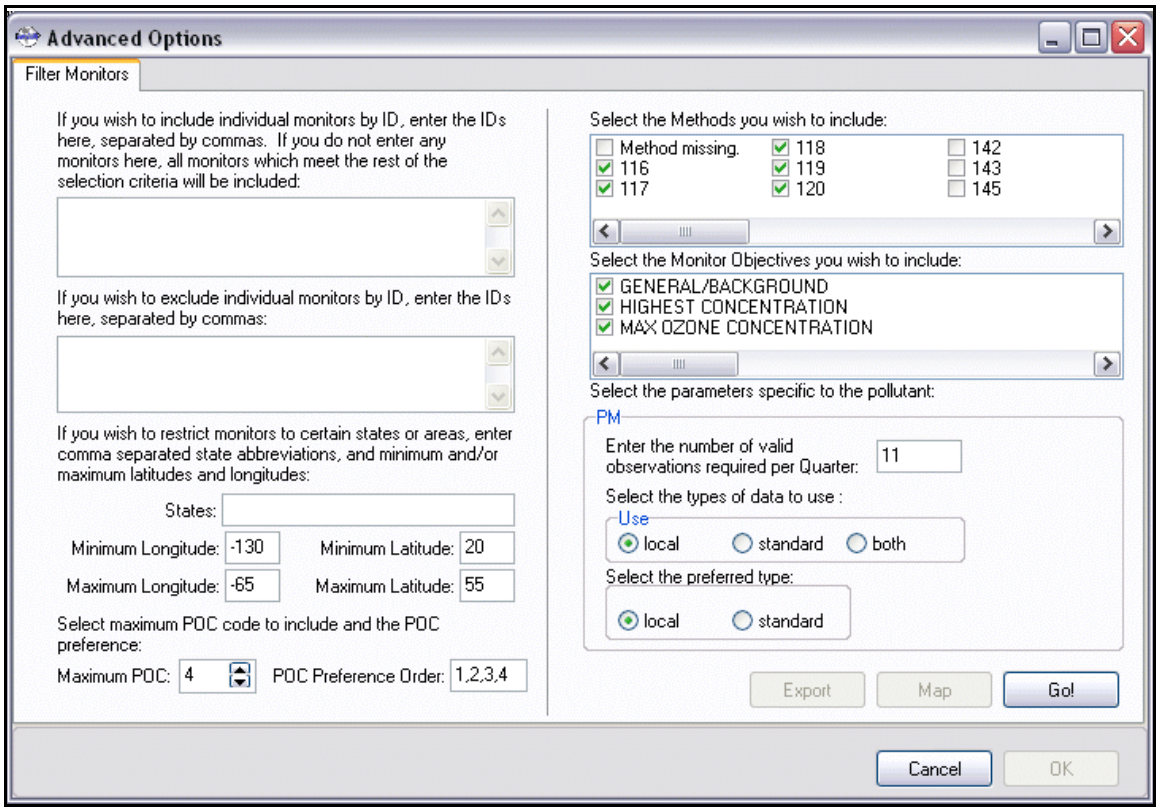


Advanced Monitor Option for $PM_{2.5}$

There are two PM2.5 specific filtering options:

1. Identify the completeness criteria for valid monitors. This involves specifying the minimum number of valid observations per quarter, where the quarters are defined as January-March, April-June, July-September, and October-December. If a monitor does not have the minimum number of observations each quarter, it will be filtered out. The default number of valid observations is eleven.

2. Data type inclusion and preference. PM2.5 monitor data comes in two types - local and standard. Monitors can be included with one type or the other, or with both types. If both types are included, a preference may be given (only one type will be used if both are present) to one or the other. The default is to only use local data.



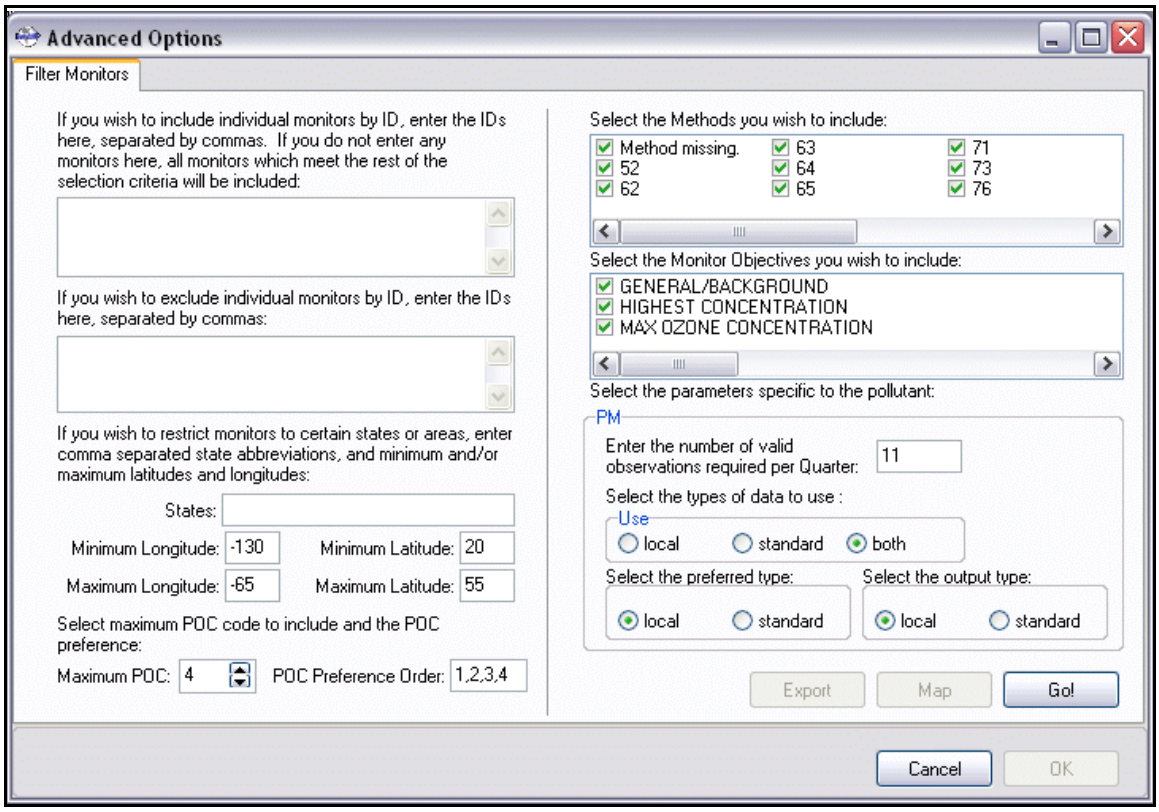
Advanced Monitor Option for PM_{10}

There are three PM10 specific filtering options:

1. Identify the completeness criteria for valid monitors. This involves specifying the minimum number of valid observations per quarter, where the quarters are defined as January-March, April-June, July-September, and October-December. If a monitor does not have the minimum number of observations each quarter, it will be filtered out. The default number of valid observations is eleven.

2. Data type inclusion and preference. PM2.5 monitor data comes in two types - local and standard. Monitors can be included with one type or the other, or with both types. If both types are included, a preference may be given (only one type will be used if both are present) to one or the other. The default is to use both local and standard data, with a preference to local.

3. Output data type. PM2.5 data can be output as either local or standard. The default is to output local data.



4.6 Questions Regarding Creating Air Quality Grids

This section answers some common questions that may arise in the creation of air quality grids.

➢When creating an air quality grid, can I use any combination of Grid Type and Pollutant?

For *Monitor Direct* grids, yes. However, BenMAP currently has limitations regarding the possible combinations of **Grid Type** and **Pollutant** for modeling data, which means that these limitations apply to both *Model Direct* and *Monitor and Model Relative* grids. *REMSAD* and *CMAQ* may be used with $PM_{2.5}$, PM_{10}, and *PMC*, and *UAM-V* and *CAMx* may be used only with *ozone*.

➢Can I export the monitor data used to make my air quality grid using the monitor direct function?

Yes. After choosing the filter options, BenMAP allows you to export a comma-delimited text file of the resulting monitor data. Simply click on the button **Export filtered data to disk**, name the file, and click **Save**.

➢Can I see the weights assigned to each monitor for each grid?

Yes. See the section on the *Neighbor File Creator* in Chapter 9.

➢Can I export my air quality results as a shapefile for use in a GIS program?

Yes. See the section on the *Shapefile Creator* in Chapter 9.

➢Can I use the *Advanced* option to filter my own monitor data?

No. Currently, this option works only for monitor data sets in the BenMAP monitor library.

➢How can I generate a map and export it?

This is explained in Chapter 8.

➢For the *Rollback to a Standard* option, why are there Interday and Intraday rollback options?

This is explained in Appendix A.

CHAPTER 5

Create and Run Configurations

In this chapter...

➢ Use the Create and Run Configuration button to specify various options for calculating incidence results.

➢ Learn about baseline and control scenarios.

➢ Learn the difference between Point Mode and the Latin Hypercube option.

➢ Select Concentration-Response (C-R) functions.

➢ Run and save a configuration.

Chapter Overview

5. Create and Run Configurations

A configuration is a record of the choices you make in estimating the change in adverse health effects between a baseline and control scenario. The choices include the following:

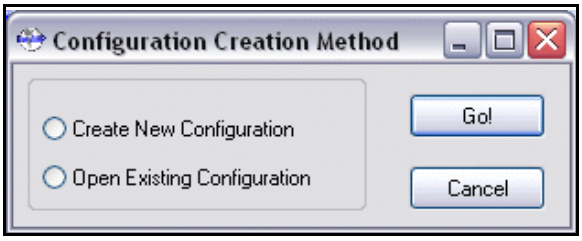
➢ The air quality grids for the baseline and control scenarios;
➢ The year for the analysis;
➢ The threshold for the analysis;
➢ Whether the analysis will focus on a single "point" estimate (**Point Mode**), or a range of results that mirror the variability in the inputs to the C-R functions (**Latin Hypercube Points**); and,
➢ The concentration-response (C-R) functions to be used in estimating adverse health effects.

.

Once these choices are made, they can be saved in a configuration file for future reuse. BenMAP gives you flexibility in the creation, editing, and saving of configuration files. You can open an already existing configuration, and proceed directly to the estimation of incidence. Or, you can create a new one, and proceed with the incidence estimation. In addition, you may save any edits made to existing or new configuration files. BenMAP saves configuration files with a ".cfg" extension. After calculating the change in adverse health effects, BenMAP saves the results in a "configuration results" file with a ".cfgr" extension.

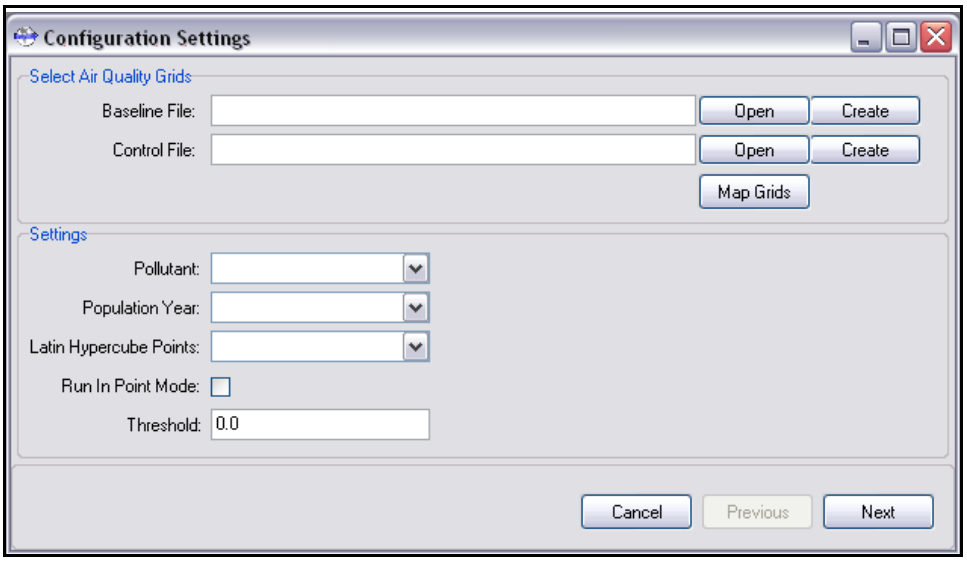
A C-R (Concentration-Response) Function calculates the change in adverse health effects associated with a change in exposure to air pollution. A typical C-R function has inputs specifying the air quality metric and pollutant, population characteristics, and the incidence rate of the health effect. The Incidence Rate gives the average number of adverse health effects per person per year.

After clicking the **Create and Run Configuration** button, you will be asked if you want to create a new configuration, or open an existing one, if you have already created one that you would like to use again. Select the desired option and click **Go!**. Below, we discuss the subsequent steps for each option.



5.1 Create New Configuration

If you choose to create a new configuration, there are two steps. First, fill out the **Configuration Settings** form with the following options: the air quality grids for the **Baseline File** and the **Control File**, the **Pollutant**, **Population Year**, **Threshold**, and whether BenMAP will run in **Point Mode** or use the **Latin Hypercube Points** option. In the second step, specify the C-R functions from lists of EPA Standard C-R Functions and your own C-R Functions.



5.1.1 The Configuration Settings Form

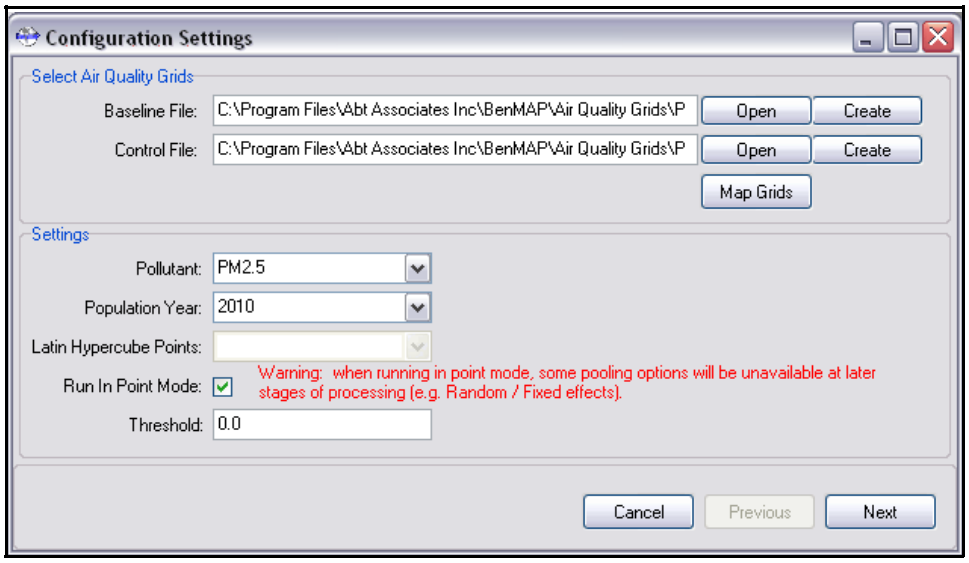
This form opens after you select *Create New Configuration* and click **Go!**. In this form, you first specify the air quality grids for the **Baseline File** and **Control File**. The baseline file contains the air quality metrics for the scenario assumed to occur without any change in policy. The control file specifies the air quality metrics assuming that some type of policy or change has been implemented. The air quality grids should be of the same pollutant, and should also be based on the same grid-type. If you choose a *REMSAD* grid for the baseline file, then a *REMSAD* grid must be used in the control file. Conversely, it would not possible to use a *REMSAD* grid-type in the baseline and a *CMAQ* file in the control file. Similar rules hold for the other grid-types.

You may choose existing air quality grids, by clicking the **Open** button, and selecting an air quality grid, which is designated with an ".aqg" extension. You may also create a new air quality grid by clicking the **New** button. This will take you to window for **Air Quality Grid Creation Method**, where you follow the same steps outlined in Chapter 4 for air quality grid creation.

The **Pollutant** you choose determines the suite of C-R functions available for the configuration. BenMAP has C-R functions for ozone, $PM_{2.5}$, PM_{10}, and PMC. When specifying the **Pollutant**, you typically choose the **Pollutant** known to have been used in the creation of the baseline and control air quality grids. If $PM_{2.5}$ data were used to create the grids, then you would specify the

Pollutant as $PM_{2.5}$. However, occasionally you may want to specify a pollutant that differs from that used in the air quality grid creation. For example, you may be interested in the change in $PM_{2.5}$ air quality, and estimate the impact of this change using PM_{10} functions. BenMAP allows you to do this.

In choosing the **Population Year**, you specify the population data that will be used in the C-R function. The values in the menu for the **Population Year** range between 1990 and 2025. These years correspond to the range covered by the Census data and the population projections built into BenMAP. You may also specify a year beyond 2025, by simply typing in the year desired - in this case, BenMAP will calculate projected populations internally. Appendix B details the sources for the population data used by BenMAP, and how the population projections are generated.



The **Threshold** indicates the minimum value that may be used in either the baseline or control air quality metrics. That is, air quality metrics below the threshold will be replaced with the threshold value. With a threshold of zero, there is no impact on the estimates generated by the C-R functions. However, as the threshold increases, then it will have a progressively larger impact on the incidence estimation. For most analyses, a threshold of zero is appropriate, as there is little evidence suggesting the existence of a threshold. For particulate matter, a review of the recent literature (Rossi et al., 1999; Daniels et al., 2000; Pope, 2000; Schwartz, 2000c) found that PM-related health effects occurred down to the lowest measured levels. Nevertheless, the **Threshold** option allows you to explore the impact of any given threshold on the incidence estimation. This is also useful for scenarios where you might want to know the incidence associated with changes in air quality occurring above a standard.

The **Point Mode** and **Latin Hypercube Points** options allow you to generate an average incidence estimate, or a range of results that mirror the variability in the inputs to the C-R

functions. With the **Point Mode** option, BenMAP uses the mean values of the inputs to the C-R functions, and generates a single "point estimate" of the change in adverse health effects.

With the **Latin Hypercube Points** option, you can generate a number of estimates that mirror the variability in the inputs to the C-R functions. The **Latin Hypercube Points** option allows you to generate specific percentiles along the estimated incidence distribution. For example, if you specify 20 points, then BenMAP will generate estimates of the 2.5th percentile, 7.5th percentile, and so on, up through the 97.5th percentile. The number of points suggested in the drop down menu varies between 10 and 100. In addition, you can simply type in the desired number of points. The greater the number of chosen points, the greater the time needed by BenMAP to process the results. The relationship between the number of points and time needed is essentially linear, so a doubling of the number of points would double the processing time.

If **Point Mode** is chosen, the number of **Latin Hypercube Points** cannot be modified and will be ignored (treated as zero). However, with the **Latin Hypercube Points** option, a point estimate will still be generated. As discussed in Chapter 6 on **Aggregation, Pooling, and Valuation**, by choosing the **Point Mode**, you have fewer pooling options. You cannot conduct fixed/random effects pooling, nor any other procedure that depends on knowing the distribution, or the range of variability of the incidence estimates.

Pooling refers to combining of different sets of data. BenMAP has several pooling methods to choose from.

5.1.2 Selecting C-R Functions for Configuration

The second step in creating a configuration is to select the C-R functions. After filling in both the **Select Air Quality Grids** and **Settings** of the **Configuration Settings** form, click **Next** to select the C-R functions. You may choose from a list of EPA Standard C-R Functions as well as any C-R functions that you may have entered in the **User C-R Functions** window. (For details on how to add your own C-R functions, see Chapter 8.)

An Endpoint group represents a broad class of adverse health effects, such as premature mortality, chronic bronchitis, and hospital admissions.

BenMAP stores the C-R functions in a tree structure that the user can expand and contract. There are eleven endpoint groups, such as *Mortality* and *Hospital Admissions, Respiratory*. Within each endpoint group are sub-groups or endpoints that are specific to each endpoint group, such as *Mortality, Long-Term, All Cause* and *Mortality, Long-Term, Cardiopulmonary*. In cases where an endpoint group has just a single endpoint, they share the same name. Exhibit 5-1 lists the endpoint groups and the associated endpoints for the EPA Standard C-R functions.

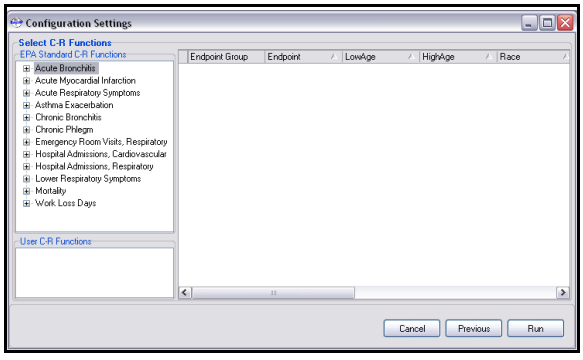
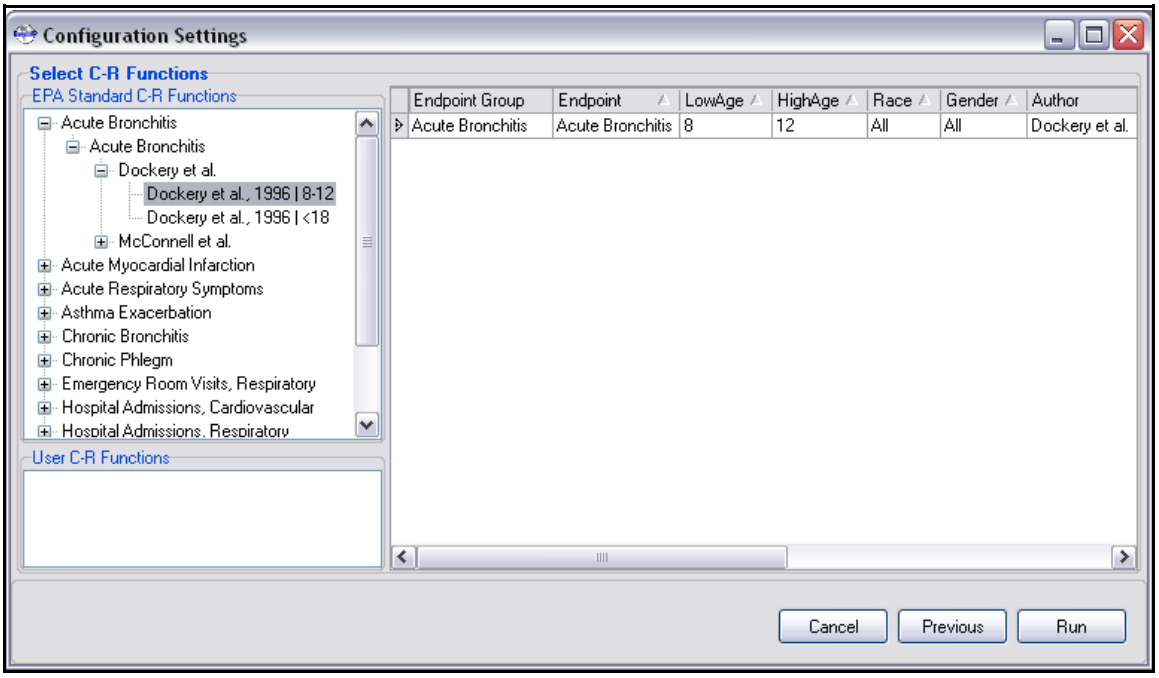


Exhibit 5-1. Classification of C-R Functions Using Endpoint Groups and Endpoints

Endpoint group	Endpoint
Acute Bronchitis	Acute Bronchitis
Acute Myocardial Infarction	Acute Myocardial Infarction, Nonfatal
Acute Respiratory Symptoms	Any of 19 Respiratory Symptoms
	Minor Restricted Activity Days
Asthma Exacerbation	Asthma Exacerbation, Asthma Attacks
	Asthma Exacerbation, Cough
	Asthma Exacerbation, Moderate or Worse
	Asthma Exacerbation, One or More Symptoms
	Asthma Exacerbation, Shortness of Breath
	Asthma Exacerbation, Wheeze
Chronic Asthma	Chronic Asthma
Chronic Bronchitis	Chronic Bronchitis
	Chronic Bronchitis, Reversals
Chronic Phlegm	Chronic Phlegm
Emergency Room Visits, Respiratory	Emergency Room Visits, Asthma
Hospital Admissions, Cardiovascular	HA, All Cardiovascular
	HA, Congestive Heart Failure
	HA, Dysrhythmia
	HA, Ischemic Heart Disease
Hospital Admissions, Respiratory	HA, All Respiratory
	HA, Asthma
	HA, Chronic Lung Disease
	HA, Chronic Lung Disease (less Asthma)
	HA, Pneumonia
Household Soiling Damage	Household Soiling Damage
Lower Respiratory Symptoms	Lower Respiratory Symptoms
Mortality	Mortality, Long-Term, All Cause
	Mortality, Long-Term, Cardiopulmonary
	Mortality, Long-Term, Infant
	Mortality, Long-Term, Lung Cancer
	Mortality, Short-Term, Chronic Lung
	Mortality, Short-Term, Non-Accidental
School Loss Days	School Loss Days, All Cause
	School Loss Days, Illness-Related
	School Loss Days, Respiratory-Related
Upper Respiratory Symptoms	Upper Respiratory Symptoms
Work Loss Days	Work Loss Days
Worker Productivity	Worker Productivity

Note: This exhibit includes the endpoint groups and endpoints for $PM_{2.5}$, PM_{10}, PMC, and ozone included in the EPA Standard C-R functions.

To add studies to your configuration, simply highlight the C-R functions of interest and drag them over to the right hand portion of the screen. You can do this for blocks of C-R functions by dragging over an endpoint group or an endpoint, or drill down and drag individual studies.



If you want to delete some of the C-R functions that you added to your configuration, just highlight the studies of interest and hit the **Delete** key on your keyboard.

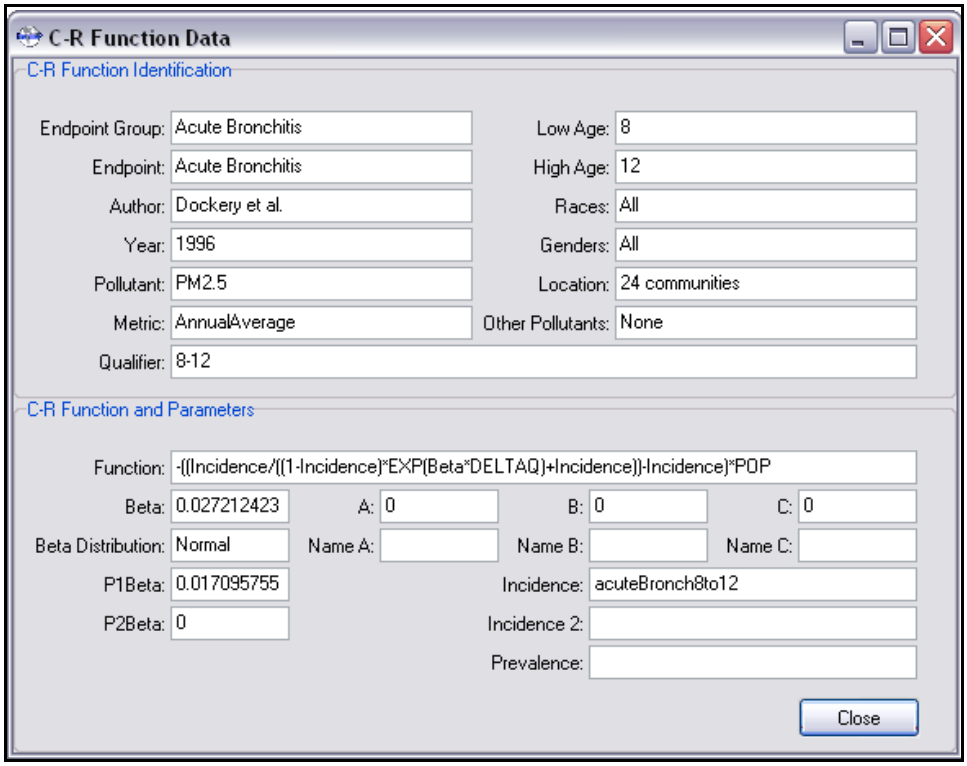
➥ TIP: **Highlighting Blocks of C-R Functions**

To highlight blocks of C-R functions that you have added to your configuration, you may use two approaches: (1) point your cursor at one end of the block, hold down the **Shift** key on your keyboard, and then point your cursor at the C-R function on the other end of the block; or (2) hold down the Ctrl-key on your keyboard, and point your cursor at all of the studies that you want to highlight – this latter approach allows you to highlight discontinuous blocks of C-R functions.

When you drag over a study, BenMAP displays the study's endpoint group, endpoint, low age, high age, race, gender, author, year, qualifier, location, and function. The low age and high age define the bounds of the population of interest, and are inclusive, so bounds of eight and twelve would include all children ages eight, nine, ten, eleven, and twelve. Race and gender also refer to the population of interest for the C-R function. Author and year refer to the epidemiological study on which the C-R function is based, and location gives the location of the study. The qualifier

helps to provide additional identifying information, such as when multiple C-R functions are derived from the same study.

The display of these variables is designed to help identify the study. For additional details, point your cursor at the C-R function and double-click your mouse – this will bring up a display box titled **C-R Function Data**, with all of the details for that particular C-R function. Note that this box is meant for information, and not for editing its contents. To edit the C-R functions that you have added to BenMAP, see Chapter 8. However, as discussed next, you can make some temporary edits right in the configuration screen.

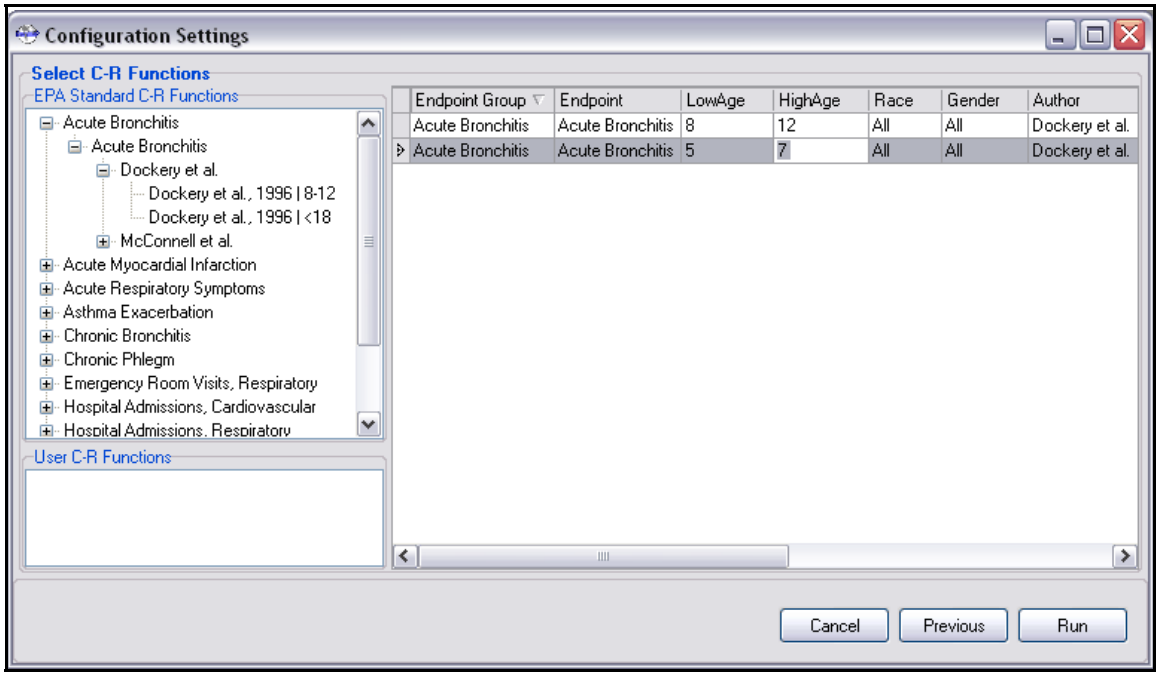


You can drag over the same study multiple times, and then make edits directly to the age, race and gender variables displayed in the configuration screen, in order to be able to calculate the impact of changes in these variables. To edit **LowAge** and **HighAge**, just highlight the appropriate cell and type in the desired age values. Keep in mind that these age represent inclusive age bounds, so if you type in *5* and *7* this will include all children ages five, six, and seven years old. If you want just a single age year, then type the same year in both the **LowAge** and the **HighAge**.

➥ TIP: **Editing C-R Functions**

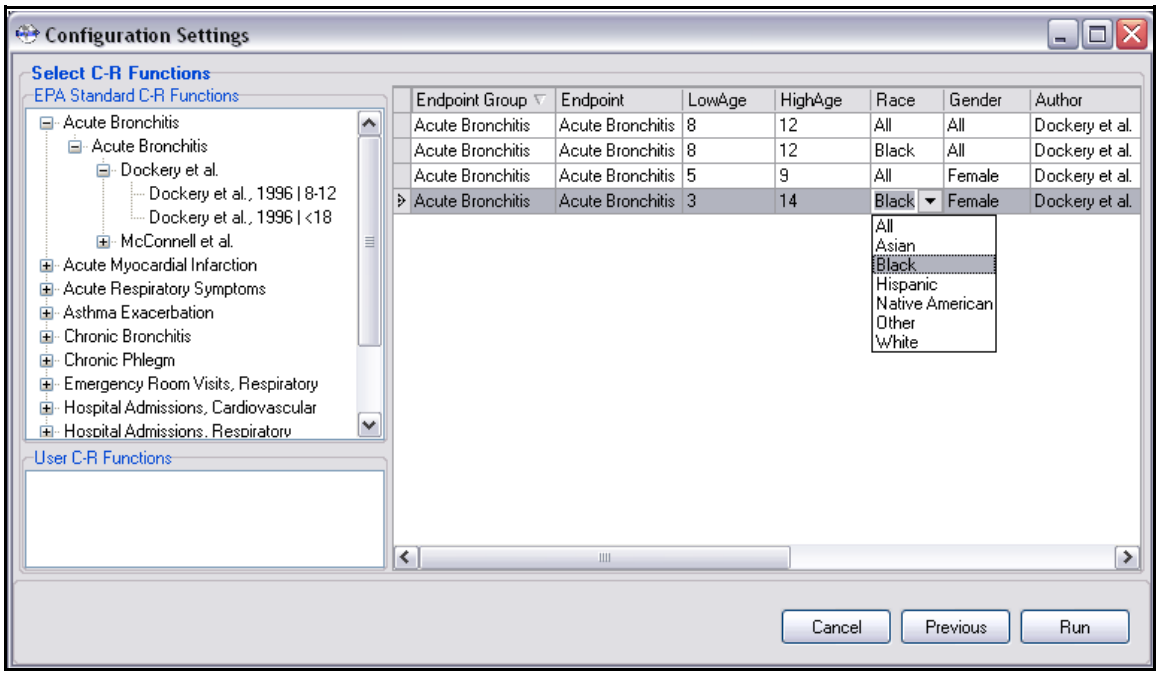
You can edit the age, race, and gender variables in the right hand portion of the **Configuration Settings** form. Any changes you make to the C-R functions will affect the current configuration ONLY. If you save the current figuration, your edited version will appear next time you open it. However, the underlying C-R function will not change.

If you want to edit C-R functions, you can create and then edit C-R functions using the **Data** menu. You cannot edit the EPA Standard functions that come with BenMAP, but you can copy them, then edit and save them as new functions in the User C-R Functions database.



At the same time you can edit the race and gender variables, by clicking on the appropriate cell, and then scrolling through the drop-down menu.

To sort the C-R functions in your configuration by a particular variable, just click the variable name in the display that you want to sort. Click the variable name again to sort it in the opposite direction (ascending or descending order).



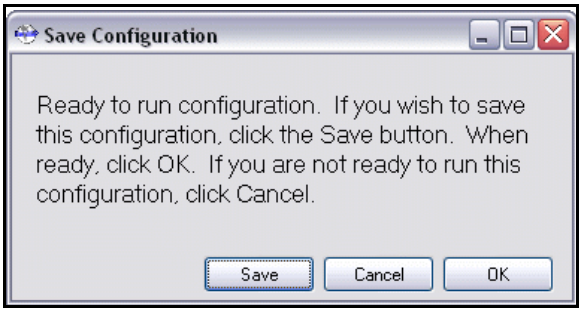
Special Note on Changing Demographic Patterns in the Future

Regarding BenMAP's population data, it is important to note that the projections used in BenMAP allow for demographic changes up through the year 2025. However, the incidence rates, as described in Appendix E, are fixed to the most recently available data, the period around the year 2000. With the population getting older in the future, the incidence rate for broad age groups, such as all individuals over the age of 30, will progressively differ the further out in the future that you estimate adverse health effects. This happens because the incidence rates vary by age, with some rates, such as mortality incidence, increasing significantly with age.

In the case of premature mortality, this can result in a large underestimate of incidence, if you estimate the impacts in the future for a broad age group, such as all individuals over the age of 30. To alleviate this problem, BenMAP provides C-R functions with relatively small age increments, for those C-R functions that have incidence rates that vary by age. For example, instead of just a single C-R function to estimate mortality for all persons 30 and older, BenMAP provides premature mortality C-R functions for small age increments, such as individuals 30-34, 35-44, 45-54, 55-64, and so on. You can then sum the results of the different age-group estimates, to estimate premature mortality for persons ages 30 and over. Appendix E provides additional data on the age groups available.

Nevertheless, this still leaves some residual problem owing to the fact that the incidence rates for individual age groups are changing over time. To the extent that the life expectancy increases over time, we will presumably see lower mortality rates in some age groups. The exact impact of this impact is difficult to predict.

5.1.3 Running the Health Effects Incidence Configuration



To begin the calculation of incidence for the C-R functions in the configuration, you click the **Run** button on the bottom right-hand corner of the **Configuration Form** that has the list of chosen C-R functions. After clicking **Run**, BenMAP allows you to save the configuration, or begin the calculation. If you wish to save the configuration for future use, click **Save** and specify a file with a ".cfg" extension. When ready to generate incidence estimates, click on the OK button. BenMAP then requires that you specify a file in which to save the results, with a ".cfgr" extension.

➥ TIP: **Saving a Configuration**

When you click **Run** at the bottom of the **Configuration Settings** form, you will get a prompt that says, "Ready to run configuration. If you wish to save this configuration, click the Save button. When ready, click OK. If you are not ready to run this configuration, click Cancel." If you click **Save**, you will be saving the **configuration**, i.e., the options and C-R functions that you have selected, so that you can open the configuration and re-run it in the future. Once you are ready to generate incidence estimates, click the **OK** button, and you will be prompted to save another file. This second file is for the **results** of the configuration run, which you will then use for aggregation, pooling and valuation and to generate reports.

5.2 Open Existing Configuration

To use the same settings as a previous BenMAP run, you can choose to open an existing configuration. After clicking the **Create and Run Configuration** button, you simply choose *Open Configuration* and click the **Go!** button. Once an existing configuration is open,, you can do all the things with it that you would do with a newly created configuration - modify settings, add and delete C-R functions, save it to a configuration file, generate results, etc.

5.3 Questions Regarding Configurations

Below are answers to some of the questions that may arise regarding the creation and use of configurations.

➢Can I use air quality grids based on different Grid Types in the baseline and control scenarios?

No. In any given analysis, you need to use the same **Grid Type** in the baseline and control scenarios.

➢Can I use air quality grids of the same Grid Type but based on different Grid Creation Methods?

Yes. In any given analysis, you may use air quality grids made with different methods. Air quality grids made with *Model Direct*, *Monitor Direct*, and *Monitor and Model Relative* may be used interchangeably, if desired. Similarly, air quality grids made with different interpolation methods may be compared. However, it generally is not recommended to create grids with different methods and use them in the same analysis.

➢Can I use air quality grids of the same Grid Type but with different Pollutants?

Yes. Air quality grids based on *REMSAD* and *CMAQ* grid-types, which have $PM_{2.5}$, PM_{10}, and *PMC* data, may in theory be mixed in BenMAP. For this reason, you need to carefully name the air quality grids, to avoid confusion. The air quality grids simply contain data for specific grid-types, with certain air quality metrics, and BenMAP does not keep track whether the air pollution data used in the generation of the air quality grids are $PM_{2.5}$, PM_{10}, or PMC. It is possible for a user to accidentally mix air quality grids of different pollutant types, and still generate incidence estimates. Generally, it is undesirable to mix air quality grids with different pollutants. However, occasionally you may want to specify a pollutant that differs from that used in the air quality grid creation. For example, you may be interested in the change in $PM_{2.5}$ air quality, and estimate the impact of this change using PM_{10} functions.

➢Can I do I an analysis with multiple pollutants?

No. Currently BenMAP analyzes one pollutant at a time.

CHAPTER 6

Aggregation, Pooling, and Valuation

In this chapter...

➢ Use the Aggregation, Pooling, and Valuation button to create a new aggregation, pooling, and valuation (APV) configuration.

➢ Sort and pool incidence results.

➢ Learn the differences between the available pooling methods.

➢ Assign economic values to incidence results.

➢ Aggregate incidence results and valuations.

➢ Save and re-open APV configurations.

Chapter Overview

6. Aggregation, Pooling, and Valuation

Once you have created a configuration results file with incidence results based on your two air quality grids (using the **Create and Run Configuration** button), you can use the **Aggregation, Pooling, and Valuation** button to combine the incidence results and place an economic value on the combined results. You have two options.

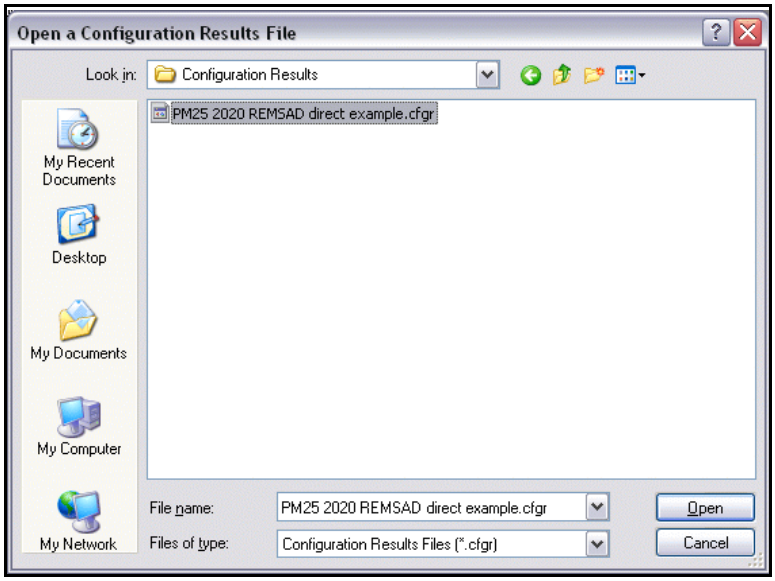
➢**Create a New Configuration for Aggregation, Pooling, and Valuation.** You can create a new type of configuration, termed an Aggregation, Pooling, and Valuation (APV) Configuration. This allows you to specify whether to aggregate incidence results at the county, state or national level, or whether to leave them at the grid cell level. In addition, you can specify how you might want to combine or "pool" the incidence results, using a variety of pooling options. Given the aggregated and pooled incidence results, you then can specify how you might want to value them – typically there are multiple valuations. These valuation results can then be further aggregated, and these aggregated valuation results can be pooled. Having made all of your selections, you may save this APV Configuration file (".apv") for future use, and then proceed to calculating the results, which are stored in an APV Results file (".apvr).

Aggregation refers to the summing of grid cell level results to the county, state or national level. Pooling refers to combining individual incidence results or valuations into groups.

➢**Open Existing Configuration for Aggregation, Pooling, and Valuation.** You can load an existing APV Configuration file, edit the configuration, save it with the same or a different name, and then proceed to calculating the results.

6.1 Creating a New Configuration for Aggregation, Pooling, and Valuation

To start you need to choose a configuration results file that contains incidence estimates at the grid cell level (created by clicking the **Create and Run Configuration** button, see Chapter 5). These incidence results are in files with a *.cfgr extension, and typically stored in the **Configuration Results** folder. Once you open this file, you can begin creating your APV configuration. You will start with selecting and pooling your incidence results, then move on to valuation. These processes are described in detail below. Advanced functions, including aggregation, are described in Section 6.14.



6.1.1 Pooling Incidence Results

After opening the configuration results file, you will find a list of **Available Incidence Results** on the left-hand side of the screen. The results are represented by the C-R Functions from which they were created, and are displayed in a tree-structure with three levels, similar to the tree-structure found in the **Configuration Settings Form** (see section 5.1.2). **Endpoint Groups** occupy the top-most level, followed by **Endpoints**, and then individual C-R Function identifiers. You will also see a pooling window at the right, where you can select pooling options.

There are several steps to pooling your incidence results:

Step 1. Select your default Advanced options

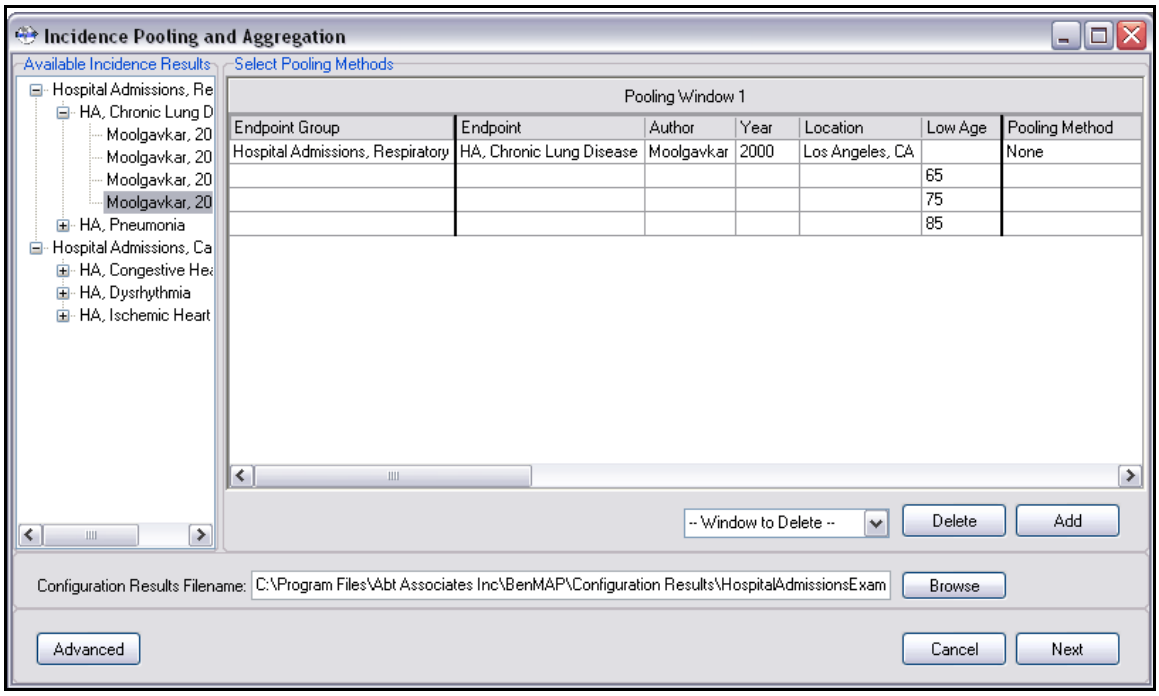
To select your default **Advanced** options, click on the **Advanced** button. There are two values which you will want to set at this point - the **Default Advanced Pooling Method** and the **Default Monte Carlo Iterations**. See Step 6, below, for a detailed discussion of these values. It is important to set them first because these default values will be applied to all incidence results added to the pooling window (see Step 2, below) *after* they are set. Once they are set, click **OK**. If you decide not to change them, click **Cancel**.

Step 2. Add incidence results to the pooling window

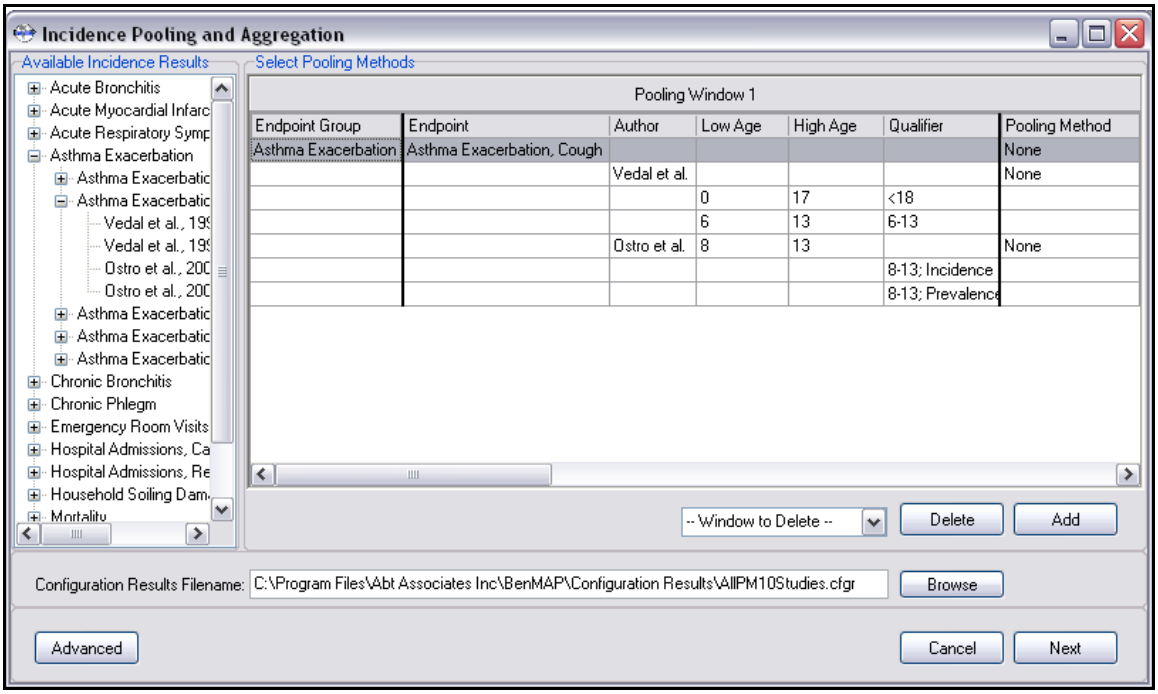
In the **Incidence and Pooling** window, you will see all the incidence results generated from your configuration (see Chapter 5) in the left hand column. You can drag individual incidence results, or groups of results at any level of the tree and drag them over to the pooling window. Note that once you drag a result or group of results into the pooling window it will still be displayed on the

left side. You do not have to drag all of your incidence results over into the pooling window, but note that only those results showing in the pooling window will be included in the pooled incidence or valuation results.

Incidence results are displayed in the pooling window in a tree structure determined by (1) the order of the columns, and (2) the values of the identifying variables of the C-R Functions from which the incidence results were generated (**Endpoint Group**, **Endpoint**, etc. - for a complete list of variables and associated descriptions, see Exhibit 8.1).



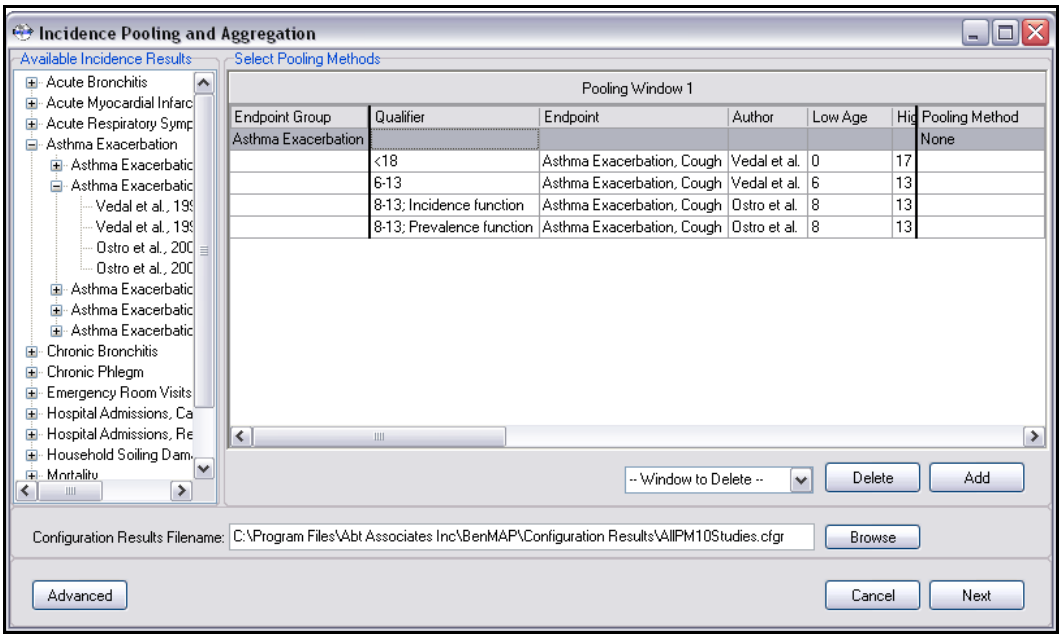
Each line in the pooling window represents a node in the tree structure, with each node representing either an individual incidence result or a collection of incidence results which have common values for their leftmost identifying variables. The tree structure is generated by comparing the leftmost values of the incidence result's identifying variables. High level nodes in the tree are formed when results have common values for identifying variables, and branches in the tree occur when the values differ.



In the above example, four incidence results have been dragged into the pooling window. Each of the four C-R Functions has **Endpoint Group** *Asthma Exacerbation* and **Endpoint** *Asthma Exacerbation, Cough*. Thus, the top line, or root of the tree structure, represents all four incidence results. A branch then occurs in the tree structure, because two studies have **Author** *Vedal et al.* while the other two have **Author** *Ostro et al.* A further branch occurs within *Vedal et al.* when the **LowAge** of the two incidence results differs. Similarly, a branch occurs within *Ostro et al*. when the **Qualifier** of the two incidence results differs. Once a node has only a single incidence result, no further branching can occur.

Step 3. Sort results

After dragging incidence results into the **Pooling Window**, you can rearrange the order of the columns (variables), and thus change the tree structure. To do this, click on a column and hold the button down as you drag it to its new location. Note that **Endpoint Group** is always the first column, and **Pooling Method** is always the last column. All the other columns can be moved. To see how the order of the columns in the pooling window affects the tree structure, consider the following example:



This example uses exactly the same incidence results as the previous example, but with the **Qualifier** column (variable) immediately after the **Endpoint Group** column. Because each result has a unique value for the **Qualifier** variable, the first branch results in four children, which each represent a single incidence result.

➥ TIP: **Sorting Incidence Results Prior to Pooling**

Click and hold your cursor on a column (variable name) in the **Pooling Window**, then drag the column either to the right or to the left. Release you cursor when you have moved the column over the desired location – BenMAP will then rebuild thee tree structure using the newly specified variable order. The **Endpoint Group** is always on the far left-hand side and **Pooling Method** is always on the right, but all of the other variables can be freely moved. Note that whenever you rearrange the tree structure any **Pooling Method** values you may have selected are reset to *None*. It is thus recommended that you sort your results prior to selecting **Pooling Method** values.

Step 4. Select pooling methods

Once the tree structure is set up in the **Pooling Window**, you are ready to select your pooling methods. Essentially each pooling method involves a different method of combining input incidence results to generate new incidence results. Results can be pooled any time a branch occurs in the tree structure - that is, any time two or more results share common values for their leftmost variables. BenMAP helps you to identify these spots by inserting a value of *None* in the **Pooling Method** column at each spot where pooling is possible.

Exhibit 6-1 summarizes the different types of pooling approaches, and Appendix I provides a detailed discussion of the approaches.

Exhibit 6-1. Pooling Approaches for Incidence and Valuation Results

Pooling Approach	Description of Pooling Approach [a]	Availability	
		Point Mode	*Latin Hypercube*
None	No pooling performed.	✔	✔
Sum (Dependent)	Results are summed assuming they are perfectly correlated. In Point Mode, this is just a simple sum. In Latin Hypercube moed, BenMAP chooses the first point from each result in the pooling and does a sipmle sum to generate the first point in the pooled result, and so on for all of the points in the distribution of results.	✔	✔
Sum (Independent)	Results are summed assuming that they are independent. A Monte Carlo simulation is used. At each iteration, a random point is chosen from the Latin Hypercube of each result, and the sum of these values is put in a holding container. After some number of iterations, the holding container is sorted low to high and binned down to the appropriate number of Latin Hypercube points.		✔
Subtraction (Dependent)	Results are subtracted assuming they are perfectly correlated. All subsequent results are subtracted from the first result (the highest result in the display - to reorder results, simply click and hold a result and then drag it to its new position). In Point Mode, this is a simple subtraction. In Latin Hypercube mode, BenMAP chooses the first point from each result in the pooling and does a simple subtraction to generate the first point in the pooled result, and so on for all of the points in the distribution of results.	✔	✔
Subtraction (Independent)	Results are subtracted assuming that they are independent. A Monte Carlo simulation is used. At each iteration, a random point is chosen from the Latin Hypercube of the first result, and then random points are chosen from the Latin Hypercube of each subsequent result and subtracted from the first. The result is put into a holding container. After some number of iterations, the holding container is sorted low to high and binned down to the appropriate number of Latin Hypercube points.		✔
Subjective Weights	Weights are specified by the user (see Step 5, below). In Point Mode, the new result is generated by a simple weighted sum of the input results. In Latin Hypercube mode, the results are combined using the user specified weights with the "Round Weights to Two Digits" *Advanced Pooling Method*. See Step 6 below for details.	✔	✔
Fixed Effects	*Pooling weights are generated automatically based on the inverse* variance of each input result, with the weights normalized to sum to one. Results with a larger absolute variance get smaller weights. Results are then combined according to the chosen *Advanced Pooling Method*.		✔
Random / Fixed Effects	BenMAP first tests if random weights should be used. If not, BenMAP uses fixed effects weights. If yes, the weights take into account both the variance within each set of results and the variance between sets of results. Results are then combined according to the chosen *Advanced Pooling Method*.		✔

[a] Appendix I details the different pooling approaches.

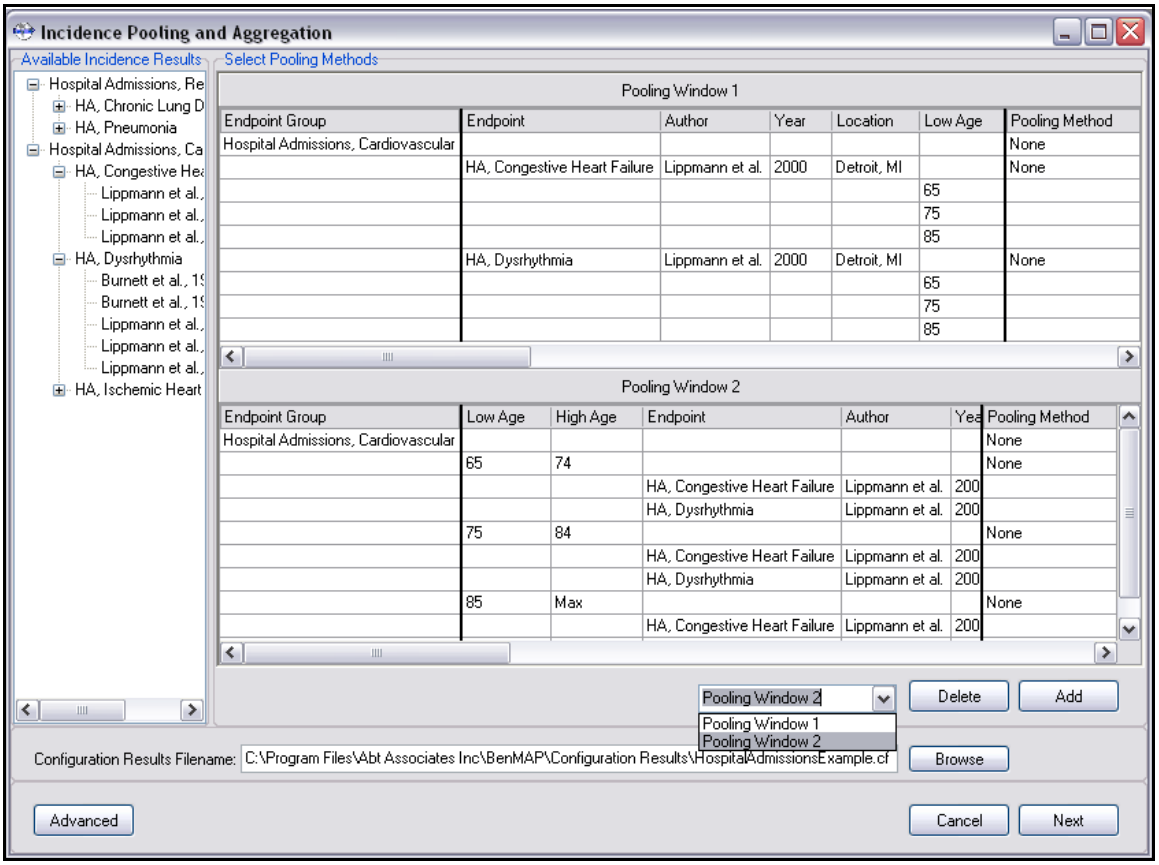
Note that some pooling methods are only available in Latin Hypercube mode. This is because these pooling methods attempt to combine distributions of results into new distributions, and no distributional information is available in Point Mode. The **Pooling Method** column will thus have different values in its drop down list depending on the mode used to generate the incidence results being pooled.

Latin Hypercube is a series of points generated by using specified percentiles in a given distribution, such as that of a C-R coefficient. It is a short-cut method designed to represent a distribution, while at the same time saving on computation time. Using the **Point Mode** means that BenMAP will use the mean value of the coefficient in the C-R function.

Step 5. Create additional pooling windows if needed

Within a given pooling window, you can have only one ordering of the columns (variables). As we have seen, however, the ordering of the columns determines the structure of the tree used to pool results. It may thus sometimes be necessary for analyses to have multiple tree structures to handle the various pooling trees they require. To facilitate this, BenMAP allows additional pooling windows to be added and deleted. To open a new pooling window, simply click on the **Add** button. You may do this as many times as needed to accommodate different sort orders. You can add the same incidence results to as many different pooling windows as you like.

As needed you can also delete a pooling window by using the "-- Window to Delete --" drop-down menu to identify the pooling window, and then hitting the **Delete** button.



Step 6. Advanced Pooling Methods, Monte Carlo Iterations

Some pooling methods have advanced options which should be set at this point. To set them, double-click the **Pooling Method** column on the row for which you wish to select advanced options. The advanced options available depend on the particular pooling method.

None, Sum (Dependent), Subtraction (Dependent). These pooling methods have no advanced options associated with them.

Sum (Independent), Subtraction (Independent). These pooling methods have one advanced option associated with them, **Monte Carlo Iterations**. As discussed in Exhibit 6-1 above and in Appendix I, these two pooling methods involve a Monte Carlo simulation. This advanced option specifies the number of iterations this simulation should go through in generating results. Its initial value is set by the **Default Monte Carlo Iterations** value from the **APV Configuration Advanced Settings** window (see Step 1, above).

Fixed Effects, Random / Fixed Effects. These pooling methods have two advanced options associated with them - **Advanced Pooling Method**, and **Monte Carlo Iterations**. **Advanced Pooling Method** can take on three different values, which are discussed next. Its initial value is set by the **Default Advanced Pooling Method** value from the **APV Configuration Advanced Settings** window (see Step 1, above).

Round weights to two digits. BenMAP rounds each weight to two digits (e.g. 0.73), and then multiplies these weights by 100 to get two digit integers. Each entire distribution (set of Latin Hypercube points) is then put into a holding container an integral number of times, according to its integral weight. This holding container is then sorted low to high and binned down to the appropriate number of Latin Hypercube points.

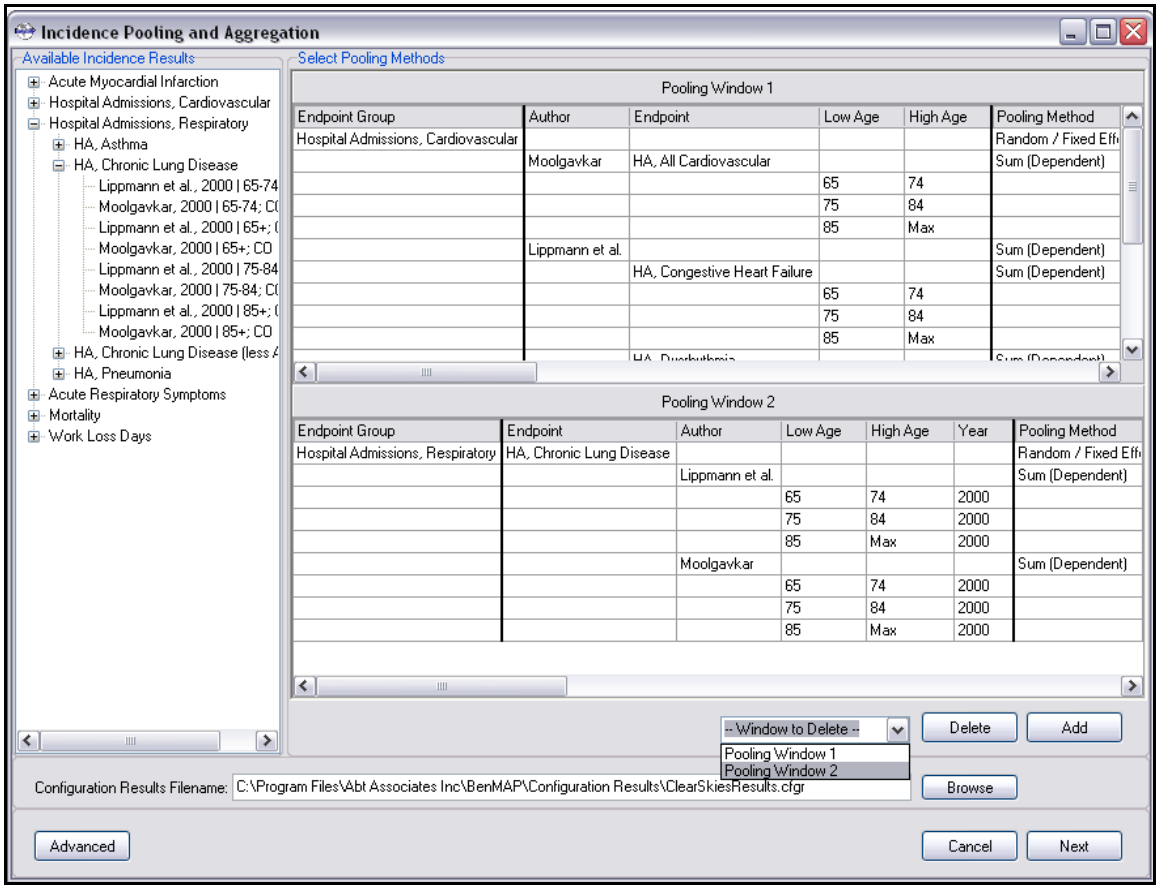
Round weights to three digits BenMAP rounds each weight to three digits (e.g. 0.732), and then multiplies these weights by 1000 to get three digit integers. Each entire distribution (set of Latin Hypercube points) is then put into a holding container an integral number of times, according to its integral weight. This holding container is then sorted low to high and binned down to the appropriate number of Latin Hypercube points.

Use exact weights for Monte Carlo. BenMAP uses exact weights and a Monte Carlo simulation. On each iteration of the procedure, a particular result is chosen with a probability equal to its weight. Once a result is chosen, one of its Latin Hypercube points is chosen at random and put into a holding container. This is done some number of times (see **Monte Carlo Iterations**, below), and the holding container is then sorted low to high and binned down to the appropriate number of Latin Hypercube points.

Monte Carlo Iterations: This drop down list is only enabled when *Use exact weights for Monte Carlo* is selected as the **Advanced Pooling Method**. It specifies the number of iterations the Monte Carlo simulation should be run (see above). Its initial

value is set by the **Default Monte Carlo Iterations** value from the **APV Advanced Settings** window (see Step 1, above).

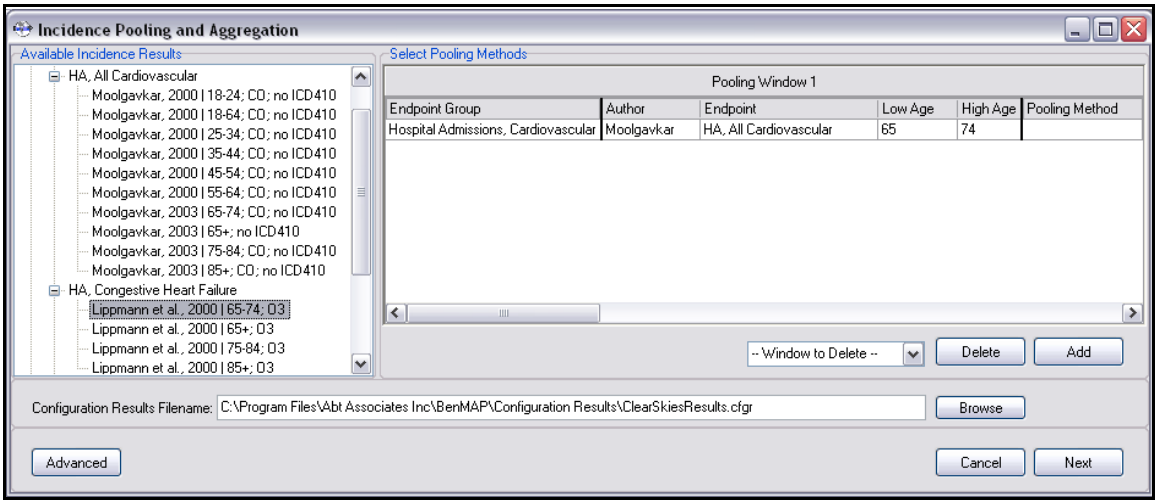
Subjective Weights. Subjective weights pooling has no advanced options associated with it - the advanced pooling method is always *Round weights to two digits* (see above). Double clicking, however, brings up a dialog through which the weights for each result can be specified. Alternatively, if you haven't set the weights before clicking the **Next** button the **Select Subjective Weights** window will come up automatically.



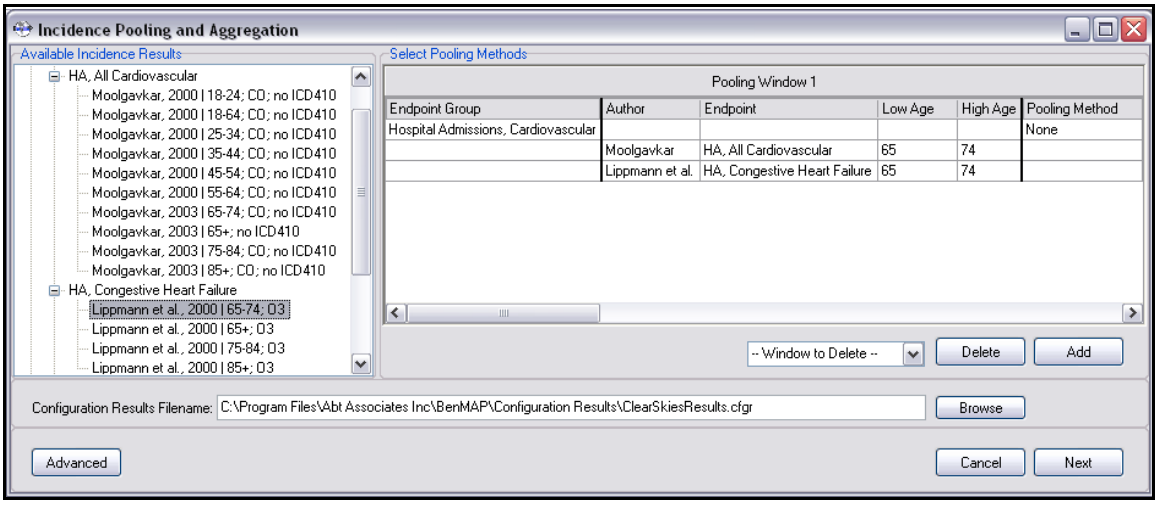
Note that the weights you enter need not add up to one - BenMAP will normalize them internally. Also note that BenMAP initializes all the weights to $1/n$, where n is the number of results being pooled.

Example: Simple sorting and pooling of incidence results

If you add a single incidence result to the right-hand window, you will see just one line, and therefore no opportunities to pool. This shown in the example below.

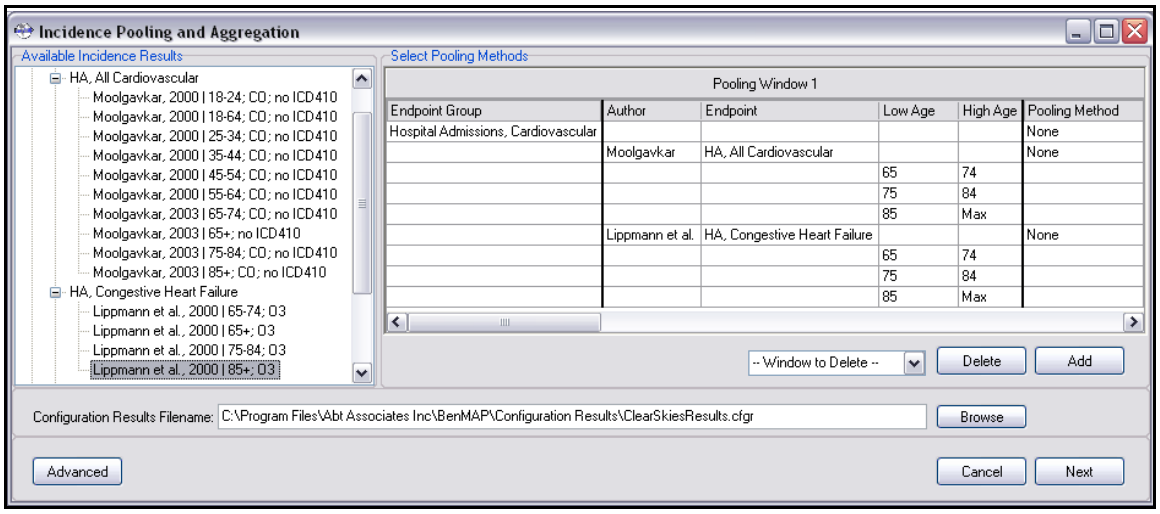


If you add a second incidence result to the window whose C-R Function has the same **Endpoint Group**, but a different **Author** and **Endpoint**, you will then have a tree with two items in it. The tree branches at the point where the two C-R Functions vary - at the **Author** column.



Note that a pooling method can now be selected for the two incidence results, since a branch has appeared. If we desired to pool these two incidence results, we would end up with a pooled result representing two *Hospital Admissions (HA), Cardiovascular* incidence results.

If you now add four more incidence results to the window whose C-R Functions have the same **Endpoint Group**, you will see the following:



Now you have many pooling options. Setting aside the issue of which pooling method to choose, there are eight different pooling options at this point, since we have three places where we can choose to pool or not to pool.

If you choose to pool at the two spots corresponding to **Endpoints** (*HA, All Cardiovascular* and *HA, Congestive Heart Failure*) you would end up with two pooled results instead of six individual incidence results.

If you choose to pool at the first place where the **Pooling Method** field says, *None*, the spot corresponding to the **Endpoint Group** (Hospital Admissions), you will end up with a single result representing all six of the original incidence results. However, you can also pool at the other spots as well, and thereby impact the final pooled result:

If you pool at all three spots:

➢First, the three Moolgavkar results are pooled to a give a single *HA, All Cardiovascular* result.

➢Next, the three Lippmann results are pooled to give a single *HA, Congestive Heart Failure* result.

➢Finally, the two results generated in the previous steps are pooled to give a single *Hospital Admissions, Cardiovascular* result.

If you pool at *Hospital Admissions, Cardiovascular* but not at the other two spots:

➢All six original results are pooled to give a single *Hospital Admissions, Cardiovascular* result.

If you pool at *Hospital Admissions, Cardiovascular* and at *HA, All Cardiovascular*, but not at *HA, Congestive Heart Failure*:

➢First, the three Moolgavkar results are pooled to a give a single *HA, All Cardiovascular* result.

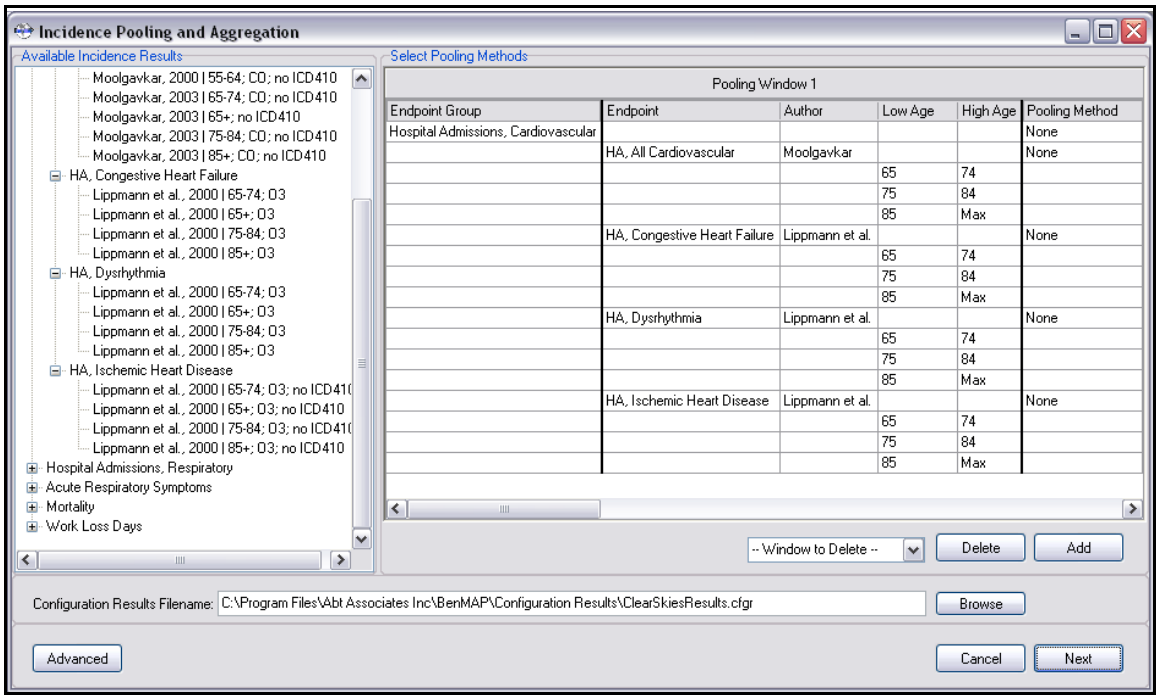
➢The result generated in the previous step is pooled with the three *HA, Congestive Heart Failure* results to give a single *Hospital Admissions, Cardiovascular* result.

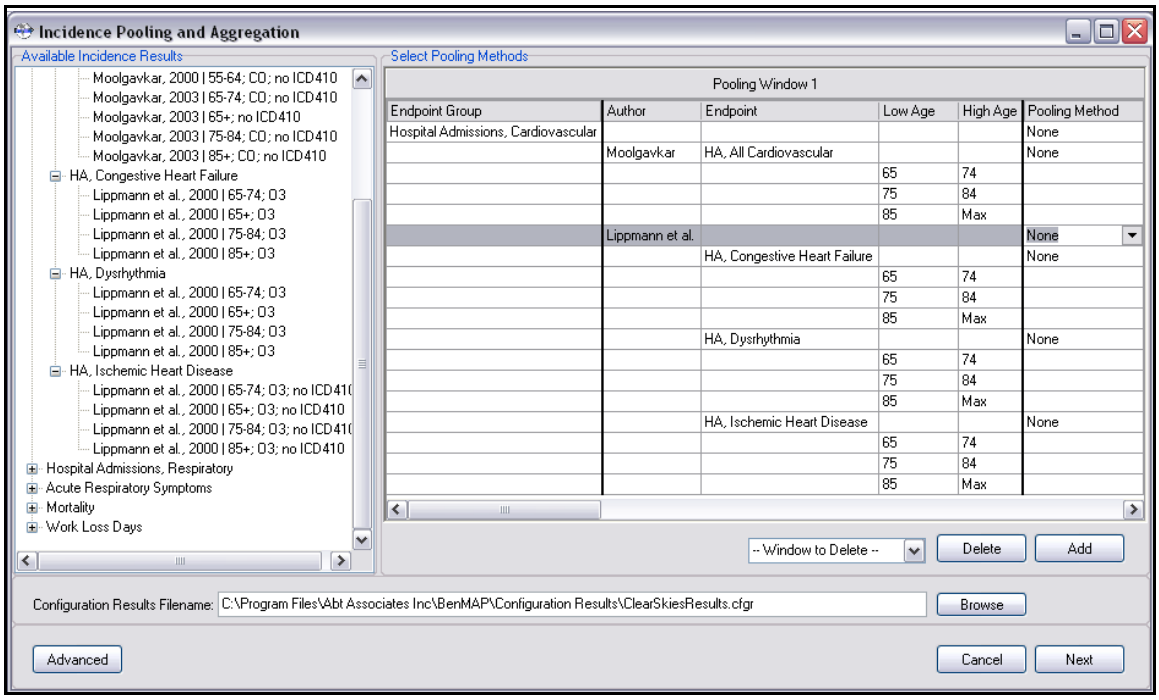
These same principles apply no matter how many incidence results are being pooled, and regardless of which pooling methods are selected.

Example: Using multiple pooling windows

As shown in the example above, there are many different ways to pool your incidence results. Sometimes you may want to look at the same results in different ways, or you may just have many results that need to be sorted by different variables. In these cases, you can open up multiple pooling windows by clicking on the **Add** button.

For example, you might want to pool all results of C-R Functions by a particular author, rather than pooling all results of C-R Functions of a particular endpoint. The examples below show the same set of incidence results, first sorted by endpoint, then sorted by author. As you can see, the pooling options are very different.





If you use two different pooling windows, each sorted as shown above, you can create results pooled by **Author**, *and* results pooled by **Endpoint**.

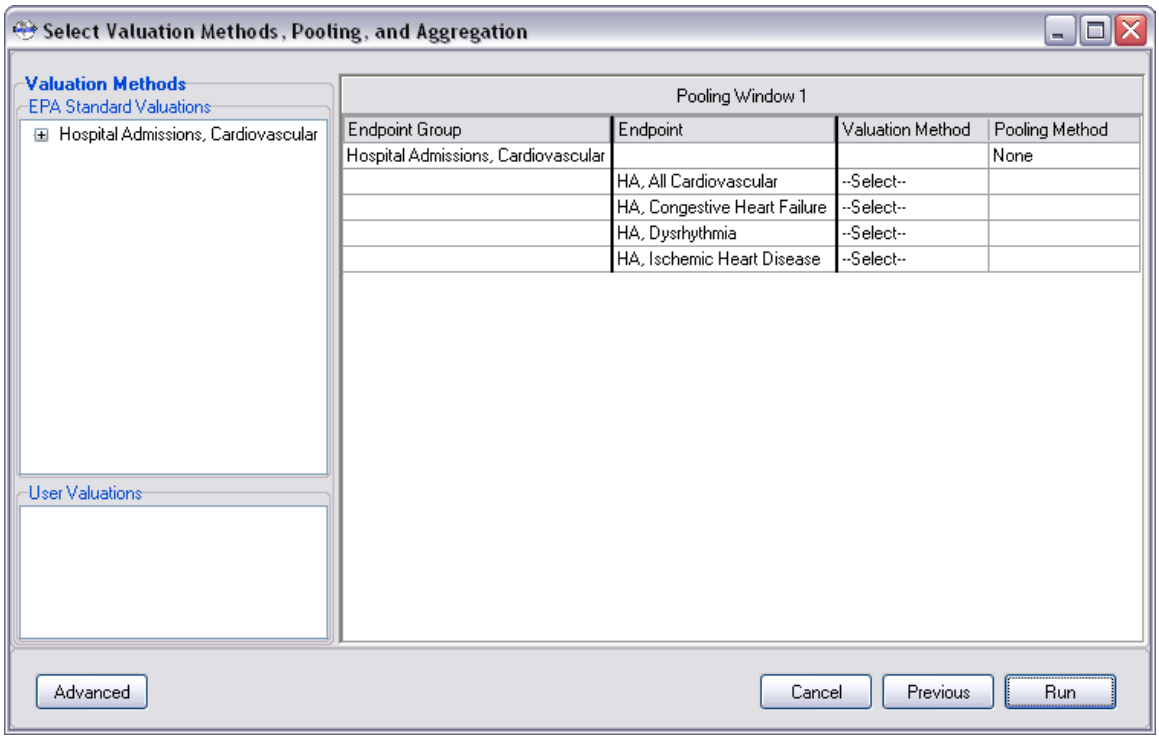
6.1.2 Valuing Pooled Incidence Results

After you have specified your incidence pooling options you can hit the **Next** button and select valuations and valuation pooling options. The **Select Valuation Methods, Pooling and Aggregation** form appears after you click **Next** on the **Incidence Pooling and Aggregation** form. This form should look quite similar to the **Incidence Pooling and Aggregation form**, with two tree views on the left side representing **EPA Standard Valuations** and **User Valuations**, and various pooling windows on the right side representing the selected valuations and pooling options.

There will be one pooling window in the **Select Valuation Methods, Pooling, and Aggregation form** for each pooling window in the **Incidence Pooling and Aggregation** form. In each pooling window, there will be one result present for each incidence result left over after all incidence pooling has occurred. Each of these results will be represented by a "-- Select --" value in the **Valuation Method** column.

The columns present in the **Select Valuation Methods, Pooling, and Aggregation** form are determined by the incidence results left after all incidence pooling has occurred. There will be exactly enough columns in each pooling window to represent the "least" pooled incidence result. That is, the columns will be in the same order they were in the **Incidence Pooling and**

Aggregation form, but the only columns present will be those up to the level of the pooled incidence result with the most columns left over after all pooling has occurred. Here is an example:



There are several steps to take in the **Select Valuation Methods, Pooling, and Aggregation** screen:

Step 1. Select your default Advanced options

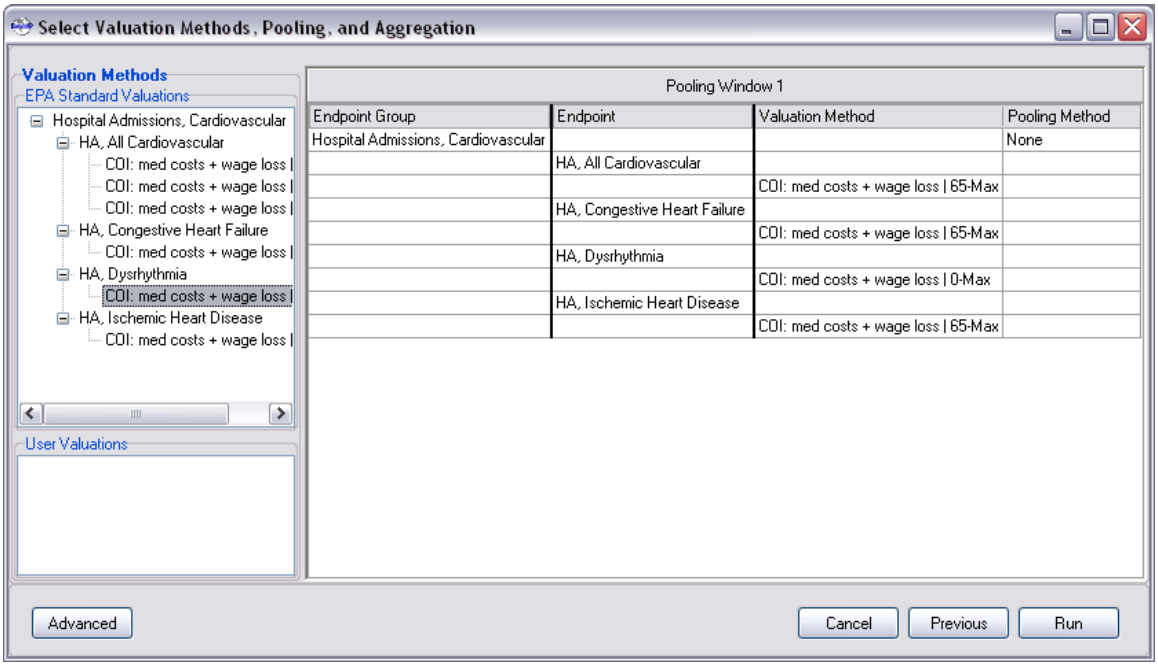
To select your default Advanced options, click on the **Advanced** button. There are two values which you will want to set at this point - the *Default Advanced Pooling Method* and the *Default Monte Carlo Iterations*. See Step 6, above, for a detailed discussion of these values. It is important to set them first because these default values will be applied to all valuation methods added to the pooling window (see Step 2, below) *after* they are set. Once they are set, click **OK**. If you decide not to change them, click **Cancel**. If you already set these values in Step 1 of pooling incidence results, you do not need to reset them here.

Step 2. Select your valuation methods

Valuation methods are specific to endpoint groups, and sometimes to endpoints as well. The only valuation methods which appear in the left side tree views are those which have the same endpoint group values as the pooled incidence results which are available to be valued. To select

a valuation method, select it in the left side tree views and drag and drop it onto the appropriate incidence result in the pooling window. Note that BenMAP will only allow you to drop valuation methods onto incidence results which have the same endpoint group value. For example, BenMAP will not allow you to drop a *Mortality* valuation on a *Hospital Admissions* incidence result. Note also that you can only drag and drop individual valuation methods, not entire groups of them. For explanations of the various valuations, see Appendix I.

If you have added any of your own **User Valuations**, using the **Data** menu (see Chapter 8), you can drag and drop them in the same way as the EPA Standard valuations.



When BenMAP runs the **APV Configuration**, it will generate a valuation result for each valuation method you select by running the valuation method's valuation functions on the incidence results for which they were selected. You do not need to select valuation methods for every incidence result - incidence results without any valuation methods will simply be ignored when valuation results are generated, aggregated, and pooled.

Because valuation functions have uncertainty associated with them, generating valuation results is fairly complicated. The procedure used depends on whether the incidence results being used were generated in **Point Mode** or with **Latin Hypercube Points** (see Chapter 5, above).

In **Point Mode**, BenMAP simply runs the valuation functions once using the point estimate of the incidence result and the mean of the valuation function (see section 9.2) as inputs.

With **Latin Hypercube Points**, on the other hand, BenMAP generates one hundred percentile points (from the 0.5th percentile to the 99.5th percentile) to represent the distribution of the inputs

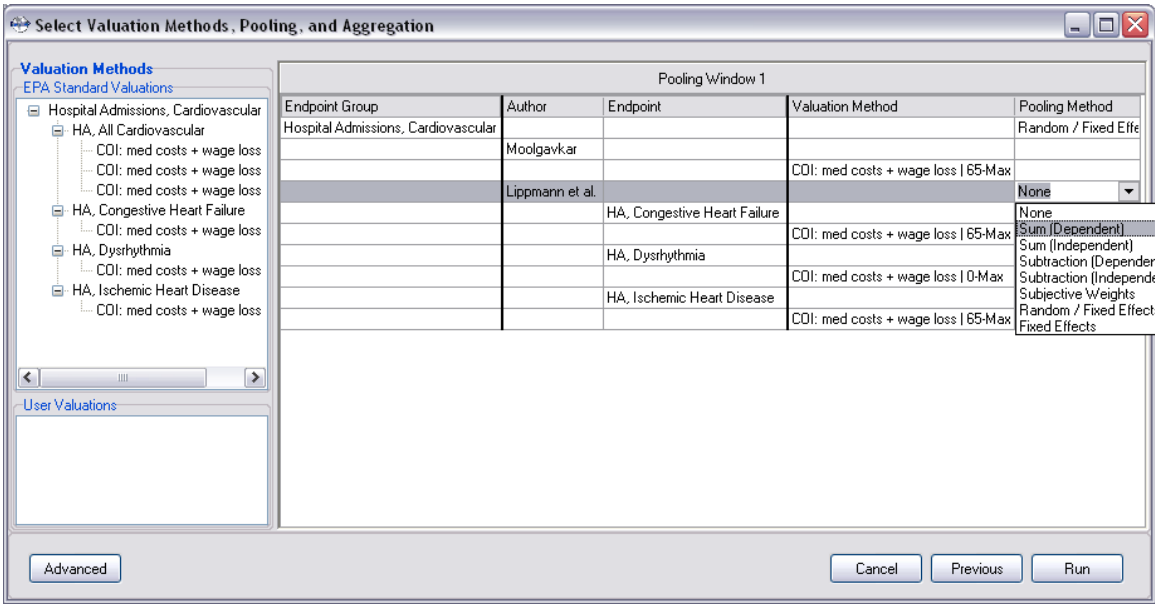
to the valuation function. It then runs the valuation function once for each combination of values from the incidence result, Latin Hypercube, and the hundred valuation points, putting the results into a holding container. Finally, the holding container is sorted low to high and binned down to the appropriate number of Latin Hypercube points, yielding a single valuation result.

Step 3. Sort results

Depending on how your incidence results were pooled, the columns in the valuation pooling windows can be resorted in the same way as the incidence pooling window columns. This resorting will have the same sort of impact on the tree structure of valuation results that it had on the tree structure of incidence results. See Step 3 of Section 6.1.1, above, for more information.

Step 4. Select pooling methods

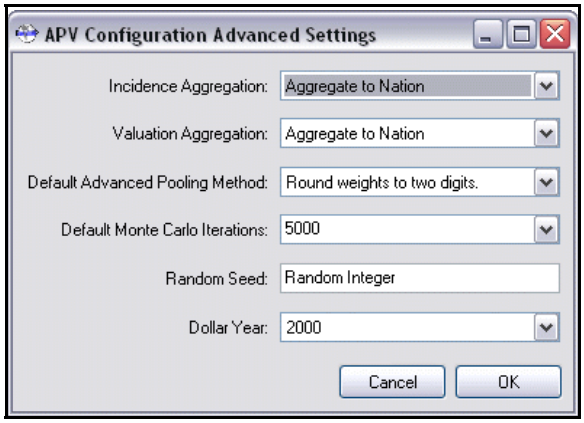
The same pooling methods are available for valuation results which were available for incidence results. See Step 4 of Section 6.1.1, above, and Exhibit 6-1, above, for more details. You should note that when more than one valuation method is selected for a particular pooled incidence result, it is possible to pool the generated valuation results.



Step 5. Advanced Pooling Methods, Monte Carlo Iterations

The same advanced pooling methods are available for valuation results which were available for incidence results. See Step 6 of Section 6.1.1, above, for more details.

6.1.3 APV Configuration Advanced Settings



At any point when specifying the incidence and valuation pooling options, you may click on the **Advanced** button on the bottom-left of either the screen for **Incidence Pooling and Aggregation** or **Select Valuation Methods, Pooling and Aggregation**. This button will open the **APV Configuration Advanced Settings** form.

The **APV Configuration Advanced Settings** form lets you choose the level of aggregation for the incidence and the valuation results. Since the valuation depends on the incidence, the level of aggregation for the valuation must equal or exceed that of incidence. For example, county-level incidence aggregation may be combined with national-level valuation aggregation, but not vice-versa.

As described above (see **Step 6. Advanced Pooling Methods, Monte Carlo Iterations**, for more details), the **APV Configuration Advanced Settings** form also allows the specification of **Default Advanced Pooling Method** and **Default Monte Carlo Iterations** values. It is recommended that these be set before any incidence results are added to the **Incidence Pooling and Aggregation** pooling windows.

The **APV Configuration Advanced Settings** form also allows the specification of a **Dollar Year** - all valuation dollar figures will be reported in this years dollars. See Exhibit 8-5 for descriptions of the variables used to adjust dollar figures to different years.

Finally, the **APV Configuration Advanced Settings** form allows the specification of a **Random Seed**. As mentioned at the beginning of this chapter (see **Open an APV Results File**, above) many of the pooling methods require the generation of sequences of random numbers (e.g. choosing a random Latin Hypercube point during a Monte Carlo simulation). Providing a specific **Random Seed** value allows the user to ensure that the same sequence of random numbers is generated as in a previous analysis, thus allowing exact results to be reproduced.

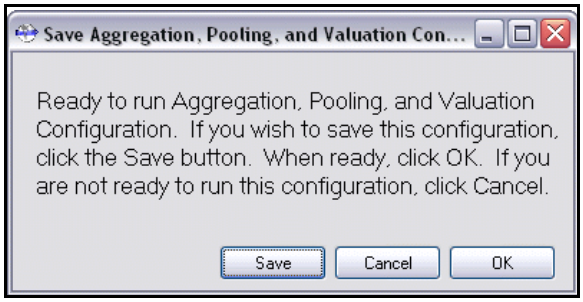
If you do not set the **Random Seed** for a particular run, one will be generated automatically from the system clock (the number generated will depend on the date and time, and should change every minute). Normally, you should *not* set the **Random Seed** value. If you need to reproduce a specific set of results, however, the random seed used to generate previous **APV Configuration Results** can be obtained from an APV Configuration Result file (*.apvr) Audit Trail Report (see Section 7.3, below).

6.2 Running the APV Configuration

After having specified the various pooling and aggregation options, you have the opportunity to save your configuration for future use. The file that you save has an “apv” extension. The

configuration that you have specified is similar in idea to the configuration that you developed for choosing C-R functions. (That configuration has a "cfg" extension.) Both files allow you to save choices that you have made, and re-run them at a later time.

You can save your APV configuration when you have finished making your valuation pooling choices. Click the **Run** button, and then choose **Save**.



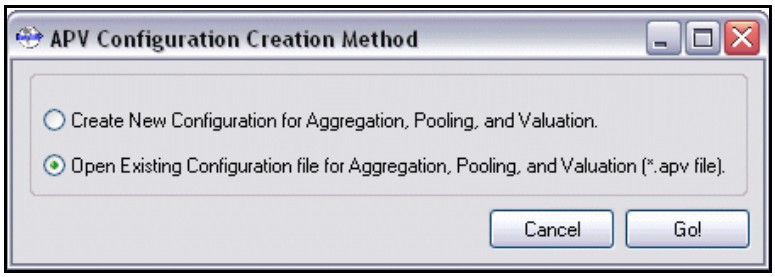
You then need to name your configuration (*.apv) file. We suggest that you save this in the Configurations folder. When ready to generate APV Configuration results, click the **OK** button. BenMAP then requires that you specify a file in which to save the results, with an ".apvr" extension.

➥ TIP: **Naming APV Configuration and APV Configuration Results files**

To keep track of your work, you may find it helpful to use the same name for your APV Configuration (*.apv) and APV Configuration Results (*.apvr) files.

6.3 Open Existing Aggregation, Pooling, and Valuation (APV) Configuration File

If you have an existing configuration (*.apv) file, you can open, and then edit it. If you have only a few changes to make to an existing configuration, it is typically much quicker to open the previous configuration, rather than entering all of your choices again. Note that the various parts of an *APV Configuration* are quite interdependent, so modifying part of the configuration may cause other parts to be reset. For



example, modifying the tree structure for incidence pooling will cause the valuation method selection and valuation pooling tree structure to be cleared and reset. Changing the *Configuration Results Filename* in the *Incidence Pooling and Aggregation* form will *not* reset the incidence or valuation pooling trees as long as the new file contains incidence results generated from the same C-R Functions as the old file. This can be quite helpful for generating new *APV Configuration Results* from several different *Configuration Results* files which were generated from different baseline / control scenarios, but with the same set of C-R Functions.

CHAPTER 7

Create Reports

In this chapter...

➢ Use the Create Reports button to create incidence and valuation reports.

➢ Find the file formats, including variable definitions and program compatibility, for each report type.

➢ Find out about using an Audit Trail report to keep track of the options and assumptions underlying each analysis.

Chapter Overview

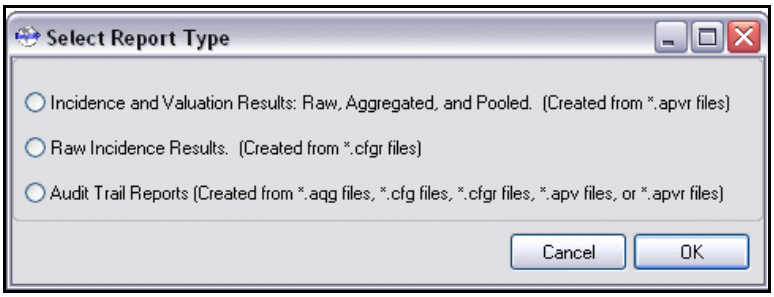
7. Create Reports

There are three types of reports that you can access by clicking on the **Create Reports** button. You will be asked which type of report you wish to create:

➢ **Incidence and Valuation Results: Raw, Aggregated, and Pooled** use an Aggregation, Pooling, and Valuation Results file (with the ".apvr" extension) to create report for incidence, aggregated incidence, pooled incidence, valuation, aggregated valuation, or pooled valuation results. These reports are comma separated values (CSV) files (*.csv) which can be read into various spreadsheet and database programs, such as Microsoft Excel.

➢ **Raw Incidence Results** use a Configuration Results file (with the ".cfgr" extension) to create reports for incidence results. These reports are CSV files.

➢ **Audit Trail Reports** provides a summary of the assumptions underlying each of five types of files generated by BenMAP: **Air Quality Grids** (with the ".aqg" extension), **Incidence Configurations** (with the ".cfg" extension), **Configuration Results** (with the ".cfgr" extension), **Aggregation, Pooling, and Valuation Configurations** (with the ".apv" extension), and **Aggregation, Pooling, and Valuation Results** (with the ".apvr" extension). These reports can be viewed within BenMAP in an expandable tree structure, or can be exported to tab-delimited text files.



7.1 Incidence and Valuation Results: Raw, Aggregated, and Pooled

Using the results in the Aggregation, Pooling, and Valuation Results file (".apvr" extension), you can create six types of reports: raw incidence, aggregated incidence, pooled incidence, valuation (of the pooled incidence), aggregated valuation, and pooled valuation. After you click on the **Create Reports** button and specify your choice, you need to specify the **APV Configuration Result File** that you want to use (see Chapter 6 for how to create an APV result file). You then need to choose a *Result Type*. Exhibit 7-1 describes the results contained in each type of report.

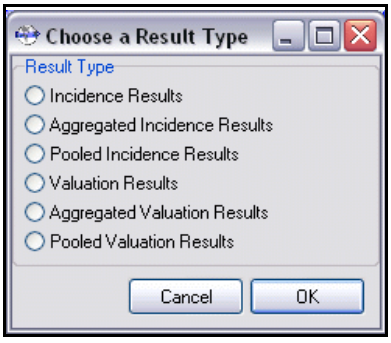
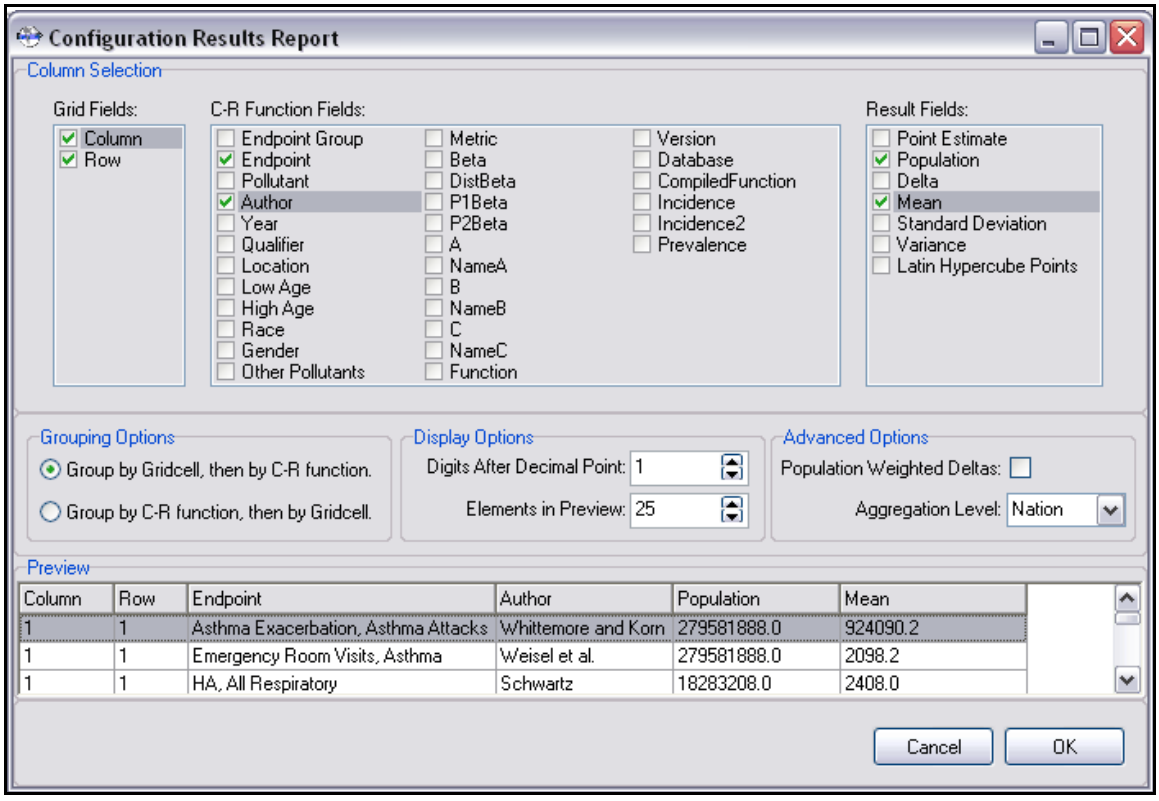


Exhibit 7-1. Summary of the Reports Generated from APVR File

Result Type	Description
Incidence Results	Incidence results for each C-R function at the grid-cell level or aggregated at the county, state, or national level.
Aggregated Incidence Results	Incidence results for each C-R function aggregated to the level you specified in the Aggregation, Pooling, and Valuation configuration file.
Pooled Incidence Results	Incidence results aggregated and pooled as you specified in the Aggregation, Pooling, and Valuation configuration file.
Valuation Results	Valuation results for the pooled and aggregated incidence results.
Aggregated Valuation Results	Valuation results aggregated to the level you specified in the Aggregation, Pooling, and Valuation configuration file.
Pooled Valuation Results	Valuation results aggregated and pooled as you specified in the Aggregation, Pooling, and Valuation configuration file.

7.1.1 Incidence and Valuation Results: Incidence Results

The **Incidence Results** report gives you the opportunity to examine the results of each C-R function at the grid-cell level, or aggregate them to the county, state, or national level. Simply select the options that you desire from the four main sections of the **Configuration Results Report** form: **Column Selection** (these include **Grid Fields**, **C-R Function Fields**, and **Result Fields**), **Grouping Options**, **Display Options**, and **Advanced Options**. As you modify your choices, the **Preview** section will be updated accordingly.



The **Column Selection** section allows you to choose the field names (and values) which will appear in the report. The *Grid Fields* section allows the inclusion of *Col* and *Row* fields, which can be helpful in identifying the grid-cell of a particular line in the report. These will not always be necessary, however - for example, when results have been aggregated to the national level. The *C-R Function Fields* section allows the inclusion of various fields which can be helpful in identifying the C-R function of a particular line in the report. Almost all of the field names have appeared previously in the preparation of a Aggregation, Pooling, and Valuation Results file. Finally, the *Result Fields* section allows the inclusion of various types of results.

Exhibit 7-2 provides a summary of the variables available in this report format.

In the **Grouping Options** section, you can change the sorting of the results, by clicking the radio buttons *Group by Gridcell, then by C-R function* and *Group by C-R function, then by Gridcell*.

In the **Display Options** section, you may set the number of digits that appear after the decimal point, and you can set the number of rows that appear in the preview window.

In the **Advanced Options**, you can set the level of aggregation at the grid cell (none), county, state, and national levels. You can also choose to generate *Population Weighted Deltas*, which BenMAP calculates at the national level for each C-R function by weighting the change in the pollution metric at each grid cell with the population of the grid cell. For example, if there are large changes in highly populated urban grid cells and relatively small changes in lightly population

rural grid cells, then the population-weighted change would reflect the large urban changes and be relatively large.

Exhibit 7-2. Selected Variables in the Reports Based on the APVR file

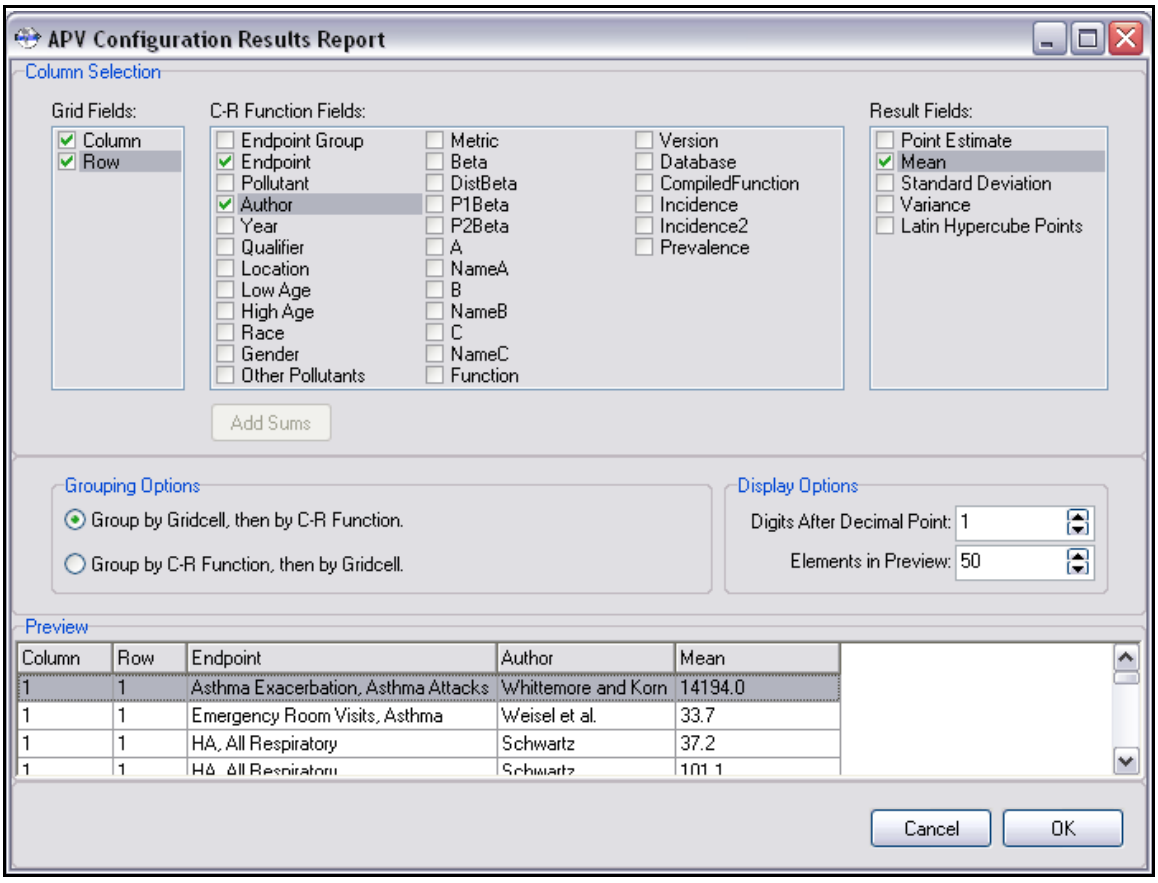
Variable	Variable Description
Col	The column of the grid cell of the result. For grid cell level results, this is the column of the grid cell. For county and state level results, this is the state FIPS code. For national results, this is always 1.
Row	The row of the grid cell of the result. For grid cell level results, this is the row of the grid cell. For county results, this is the county FIPS code. For state and national results, this is always 1.
Endpoint Group	The endpoint group of the result (from the C-R function and/or valuation method). See Exhibit 5-1 for a list of endpoint groups.
Endpoint	The endpoint of the result (from the C-R function and/or valuation method). See Exhibit 5-1 for a list of endpoints.
Author	Author of the study used to develop the C-R function associated with the result.
Year	Year of the study used to develop the C-R function associated with the result.
Location	Location of the study used to develop the C-R function associated with the result.
Qualifier	For incidence results, the qualifier is a description that uniquely identifies a C-R function when combined with the endpoint group, endpoint, author, and year. For valuation results, the qualifier is a description that uniquely identifies a valuation function when combined with the endpoint group, endpoint, and age range.
Other Pollutants	Other pollutants that were simultaneously included in the original study used to develop the C-R function associated with the result.
Metric	Air quality metric used in the C-R function associated with the result.
Function	The C-R function associated with the result.
Compiled Function	To run faster, BenMAP uses C-R functions that have already been compiled. There are approximately 25 types of compiled in functions that are numbered starting with zero.
Version	The version of the C-R function associated with the result. Version indicates the number of times that a particular C-R function has been included in a configuration. Typically the different versions of a C-R function will have different population variables (Low Age, High Age, Race, and/or Gender).
ValuationMethod	A combination of the qualifier and age range of the result (from the associated valuation method(s)).
Population	Population provides the number of persons used in the C-R function calculation.
Delta	The difference between the baseline and control scenarios for the metric used in the C-R function. Calculated by subtracting the metric value in the control scenario from the metric value in the baseline scenario.
Point Estimate	The point estimate for this result.
Mean	Mean of the points in the Latin Hypercube for this result.
Std Dev	Standard deviation calculated based on the points in the Latin Hypercube for this result.
Latin Hypercube Points	The number of percentiles depends on the number of points in the Latin Hypercube for this result.

7.1.2 Incidence and Valuation Results: Aggregated Incidence Results

The **Aggregated Incidence Results Report** presents the incidence results at the aggregation level that you have previously specified in the Aggregation, Pooling, and Valuation Configuration file. If you want to change the level of aggregation, you need to revise your choices in the **APV Configuration Advanced Settings** screen (see Chapter 6).

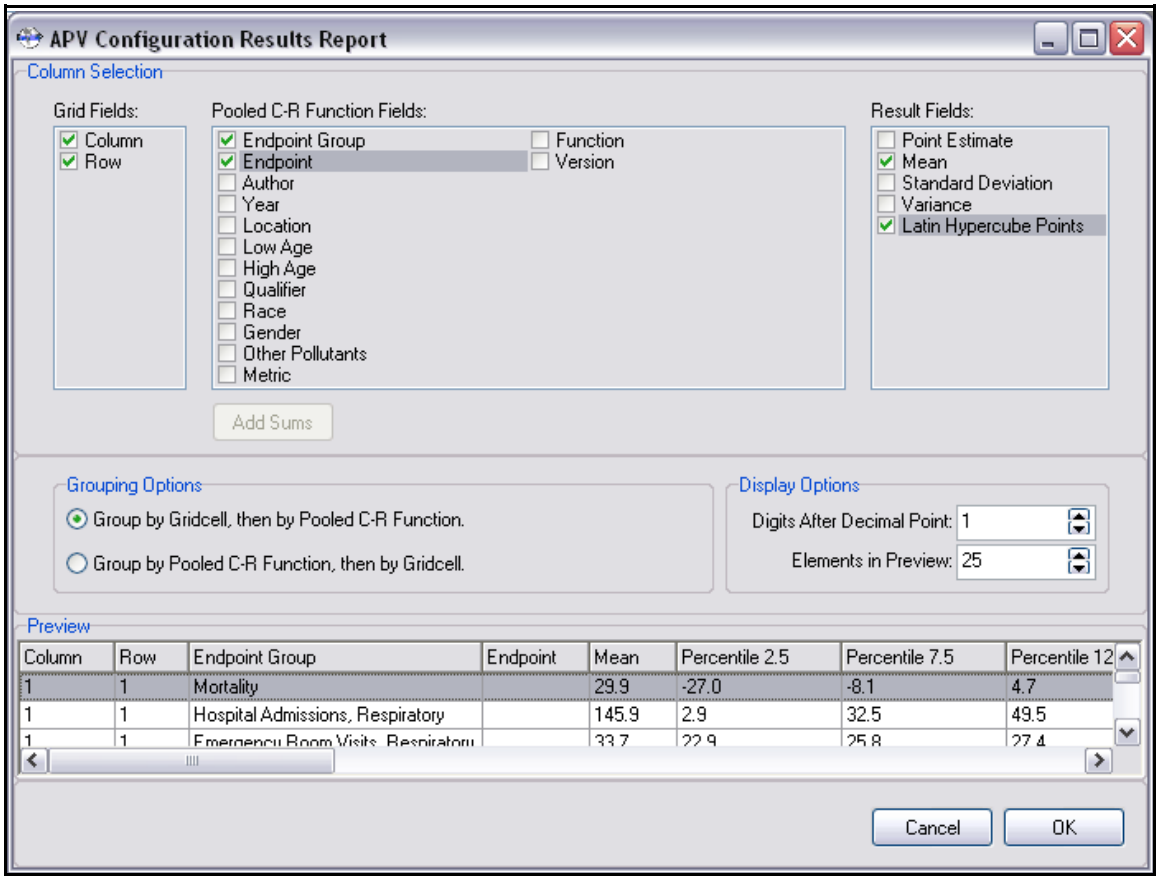
Aggregation refers to the summing of grid cell level results to the county, state or national level. Pooling refers to combining individual incidence results or valuations into composite results.

The **Column Selection** section looks largely the same as in the **Incidence Results Report**, except that the *Population* and *Delta Result Fields* are not available. The **Grouping Options** section and **Display Options** section are exactly the same. The **Advanced Options** section no longer exists, as the options in it do not apply to aggregated incidence results.



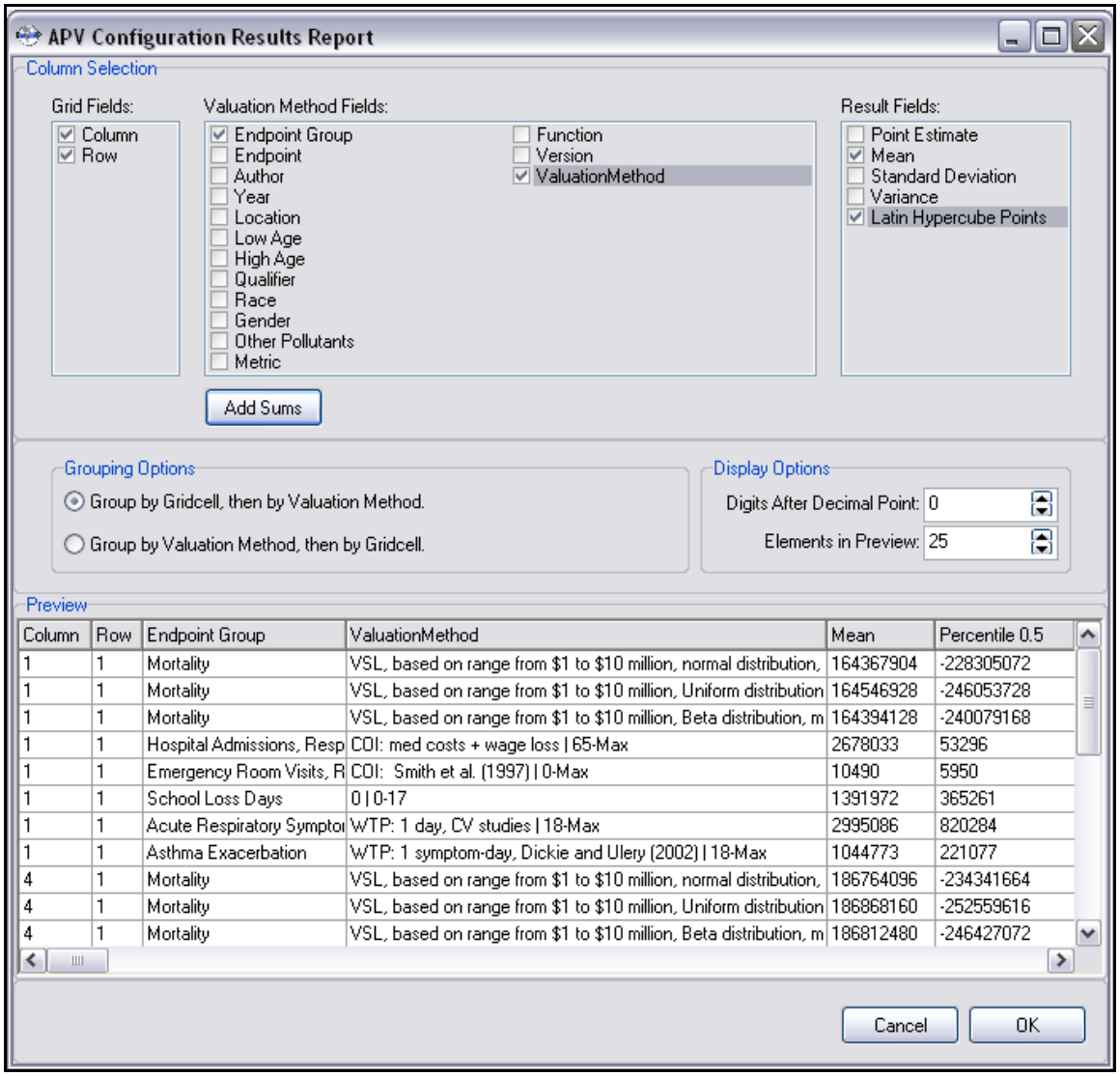
7.1.3 Incidence and Valuation Results: Pooled Incidence Results

The **Pooled Incidence Results Report** provides results aggregated and pooled to the level that you previously specified in the Aggregation, Pooling, and Valuation Configuration file. This report looks largely the same as the **Aggregated Incidence Results Report**, except that fewer *C-R Function Fields* are available, and values for others will be blank. This is because after pooling, only enough fields are retained to uniquely identify individual results.



7.1.4 Incidence and Valuation Results: Valuation Results

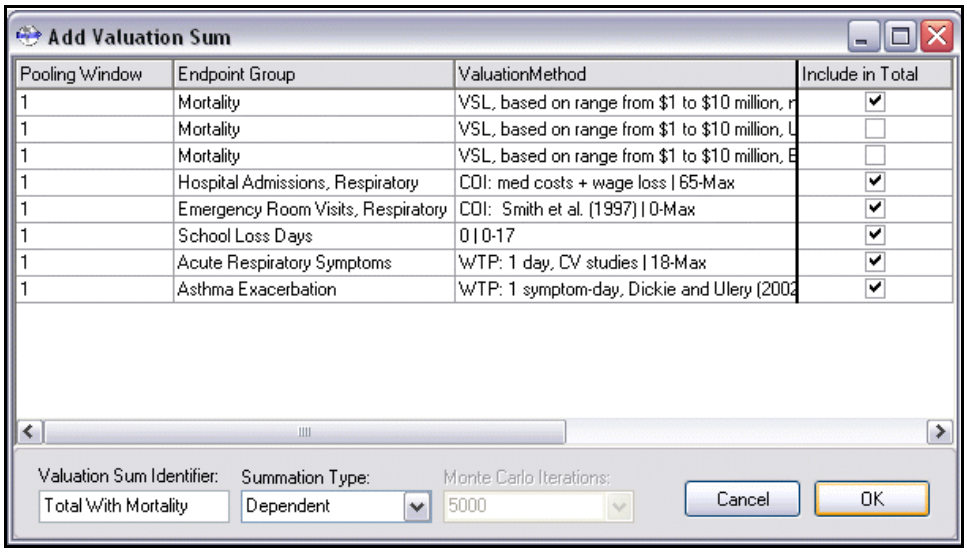
The **Valuation Results** report gives you the opportunity to examine the valuation results for the pooled and aggregated incidence results. In the example below, the incidence results were aggregated to the state level, so in the **Column** field, which represents the state FIPS code, you can see the value changing from *1* (Alabama) to *4* (Arizona).



Add Sums Button

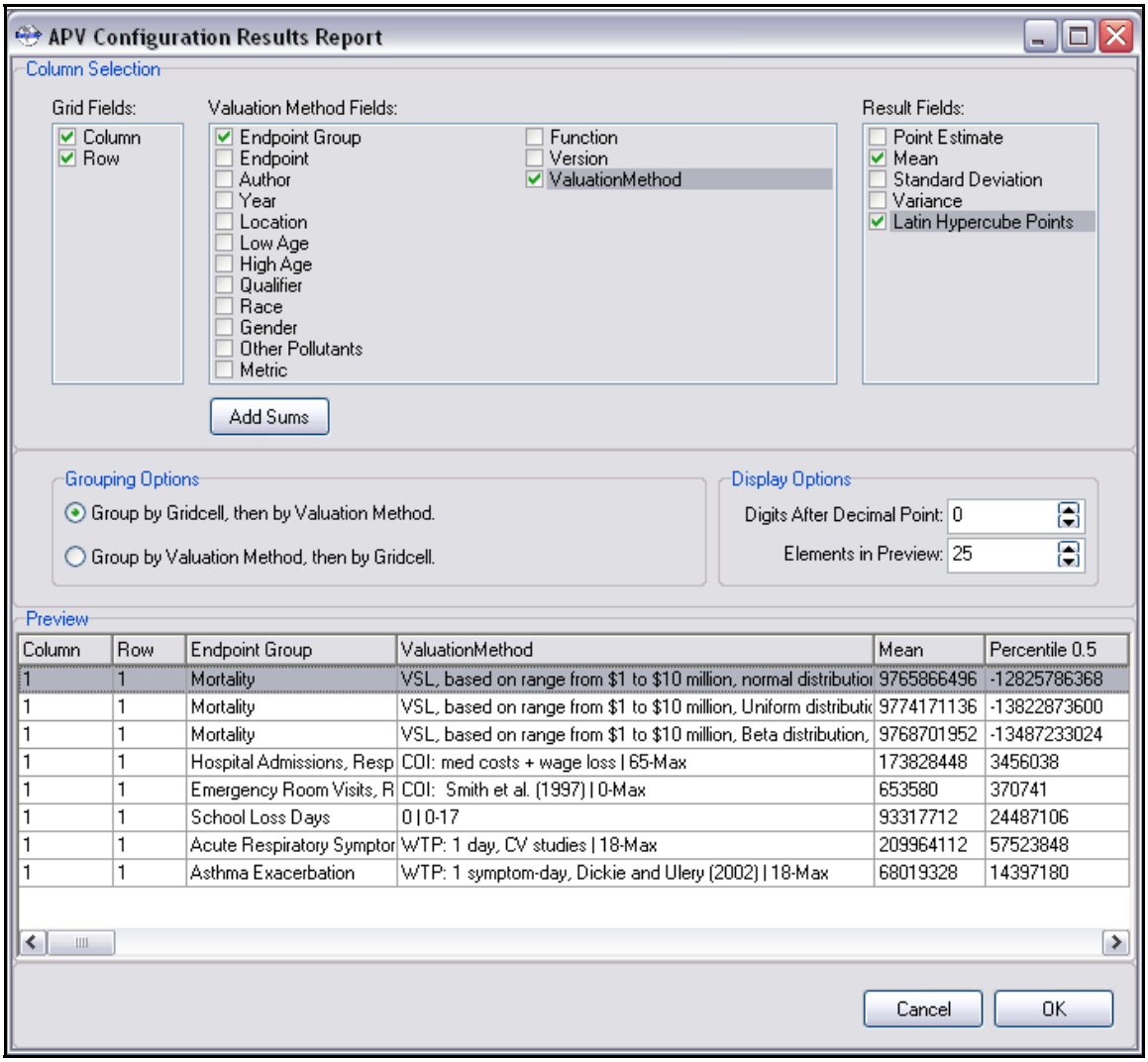
To calculate the total value of different groups of effects, click on the **Add Sums** button in the **APV Configuration Results Report** screen. On the next screen you can specify the effects you want to sum and how you want to sum them. In the **Include in Total** field, simply check the effects that you want to sum together, and then name the sum in the **Valuation Sum Identifier** field. You also need to specify whether you are going to use the *Dependent* or *Independent* sum approach. (See Chapter 6 for a discussion of these two summing approaches.) In the **Summation Type** use the drop-down menu, choose the approach that you want to use. If you choose the *Independent* sum, then you also need to specify the number of Monte Carlo draws that you want BenMAP to use. The default is 5,000. You may repeat this procedure to generate as many totals as you like.

Note that the **Add Sums** button is only enabled for reports involving monetary valuations, not those involving incidence estimates. Typically, incidence estimates should not be summed across endpoint groups (for example, Mortality and Hospital Admissions, Respiratory). Within endpoint groups, incidence estimates can be summed - you may do this in Aggregation, Pooling and Valuation Configurations (see Chapter 6). Once results are in monetary values, however, summing across endpoint groups can be useful in calculating aggregate benefits.



7.1.5 Incidence and Valuation Results: Aggregated Valuation Results

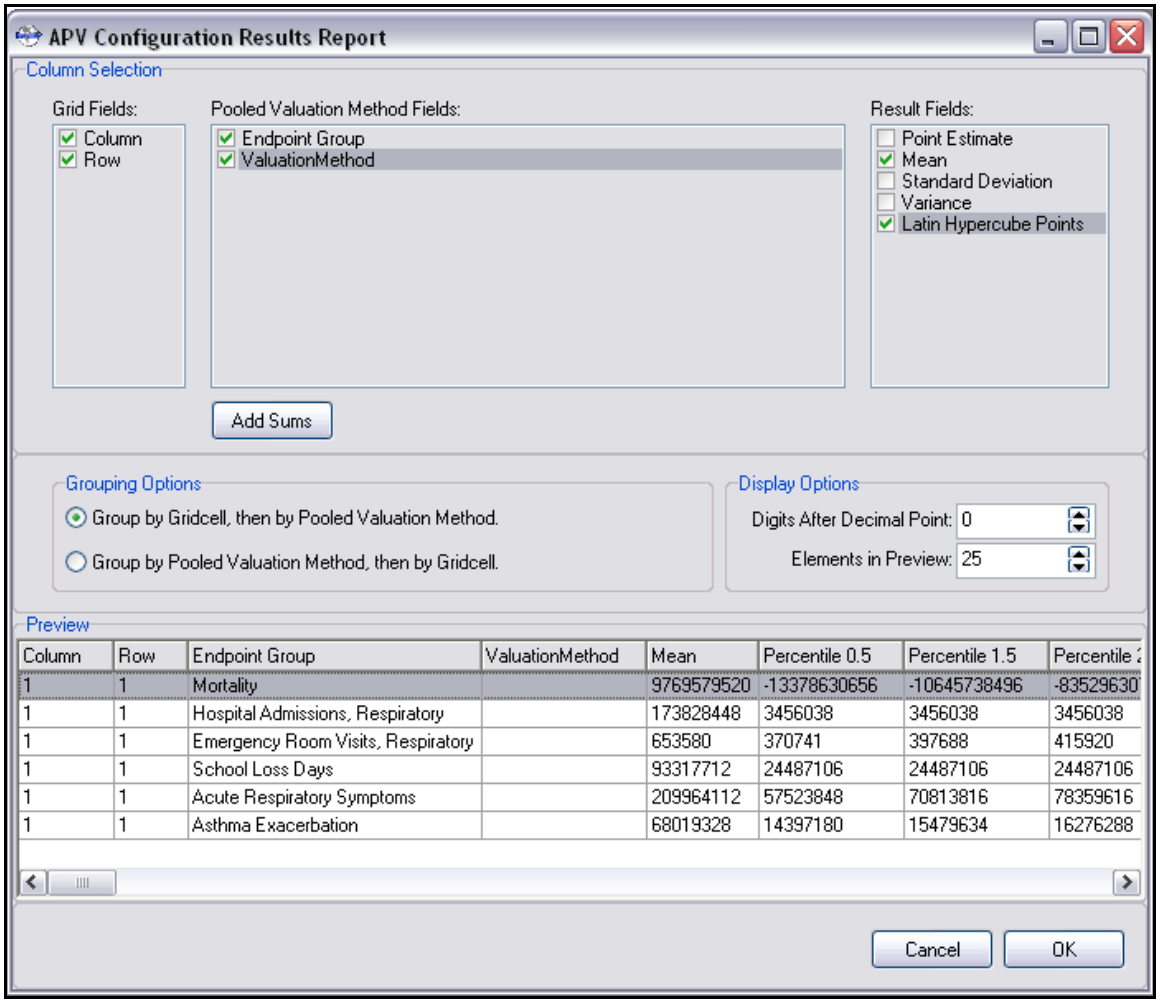
The **Aggregated Valuation Results Report** presents valuation results aggregated to the level you specified in the Aggregation, Pooling, and Valuation configuration file. In the example below, the valuation results are aggregated at the national level, so all of the results have the same value (*1*) in the *Column* and *Row* fields. As with the **Valuation Results Report**, you can use the **Add Sums** button to create totals with various valuation results.



7.1.6 Incidence and Valuation Results: Pooled Valuation Results

The **Pooled Valuation Results Report** presents valuation results aggregated *and pooled* to the level you specified in the Aggregation, Pooling, and Valuation configuration file. (Recall that aggregation refers to the geographic level that you have combined your results, and pooling refers to how you have combined the results of different C-R functions / valuations.) As with the **Pooled Incidence Results Report**, fewer *Pooled Valuation Method Fields* are available, because only enough fields are retained to uniquely identify individual results.

In the example below, several different valuation approaches have been combined when valuing mortality, so the **Valuation Method** field is blank. As with the other valuation reports, you can use the **Add Sums** button to create totals with various valuation results.



7.2 Raw Incidence Results

The *Raw Incidence Results* report gives you the opportunity to examine the results of each C-R function at the grid-cell level, county, state, or national level. It is based on the Configuration Results file (with the *.cfgr extension), and is otherwise identical to the **Incidence Results Report** generated from the Aggregation, Pooling, and Valuation results file (with the *.apvr extension). BenMAP includes both versions to increase the reporting flexibility for users. (See Section 7.1.1 for a description of the options available for this report type.)

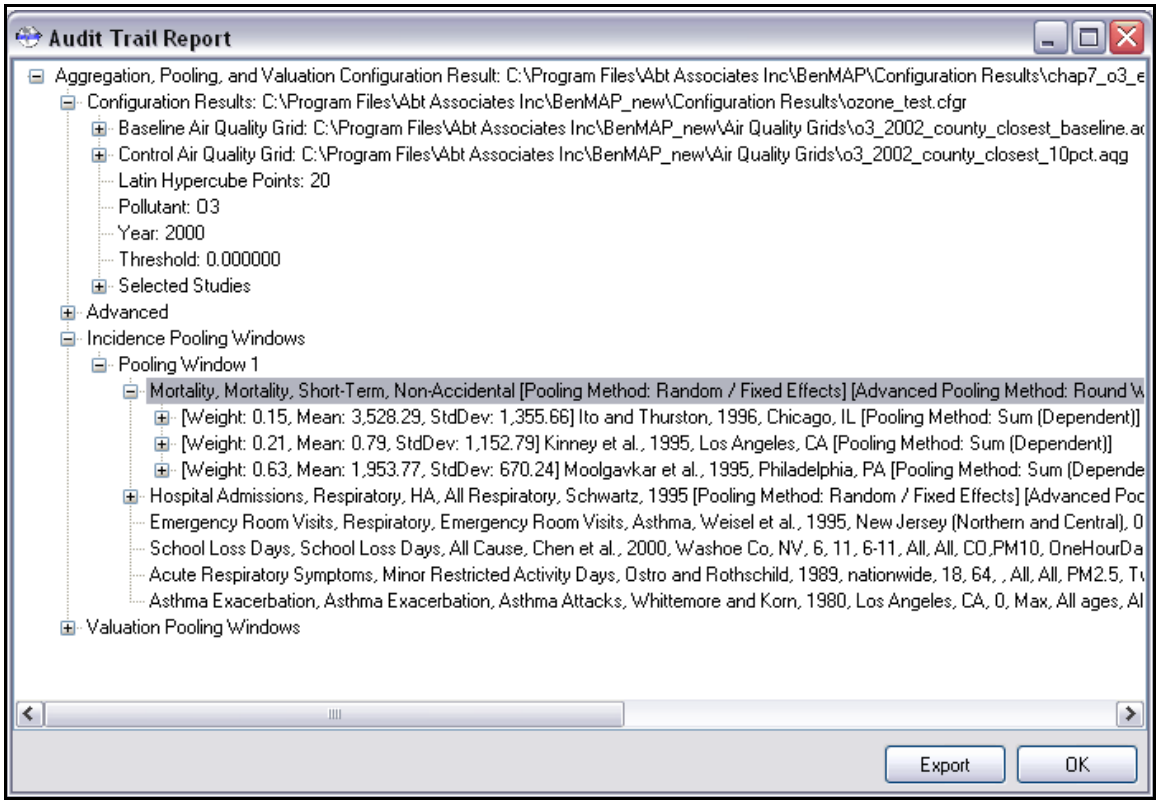
7.3 Audit Trail Reports

The **Audit Trail Reports** provide a summary of the assumptions underlying the various parts of the analysis. You may generate an audit trail with any of the file types used in BenMAP: Air Quality Grids (with the ".aqg" extension), Incidence Configurations (with the ".cfg" extension),

Configuration Results (with the ".cfgr" extension), Aggregation, Pooling, and Valuation Configurations (with the ".apv" extension), and Aggregation, Pooling, and Valuation Results (with the ".apvr" extension). The report itself has a tree structure that lets you easily find the information that you are seeking. Below is an example of an **Audit Trail Report** for a Aggregation, Pooling, and Valuation Results file.

Note that each successive step in an analysis contains a summary of its assumptions, and those of each previous step in the analysis. For example, in the below report the assumptions of the Configuration Results file used to generate the APV Results are present in the *Configuration Results* node. Similarly, the assumptions of both the baseline and control air quality grids are present under the *Configuration Results* node.

Note that you can export **Audit Trail** as a text file . Each branch in the tree structure will be converted to a tab in the exported file, allowing for easy viewing in Excel, WordPad, and a variety of other programs. Simply click on the **Export** button, name the file, and click **Save**.



7.4 Questions Regarding Creating Reports

Below are answers to some of the common questions that may arise in the creation of reports.

➢When creating reports from *.cfgr and *.apvr files, why do some of the variables that I have checked appear as blanks?

When results are pooled together, some of the identifying information for individual C-R functions gets lost. For example, when pooling together endpoints within the same endpoint group, such as "HA, Pneumonia" and "HA, Chronic Lung Disease" (both within "Hospital Admissions, Respiratory"), there is no longer a unique endpoint name for the pooled result. So, BenMAP would leave the endpoint name blank.

CHAPTER 8

Mapping

In this chapter...

➢ Learn about BenMAP's mapping functions.

➢ Map configuration inputs like air quality grids.

➢ Map incidence and valuation results.

➢ Map different variables and modify the map display.

Chapter Overview

8. Mapping

BenMAP features integrated mapping capabilities which you can access at several points in the model. The main *Mapping / GIS* tool, available via the **Tools** drop-down menu in the main screen, allows you to map all types of files and data associated with an analysis, including air quality grid files, monitor data, population data, and incidence and valuation results. In addition, at several points from within the program, you can view data being used in an analysis. You can view maps when filtering monitor data, creating air quality grids, and when creating incidence configuration files (with the *.cfg extension).

8.1 Overview of Mapping Features

To access the main mapping capabilities within BenMAP, go to the **Tools** drop-down menu, and choose the *Mapping / GIS* option. A blank screen will appear, with buttons at the top for managing files and navigating the map. To see the name of each button, simply hold the cursor over it. Use the **Open a file** button in the top-left corner to choose the file (or other type of data) that you want to view. You can view all of the files and data underlying an analysis. Each map that you select and load into the GIS viewer will appear on the left-hand side of the screen as a layer. Exhibit 8-1 describes each type of map you can view.

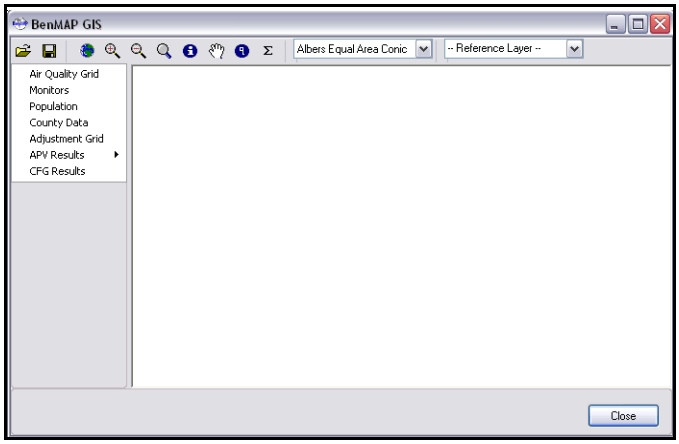


Exhibit 8-1. Description of Mapping File Types

File Type	Variables Available	Source File Used	Notes
Air Quality Grid	Annual summary values (average, median or other metric where available) within each grid cell.	Air quality grids created by BenMAP (*.aqg)	
Monitors	Annual summary values (average, median or other metric where available) for each monitor.	Select monitors from the BenMAP library or use your own by loading a file.	Customize the monitor filtering options or use the default settings.

File Type	Variables Available	Source File Used	Notes
Population	Total population, and cross-tabulations of age and gender with race/hispanic origin within each grid cell (for the Year and GridType you specify).	Internal BenMAP population files and projections, based on Census block-level data.	Variable names for the cross-tabulations are abbreviated: for example, *a_m_35to39* is Asian males 35 to 39 years of age. Total population and single variables are at the bottom of the list.
County Data	Incidence rates by age group; some household, wage, income and employment data; total population. At the county level.	Internal BenMAP incidence data.	Variables are listed in alphabetical order; like variables are not grouped together.
Adjustment Grid	**PM grids**: each variable represents a seasonal adjustment factor. Daily average concentrations for each season are sorted and placed in 5 bins. The average of each bin is used as the adjustment factor: *s1_b1* is the average for season 1's first bin. **Ozone grids**: each variable represents one of 10 decile adjustment factors.	BenMAP adjustment files (*.adj)	Adjustment files are created using the *Adjustment Factor Creator* under the **Tools** menu, or during the Air Quality Grid creation process using the *Model and Monitor Relative* option.
APV Results: Incidence	For each incidence result, the **Result**X variable is the mean increase in cases between the control and baseline scenarios; the **DELTA_X** variable is the difference between the baseline and control scenarios for the metric used in the C-R function; and the **POP_X** variable is the number of persons used in the C-R function.	Aggregation, Pooling and Valuation Results files (*.apvr)	Variables are as calculated between the baseline and control scenarios before pooling and aggregation. Each incidence result is given a number (e.g. Result0, Result1). Result variables can be renamed in the **Edit GIS Field Names** window.
APV Result: Aggregated Incidence	For each aggregated incidence result, the **ResultX** variable is the mean increase in cases between the control and baseline.	Aggregation, Pooling and Valuation Results files (*.apvr)	Results are aggregated to the level specified in the APVR file. If results were aggregated to the national level, the US map will only show one color. Variables can be renamed in the **Edit GIS Field Names** window.
APV Results: Pooled Incidence	For each pooled incidence result, the **ResultX variable** is the mean increase in cases between the control and baseline.	Aggregation, Pooling and Valuation Results files (*.apvr)	Results are pooled and aggregated as specified in the APVR file. Variables can be renamed in the **Edit GIS Field Names** window.
APV Results: Valuation	For each valuation result, the **ResultX** variable is the mean economic value placed on the increase in cases between the control and baseline. You can also map any new sums created with the **Add Sums** button.	Aggregation, Pooling and Valuation Results files (*.apvr)	Only incidence results assigned a value in the APVR file will appear in the variable list. You can create new groups of results and add their valuations together using the **Add Sums** buttons. Variables can be renamed in the **Edit GIS Field Names** window.

File Type	Variables Available	Source File Used	Notes
APV Results: Aggregated Valuation	For each aggregated valuation result, the **ResultX** variable is the mean economic value placed on the increase in cases between the control and baseline. You can also map any new sums created with the **Add Sums** button.	Aggregation, Pooling and Valuation Results files (*.apvr)	Values are aggregated to the level specified in the APVR file. Variables can be renamed in the **Edit GIS Field Names** window.
APV Results: Pooled Valuation	For each pooled valuation result, the **ResultX** variable is the mean economic value placed on the increase in cases between the control and baseline. You can also map any new sums created with the **Add Sums** button.	Aggregation, Pooling and Valuation Results files (*.apvr)	Values are pooled and aggregated as specified in the APVR file. Variables can be renamed in the **Edit GIS Field Names** window.
APV Results: All	Variables as described above in each report type.	Aggregation, Pooling and Valuation Results files (*.apvr)	BenMAP opens each file type as a separate layer. You can widen the left-hand panel to see the full layer names.
CFG Results	Variables are the same as **APV Results: Incidence**.	Configuration Results files (*.cfgr)	This produces the same map and variables as the *Incidence* option under *APV results*. Variables can be renamed in the **Edit GIS Field Names** window.

8.1.1 Display Options

After choosing the file or data type that you want to map, you must select the variable you want to map, and how you want it displayed. Double-click on the layer name on the left-hand side of the screen. A small box will appear with the **Display Options** for the layer. This box allows you to set the following options:

Variable. The drop-down menu lists all of the variables contained in the layer you chose. Select the variable that you want to view. If you want to view multiple variables in the same map at the same time, you must load multiple versions of the same layer. Each layer can only show one variable at a time.

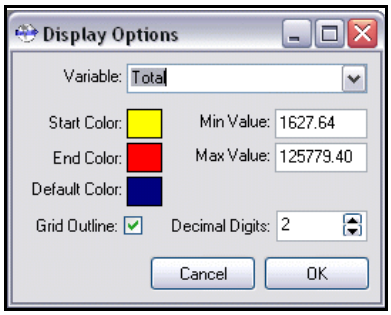
Start Color and **End Color**. These are the colors that represent the gradations in the selected variable. BenMAP uses 10 equal-sized increments for the variable, with a gradual transition between the **Start Color** and the **End Color**. To change either color, click on the colored square and select a new color.

Default Color. This is the color is used for values that fall outside of the range of the **Min Value** and the **Max Value.**

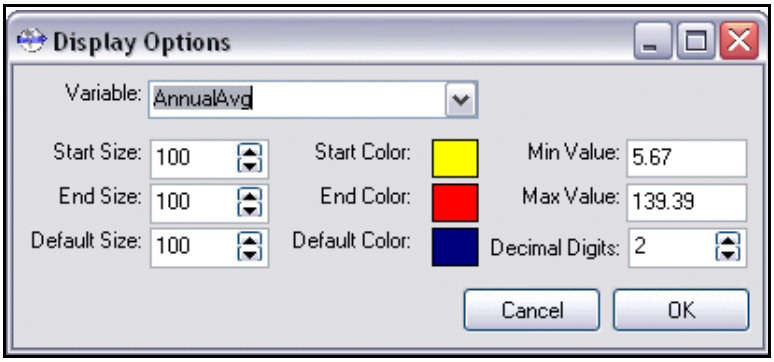
Min Value and the **Max Value.** These options define the range of the selected variable and are automatically set to the minimum and maximum of the variable. However, you may wish to enter other minimum and maximum values, such as in the case where there is one outlier that is dominating the color scale. For example, some urban grid-cells such as those in Los Angeles have extremely large values. Since BenMAP creates 10 equal-sized increments between the minimum and the maximum, it is not uncommon to have most of the grid values in just a few of the increments. If you change the minimum and/or maximum values to exclude outliers from the color scale, the outliers will then appear in the **Default** color.

Decimal Digits. This specifies the number of digits used in displaying the results. The default is two decimal points. However, for variables with values much smaller than one, such as some of the incidence rate data, you will want to increase the number of digits that you display.

Grid Outline. This appears as a fine white line around the border of each grid cell. For some grid types, this outline can make the map too complicated, and you may want to uncheck this option. An alternative is to use the **Reference Layer** drop-down window, which lets you add a blue outline for the nation, states, or counties.



Note that if you have specified point data, such as monitors, the **Display Options** window includes the ability to set the size of the points on the map. The **Start Size**, **End Size**, and **Default Size** have a default value of *100*, which you then edit. For example, if you want smaller values to be represented by smaller points, you might leave the **End Size** unchanged, and reduce the value for the **Start Size**.



8.1.2 Taskbar Buttons

There are a number of standard buttons used in most map viewing programs which you can use to navigate and customize the map view.

Open a file. Use this to open maps for viewing.

Save active layer to file. Creates a shape file for use in other map-viewing programs. The active layer is the topmost *visible* layer.

Zoom to full extent. Allows you to view the whole map that you are viewing.

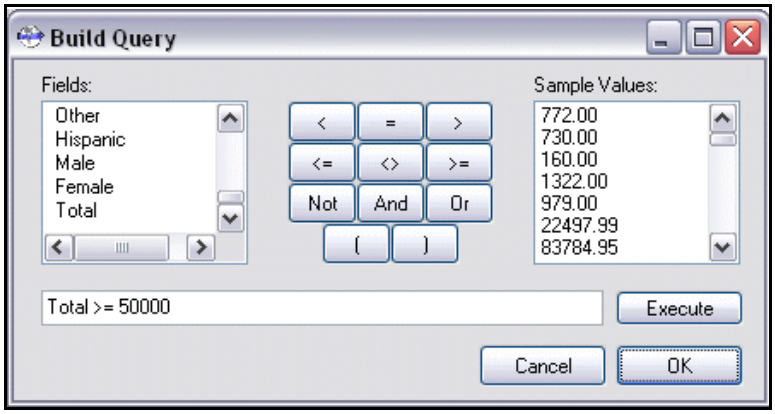

Increase zoom and **Decrease zoom.** Allows you to zoom in and out.

Select a region for zooming. Allows you to select a region to view.

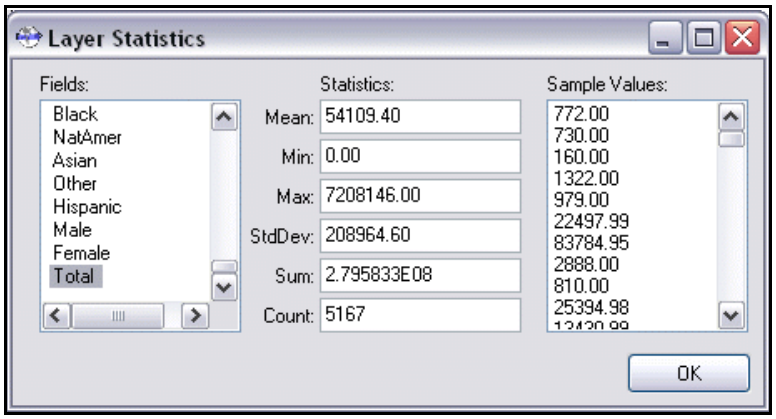
Drag mode. Allows you to manually move the map by clicking and dragging.

Click to display info for the cell under the mouse. Allows you to display info (all the variable values) for individual cells or points by clicking on them.

Build Query. Allows you to view grid-cells that satisfy certain criteria. Hitting the **Execute** button will produce a map of the cells that meet the criteria that you have specified. Below is an example of how you might use this function.



Layer Statistics. Provides information about the active layer. In the **Fields** section simply choose the variable of interest, and BenMAP will display statistics and sample values for that variable.

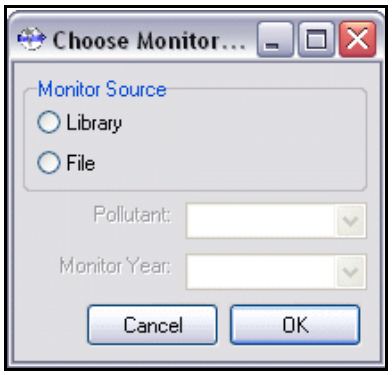


The drop-down menu provides alternative projections for displaying the data, with the default set to *Albers Equal Area Conic.*

8.1.3 Mapping Different File Types with the Tools Menu

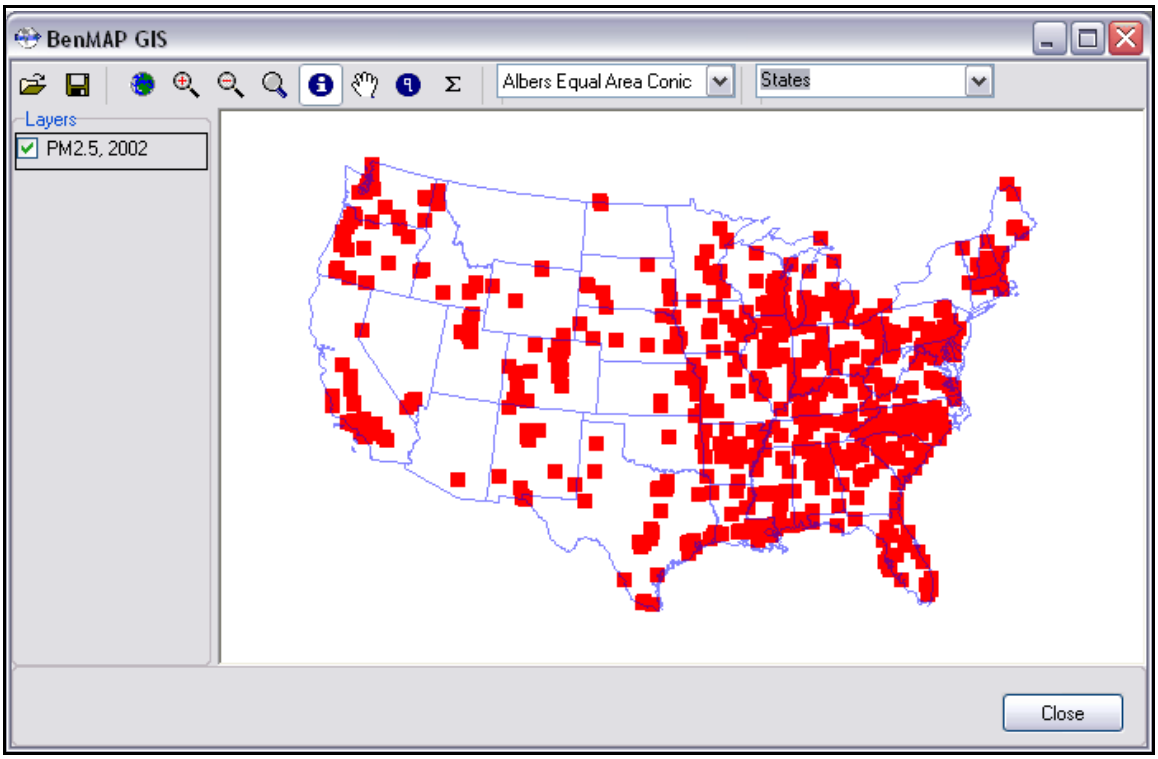
The **Open a file** button gives the option to open a variety of file and data types. All of the options are straightforward, though some require a few more steps than others. See Exhibit 8-1 above for information about each type. Below we give some step-by-step examples for some of the types.

Example 1: Mapping Monitor Data

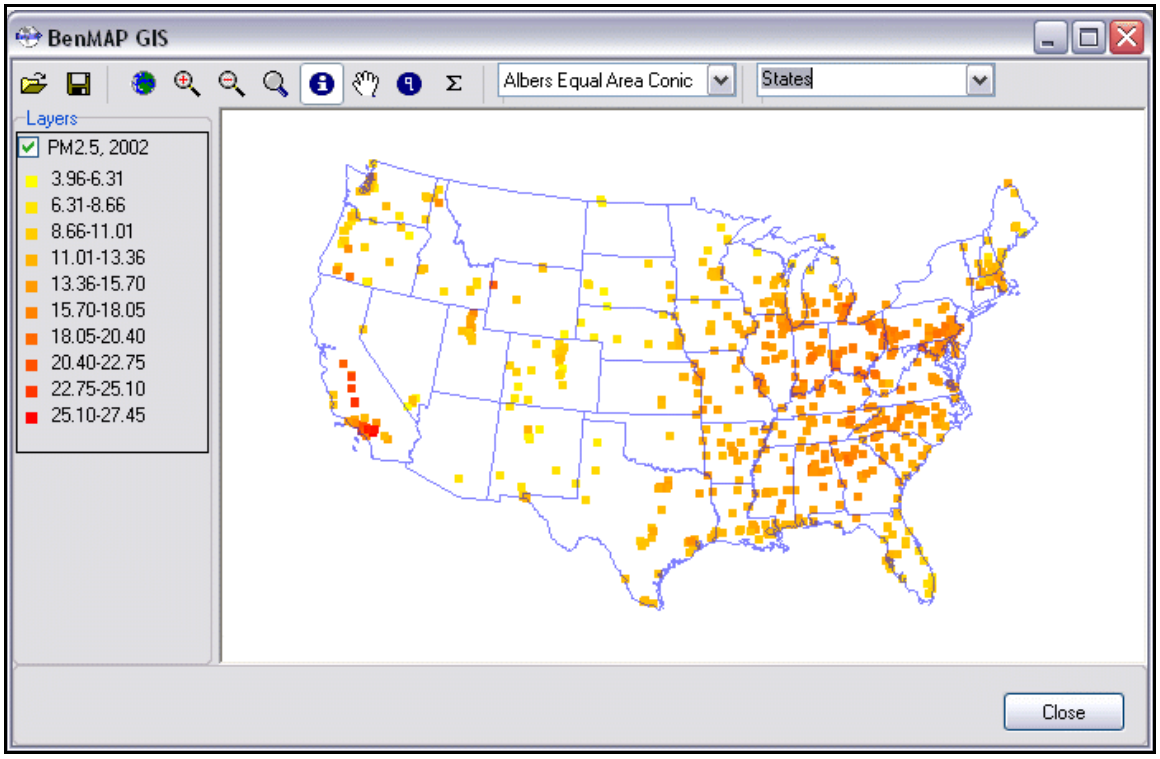


You can map monitor locations and concentration levels using the *Monitors* option. Click the **Open a file** button and select *Monitors* from the drop-down menu. A small window will appear in which you can select a **Monitor Source**. You can map data from the *Library* that comes with BenMAP, or you may map your own monitor data, so long as it follows the format in Exhibit 4-3.

If you map data from the BenMAP library, you need to specify the **Pollutant** and the **Monitor Year**. Click **OK**. BenMAP will then present the **Advanced Options**, where you need to choose the filtering options that you want to use (See Section 4.4 for more information on filtering). Click **Go!**, and BenMAP will filter the data. Then click **OK**, and BenMAP will generate a monitor layer that you can view.



In this map, each red square is a monitor location. To see the monitor values displayed with colors varying by the level of the monitor, double-click on the layer on the left side of the map, and follow the steps outlined above for setting the display options. Values shown using the *Monitors* mapping option are annual average values ($\mu g/m^3$ for PM and ppb for ozone).



Example 2: Mapping County Data

You can also map county incidence rates and demographic characteristics. Included are incidence rates by age and race in some cases, total population and population by age, and some housing and income variables.

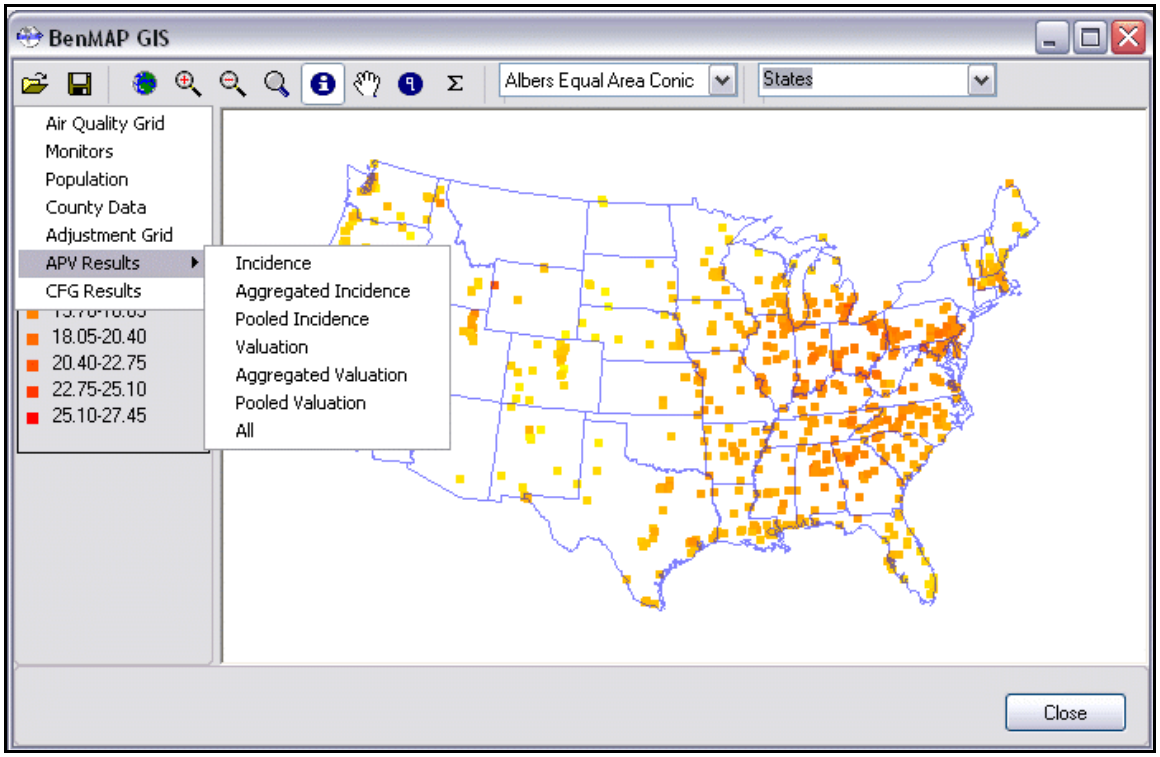
Select *Mapping / GIS* from the **Tools** menu on the main BenMAP screen. Click the **Open a file** button and select *County Data* from the drop-down menu. A map of the continental United States will appear, with all of the counties filled with gray. In the **Layers** field on the left side of the screen, double-click on **CountyIncidenceData** label. In the next screen, you can choose the variable of interest and your display options. See Appendix E for more information on incidence data, including sources.

Variables for County Incidence Data are shown in alphabetical order by variable name, not by type; for instance, *total population* is not grouped with other population variables.

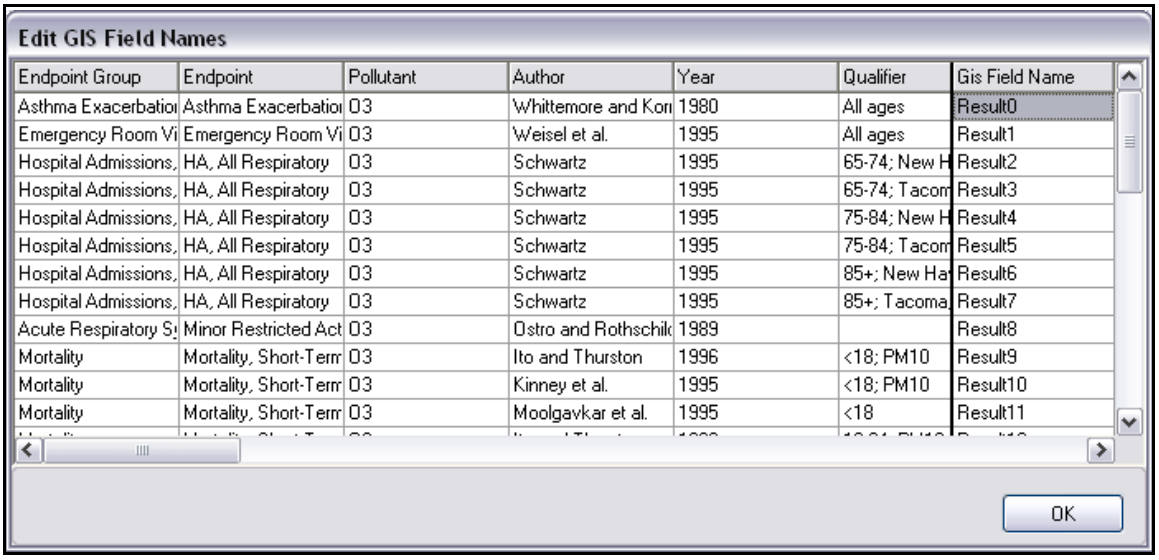
Example 3: Mapping Aggregation, Pooling, and Valuation Results

You can map each of the six types of results stored in *APV Results* files (*.apvr): *(1) Incidence, (2) Aggregated Incidence, (3) Pooled Incidence, (4) Valuation, (5) Aggregated Valuation,* and *(6) Pooled Valuation.* Finally, you can select *(7) All*, if you wish to map all of the types of results simultaneously. All of the options use the information contained in the **Aggregation,**

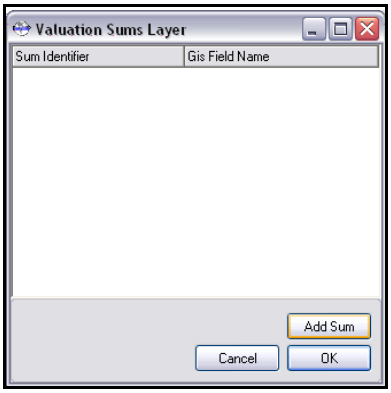
Pooling and Valuation Results files (*.apvr), which you can create using the **Aggregation, Pooling, and Valuation** button on the main screen.



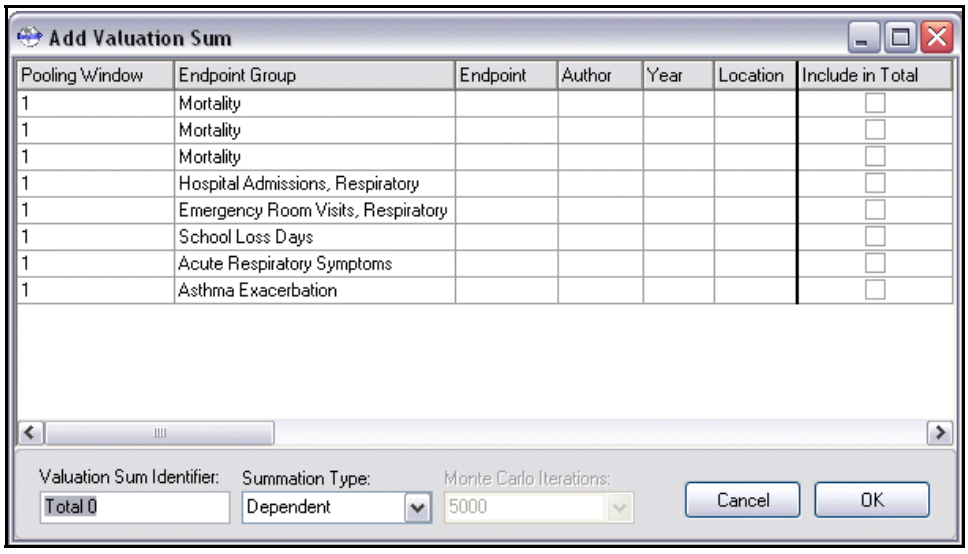
Click the **Open a file** button and select *APV Results* and one then one of the six results types from the drop-down menu. A window will appear with identifiers for each of the variables. On the right-hand side is the **GIS Field Name**. Mapping programs typically have a default length of 10 characters for variable names, so it is necessary to specify your own names or to use the default name given in the **GIS Field Name** column.



For each of the three valuation options (*Valuation, Aggregated Valuation,* and *Pooled Valuation*), a second window appears that allows you to specify variables that you want to add together (if you have selected *Incidence*, *Pooled Incidence*, or *Aggregated Incidence*, BenMAP will go directly to loading the new layer).



Click on the **Add Sum** button. Then in the far-right column, **Include in Total**, check the variables that you want to include in the total. In the lower-left corner, give a name (no longer than 10 characters) for the total in the **Valuation Sum Identifier**, and then choose whether to use a *Dependent* or *Independent* sum. If you choose the latter, then you also need to choose the number of Monte Carlo draws. To finish, click **OK**. You may repeat this step as many times as desired.



Example 4: Mapping Multiple Layers of Data

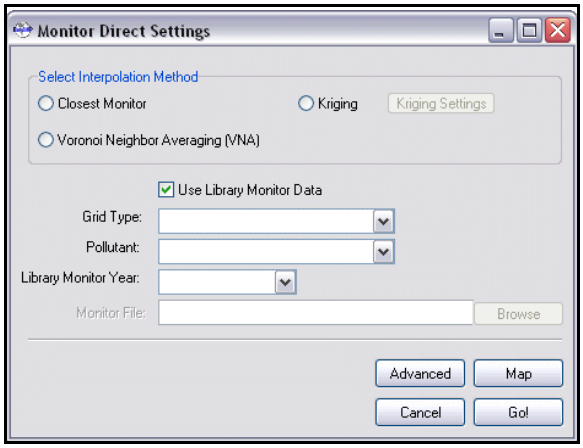
It is possible to map multiple layers simultaneously. The layer that you have opened most recently appears on the top of the list, and in the map its values lie on top of the other layers. By right-clicking on any given layer, you can move its position within the list (select *Move Up* or *Move Down*). By checking the box to the left of the layer name, you can turn a layer's visibility off and on. For instance, if you want to see the second layer in the list, simply un-check the box next to the first layer in the list. The second layer will then be on top and be visible. To delete a layer, right-click on the layer name and select *Delete*.

8.2 Viewing Maps in a BenMAP Analysis

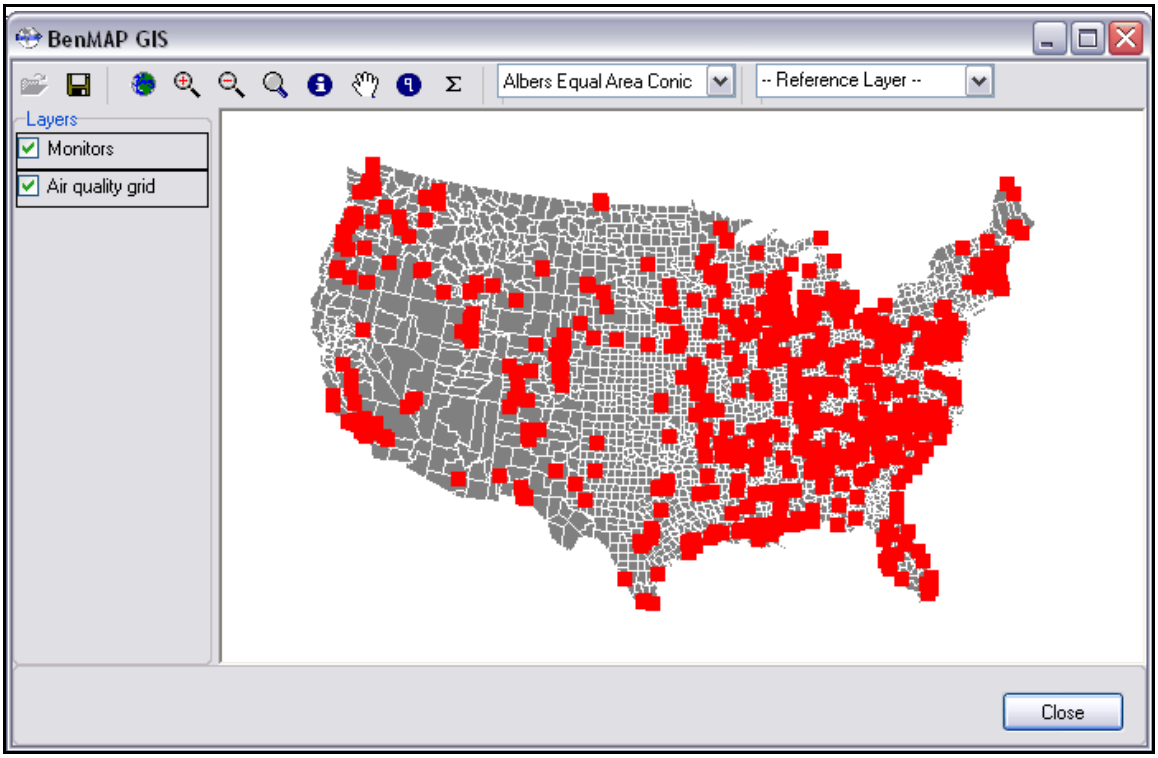
In addition to the **Tools** menu, BenMAP provides several places where you can map the input data that you are using for a particular analysis. We provide examples of these mapping options below.

Example 5: Mapping Monitor Direct Air Quality Grids

When generating air quality grids there is an option to map the grid and the data used to create it. For example, when generating a *Monitor Direct* grid, click on the **Map** button on the **Monitor Direct Settings** page.

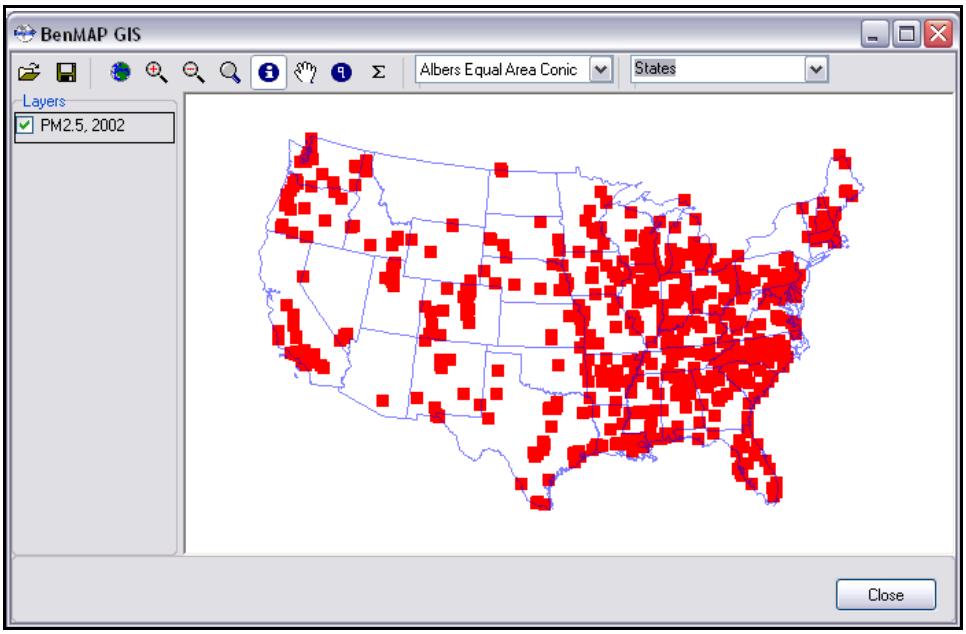


BenMAP will then generate the grid, generate a map with both the monitors and the monitor data interpolated to the grid cells. You can then use the display options to generate the map you desire.

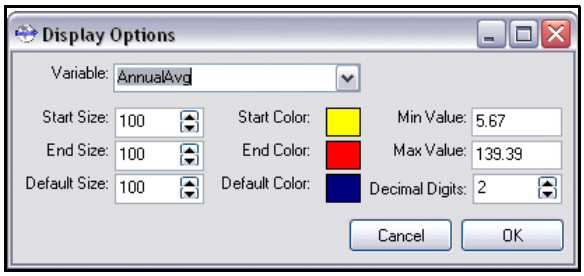


Example 6: Mapping Air Quality Monitors from the Advanced Monitor Filter

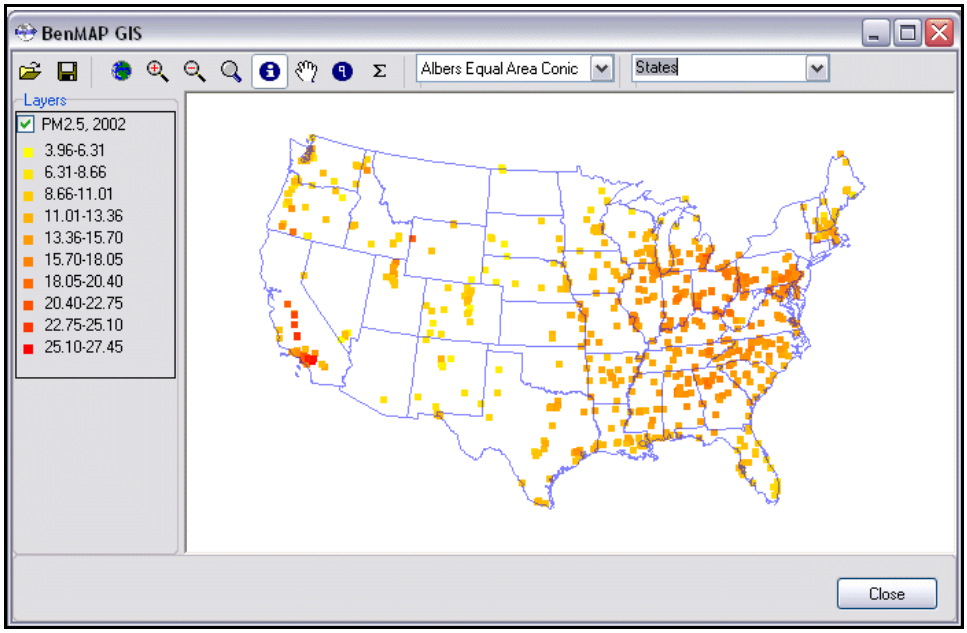
You can access mapping through the advanced monitor filter (discussed in Section 4.4) when generating air quality grids. Click on the **Create Air Quality Grids** button and choose either *Monitor Direct, Monitor and Model Relative,* or *Monitor Rollback*. On the **Settings** page, click on the **Advanced** button, which brings up the **Advanced Options** page. Choose your filtering options and click **Go!** Then click the **Map** button, which becomes active after filtering. An initial map appears with each monitor location identified by a red square. To immediately provide some context, you may want to choose, say, the **States** reference layer.



Double-click on the layer to the left of the map. Choose the display options that you would like, and then click **OK**.

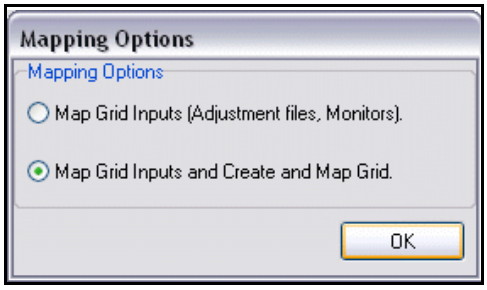


A map with your display options will then appear. You may then use the various mapping options available, such as zooming in to an area of interest, getting information on particular monitors, querying the map to display monitors with certain characteristics, and saving the map as a shapefile and viewing it in another map viewer.



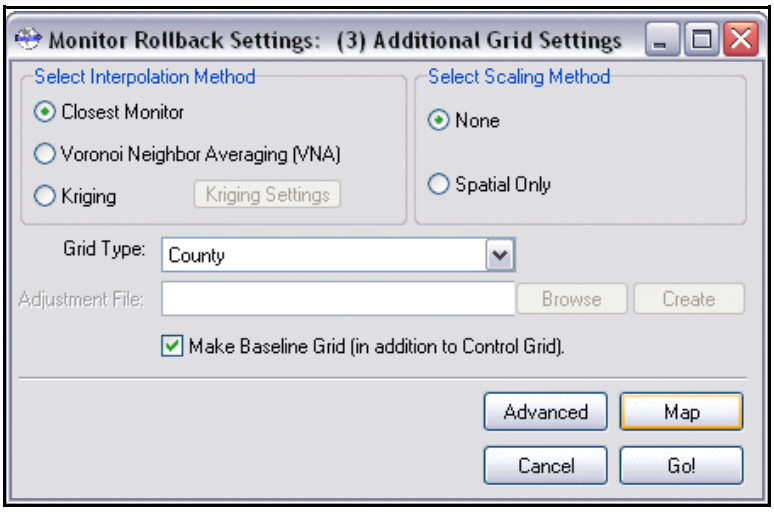
Example 7: Mapping Monitor Rollback Inputs and Outputs

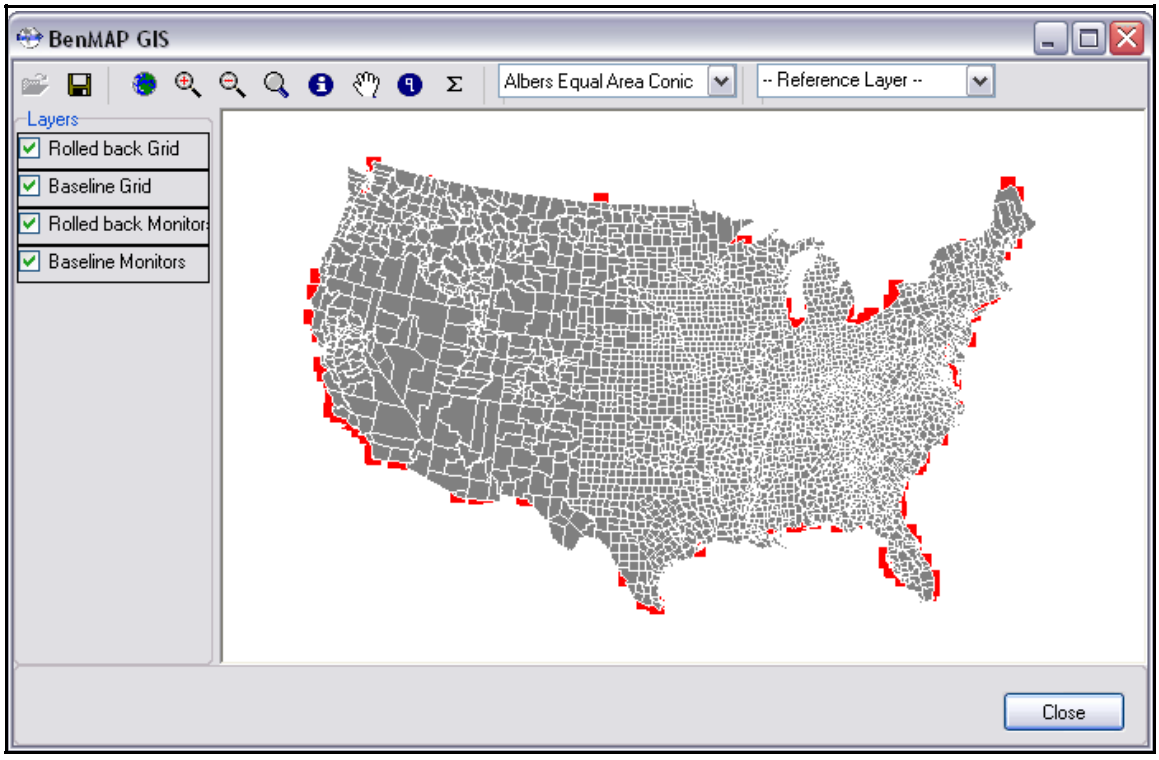
When using the *Monitor Rollback* option, you can generate two types of maps. First, you can map the inputs to the rollback – the monitor data and any spatial adjustment file that you may have used. Second, you can map the inputs and at the same time map the grid based on these inputs. To start, click the **Map** button.



You will then have the option to map the inputs, or map the inputs and the resulting grids.

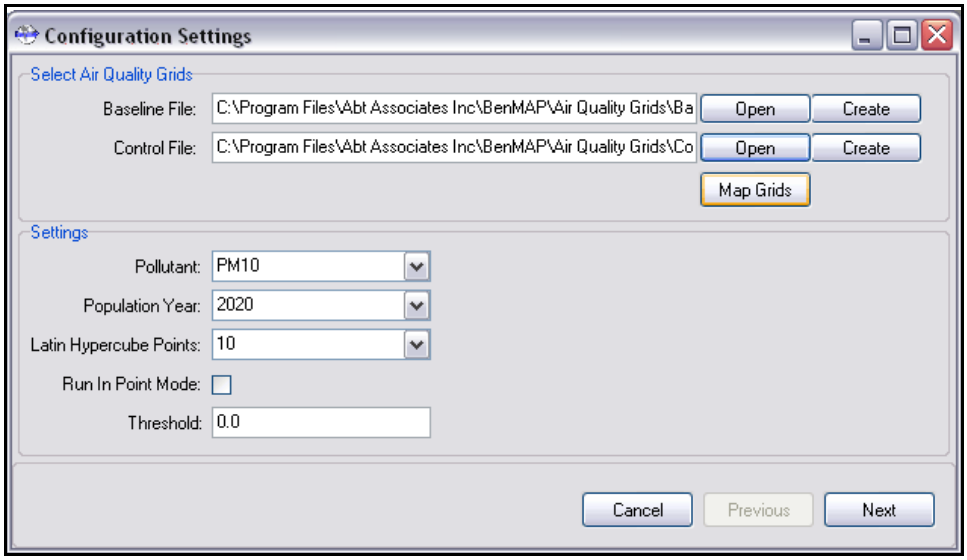
In this example, BenMAP will produce a map with both the inputs, as well as the baseline and control grids. (The baseline grid gets created because we checked the box for *Make Baseline Grid (in addition to Control Grid*.) As before, you may use the various display options to generate the map that you desire. Recall that the topmost layer lies on top of the other mapped layers. To change the ordering of the layers, simply right-click on the layer that you want to move (up or down).


Monitor Rollback Settings: (3) Additional Grid Settings
Select Interpolation Method
Closest Monitor
Voronoi Neighbor Averaging (VNA)
Kriging
Kriging Settings
Select Scaling Method
None
Spatial Only
Grid Type:
County
Adjustment File:
Browse
Create
Make Baseline Grid (in addition to Control Grid).
Advanced
Map
Cancel
Go!

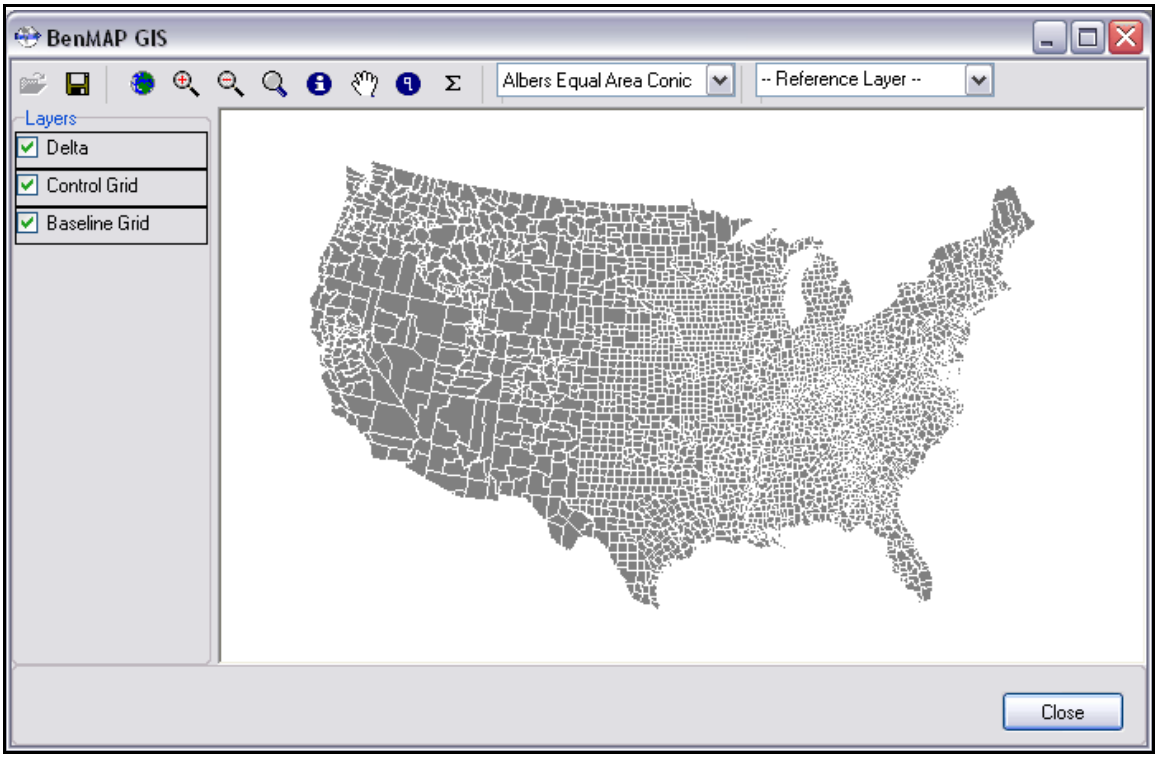

BenMAP GIS
Albers Equal Area Conic
-- Reference Layer --
Layers
Rolled back Grid
Baseline Grid
Rolled back Monitors
Baseline Monitors
Close

Example 8: Mapping Air Quality Deltas

When developing a *Configuration* file (with the *.cfg extension), you can map the baseline and control air quality grids, as well as the difference between the two. In the *Configuration Settings* window, after choosing the grids that you want to use, click on the **Map Grids** button.



This will generate layers with the delta (baseline minus control), baseline, and control values for each grid cell.



8.3 Questions Regarding Mapping

➢Why is the *Open a File* menu button disabled?

This happened because you did not use the **Tools** button and choose *Mapping / GIS*. When viewing maps while generating air quality grids, filtering monitor data, and other activities within BenMAP you do not have access to other types of maps. This is to avoid too many competing activities.

➢All of the mapped values have the same color. How do I avoid this?

This can happen when the values are extremely small and you have not specified a sufficient number of decimal points. Go to the *Display Options* window, and change the *Decimal Points*. This can also happen when one of the grid cells is an outlier, with either a very low or very small value. You can go to the *Display Options* window, and change either the *Min Value* or the *Max Value*. Finally, this can happen if you have mapped national data. In this instance, you should expect all areas to have the same color, since there is only a single national number for display, such as when mapping national results, or mapping incidence rates that do not vary by region (e.g., MRAD incidence rate).

➢Can I map air quality for individual days?

No. BenMAP only maps annual averages. In the case of hourly metrics, such as the one-hour daily maximum for ozone, BenMAP will map the average of the metric for the available days.

➢Can I print maps in BenMAP?

No. BenMAP does not currently allow printing directly from the program. However, you can export shapefiles, and then read these shapefiles into a program that does support printing, such as ArcView.

CHAPTER 9

Viewing and Editing C-R and Valuation Functions

In this chapter...

➢ View EPA Standard and User-defined C-R and valuation functions.

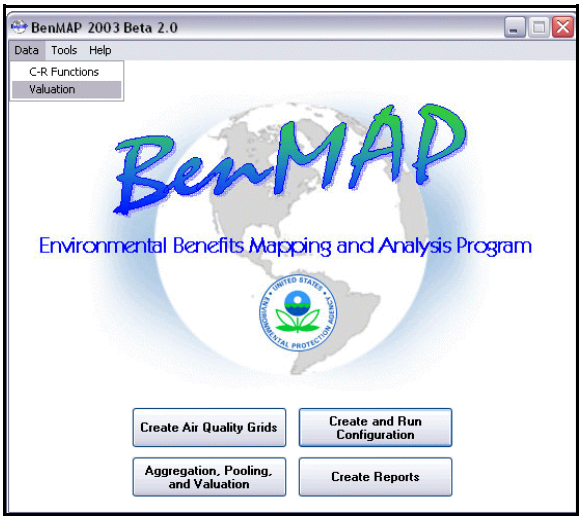
➢ Add and edit User-defined C-R and valuation functions.

➢ Find definitions for variables used in C-R and valuation functions.

Chapter Overview

9. Viewing and Editing C-R and Valuation Functions

You can view the EPA Standard (or EPA Approved) C-R and valuation functions by using the **Data** drop-down menu on BenMAP's main screen. The EPA Standard functions cannot be edited, but you can create your own database of user C-R and valuation functions which you can add to, view and edit.



9.1 C-R Functions

BenMAP includes a large number of C-R functions that quantify the relationship between the change in exposure to air pollution and the adverse effects caused by specific pollutants. In particular, BenMAP has C-R functions for ozone, $PM_{2.5}$, PM_{10}, and PMC. See Appendices F and G for details on the functions and their derivations.

The C-R functions are organized in the left part of the display into two tree views - **EPA Standard** and **User**. Within each tree view the C-R functions are organized by **Endpoint Group**, then by **Endpoint**. C-R functions are represented within these groups by their **Study ID** value, which combines the **Author**, **Year**, and **Qualifier** fields.

A C-R (Concentration-Response) Function calculates the change in adverse health effects associated with a change in exposure to air pollution. A typical C-R function has inputs specifying the air quality metric and pollutant, population characteristics, and the incidence rate of the health effect.

9.1.1 Viewing C-R Functions

To view a C-R function, locate and highlight it in the left tree view. Note that when an EPA Standard C-R function is highlighted in this manner, the right side of the display is disabled. As mentioned above, the EPA Standard functions cannot be edited. When a User C-R function is highlighted, the right side of the display is enabled, and all values of User C-R functions can be edited. Editing User-defined C-R functions is described below.

Exhibit 9-1 presents brief descriptions of the variables shown for each C-R function.

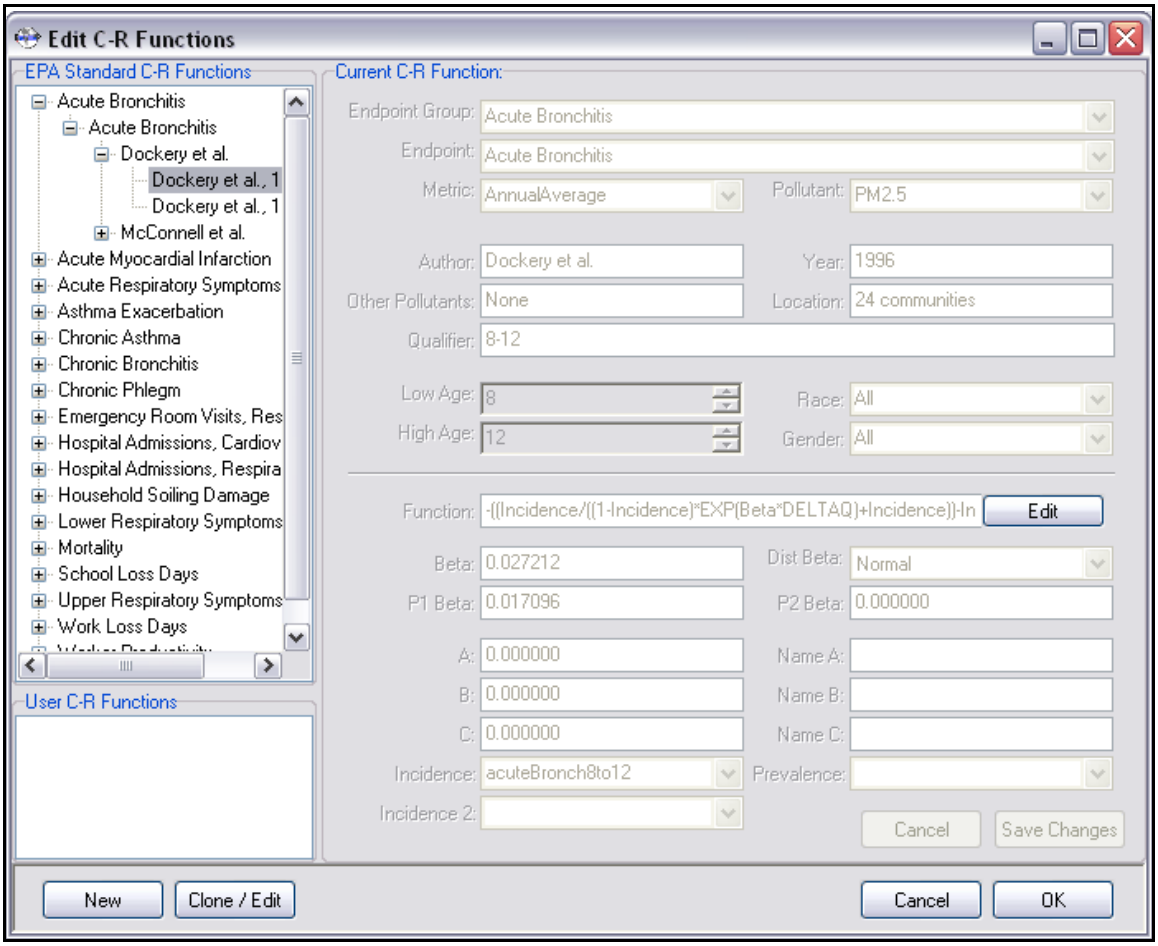


Exhibit 9-1. Selected Variables in the C-R Function Database

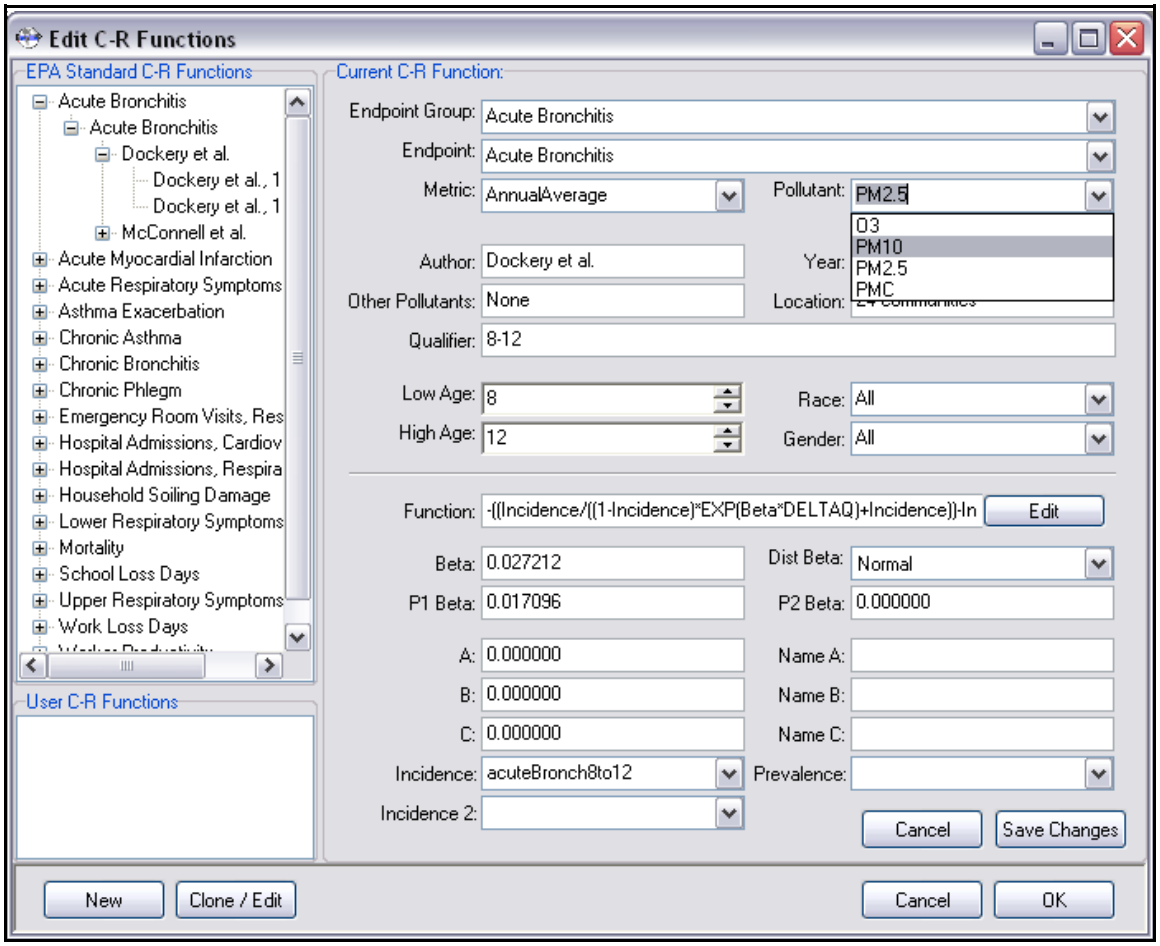
Variable	Description
Endpoint Group	An endpoint group represents a broad class of adverse health effects, such as premature mortality, chronic bronchitis, and respiratory hospital admissions.
Endpoint	An endpoint represents a relatively small class of adverse health effects, such as hospital admissions for pneumonia. In some cases the endpoint and the endpoint group are the same.
Metric	Air quality metric used in the C-R function.
Pollutant	Pollutant used in C-R function to estimate adverse effect.
Author	Author of the study used to develop the C-R function.
Year	Year of the study used to develop the C-R function.
Other Pollutants	Other pollutants that were simultaneously included in the original study used to develop the C-R function.
Location	Location of the study used to develop the C-R function.
Qualifier	Description that uniquely identifies a valuation function, when combined with the endpoint group, endpoint, and age.
Low Age	Lower bound of the age range included in the C-R function.
High Age	Upper bound of the age range included in the C-R function.
Race	Race of the population in the C-R function: American Indian, Asian American, Black, Other, and White.
Gender	Gender of the population used in the C-R function.
Function	Function refers to the form of the C-R function.
Beta	Variable for C-R function.
Dist Beta	Distribution for the variable Beta. The distribution may be chosen from a drop-down menu.
P1 Beta	Parameter used to describe distribution of Beta.
P2 Beta	Parameter used to describe distribution of Beta.
A	User-defined scalar.
Name A	User-defined description for scalar A.
B	User-defined scalar.
Name B	User-defined description for scalar B.
C	User-defined scalar.
Name C	User-defined description for scalar C.
Incidence	Number of adverse effects per unit time per person.
Incidence2	Number of adverse effects per unit time per person. In some instances, two incidence rates are required, such as when calculating cardiovascular hospital admissions minus the contribution of myocardial hospital admissions.
Prevalence	Percentage of the population subject to being affected by adverse health effect (e.g., asthmatic population subject to emergency room visits for asthma).

9.1.2 Adding C-R Functions

There are two approaches to adding C-R functions: (1) cloning existing C-R functions, and (2) adding completely new C-R functions.

Cloning C-R Functions

You can clone, or copy, both EPA-Standard and User C-R functions in the same manner. Cloning is a quick way to make minor changes to existing C-R functions. To clone a particular function, first select the desired function as you would to view it (see above). On the bottom-left of the screen, click the **Clone / Edit** button - the right side of the screen should now be enabled for editing. Edit the values you wish to change (see below for information on editing a C-R function), and then click the **Save Changes** button. This will save your newly cloned and edited C-R function to the User C-R Functions database. Alternatively, click the upper **Cancel** button to cancel your changes without saving them to the User C-R Functions database.



Adding New C-R Functions

You can add completely new C-R functions by clicking the **New** button on the bottom-left of the screen. The right side of the screen should then be enabled for editing, but all the values will be blank. You can now fill in values for each field (certain fields *must* have values - **Endpoint Group**, **Endpoint**, **Metric**, **Pollutant**, etc.) and click **Save Changes** or **Cancel**, as with **Clone / Edit**.

9.1.3 Editing C-R Functions

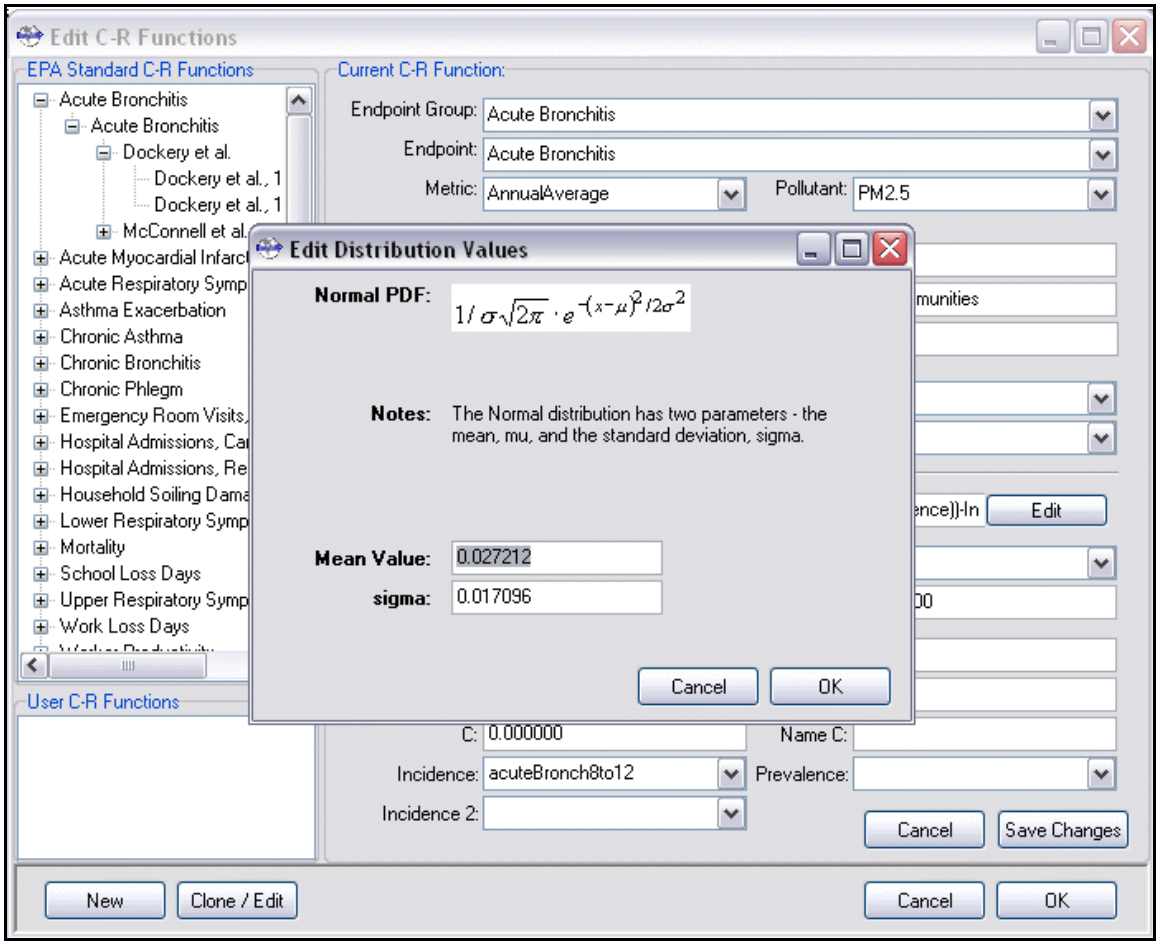
You can edit C-R functions in various ways - by adding new functions (see above for the two methods of adding new functions), or by selecting functions in the User C-R Functions database and editing their values. The following section applies to functions edited in any of these ways. When done editing, simply click **Save Changes** or **Cancel** to save the new/modified C-R function to the User C-R Functions database or cancel changes, respectively.

Many C-R function variables can be edited directly by typing into the appropriate text boxes or selecting values from the appropriate drop down lists. Examples of text boxes include the **Author** and **Qualifier** variables. Examples of drop down lists include the **Endpoint Group** and **Pollutant** variables.

The **Beta** variable and its associated variables, **Dist Beta**, **P1 Beta**, and **P2 Beta** require some special editing, as does the **Function** variable. Users are advised to familiarize themselves with Appendix D before editing these variables.

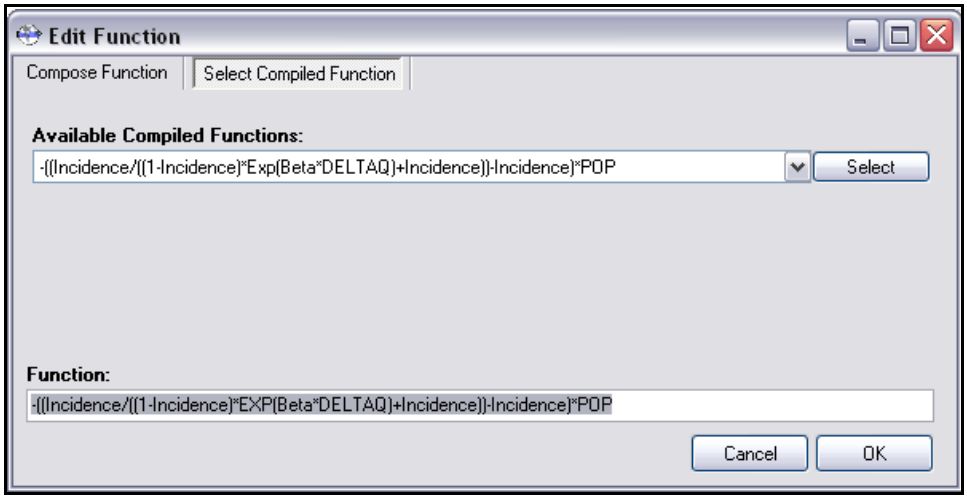
The C-R function can have a single variable, **Beta**, which has a distribution associated with it. This distribution is specified by the **Dist Beta** variable, and can take on sixteen different distribution types: *None* (no uncertainty), *Normal*, *Triangular*, *Poisson*, *Binomial*, *Log Normal*, *Uniform*, *Exponential*, *Geometric*, *Weibull*, *Gamma*, *Logistic*, *Beta*, *Pareto*, and *Cauchy*.

Each of these distributions takes different numbers and types of parameters which need to be put into the **P1 Beta** and **P2 Beta** variables. To facilitate this, BenMAP includes special dialogs for each distribution type which contain the probability distribution function of the distribution, some notes about the distribution type, and editable text boxes through which the parameters of the distribution can be set. When you click **OK** after filling in values for each parameter in any of these dialogs, BenMAP fills in the appropriate values for **P1 Beta** and **P2 Beta**. Note that **P1 Beta** and **P2 Beta** cannot be edited directly.

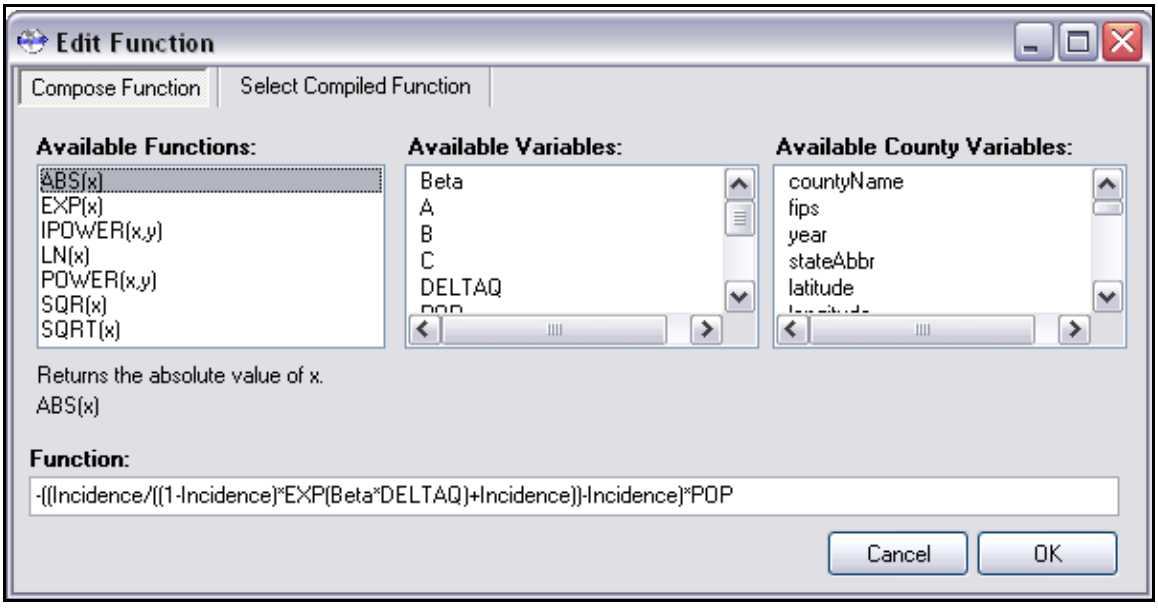


The **Function** variable can be complicated to edit, so BenMAP includes a special function editing dialog which can be accessed by clicking the **Edit** button to the right of the **Function** text box. In general the function should be a mathematical expression which calculates the adverse effects associated with a change in air quality. You can either select from a set of previously "compiled" functions, or you can compose your own function. Choosing from among the 25 or so compiled functions is the preferred option, since already compiled functions can be processed by BenMAP significantly faster.

Note that if you make any changes to the compiled functions (adding variables, deleting variables, etc.) your function will no longer be treated as compiled and will not have the speed advantages associated with the compiled functions. Similarly, if you manually type in a function which looks just like a compiled function it will not be treated as compiled.



If you choose to compose your own function string, you have a great deal of flexibility in choosing from functions, variables, operators, and numbers. The functions available to be used in the **Function** expression are listed in the leftmost of the upper windows in the function editing dialog. They include ABS (absolute value), EXP (*e* to the *x*), IPOWER (*x* to the power *y*, *y* an integer), LN (natural logarithm), POWER (*x* to the power *y*, *y* a real number), SQR (square), and SQRT (square root). To see a description of one of the **Available Functions**, simply highlight it - a description will appear below the **Available Functions** window. To insert the function into the **Function** text box at the current cursor position, simply double-click it.



The variables available to be used in the **Function** expression are listed in the two other upper windows in the function editing dialog, **Available Variables** and **Available County Variables**.

The variables in the **Available Variables** window are the main variables used in **Function** expressions, and are described in Exhibit 9-2.

Exhibit 9-2. Available Variables

Variable	Description
Beta	The variable Beta specifies the C-R Function Database entry variable Beta, with its associated distribution and parameters. As such, during a point-mode run this variable simply takes on the value specified for Beta in the C-R Function Database. During an uncertainty run, however, this variable takes on the various Latin Hypercube values from the distribution specified by Dist Beta, as described in Chapter 5.
A	The variable A specifies the C-R Function Database entry variable A (a user-defined scalar).
B	The variable B specifies the C-R Function Database entry variable B (a user-defined scalar).
C	The variable C specifies the C-R Function Database entry variable C (a user-defined scalar).
DELTAQ	The variable DELTAQ represents the air quality delta (defined as baseline minus control), as defined by the C-R Function Database entry Metric variable. This variable takes on different values for each grid cell processed.
POP	The variable POP represents the population, as specified by the C-R Function Database entry variables Low Age, High Age, Race, and Gender. This variable takes on different values for each grid cell processed.
Incidence	The variable Incidence represents a single county incidence variable (see below for details on county variables), and is mainly present for readability. It is replaced at runtime with the county incidence variable specified by the C-R Function Database entry variable Incidence.
Q0	The variable Q0 represents the control air quality value. This variable takes on different values for each grid cell processed.
Q1	The variable Q1 represents the baseline air quality value. This variable takes on different values for each grid cell processed.

The variables in the **Available County Variables** window are treated differently from the **Available Variables**. These variables are present in a database, and take on different values for each county in the United States. This allows BenMAP to calculate more accurate estimates of adverse effects. Exhibit 9-3 contains a selection of variables from the **Available County Variables** and associated descriptions.

Exhibit 9-3. Selection of Available County Variables

Variable	Description
hospAllCardioUnder18	This variable represents the county-specific count of hospitalizations for cardiovascular disease per person under the age of 18 per day.
mortAllCause25Up	This variable represents the county-specific count of all cause mortality per person over the age of 25 per year.
count_farm_employed	This variable represents the county-specific count of persons employed on farms.
pctAsthma5to17Black	This variable represents the county-specific percentage of total asthma cases which affect black persons ages 5 to 17.

For a given grid cell, BenMAP calculates a value for each county variable in the following manner:

➢ BenMAP has a database which contains the percentages of each grid cell's population which comes from each county. For example, the REMSAD grid cell (79, 26) gets 2.5% of its population from county 01085, 47% from county 01047, and 50.5% from county 01001.
➢ For each county which contributes population to a grid cell, BenMAP multiplies the percentage of the population contributed by that county by that county's value for the county variable.
➢ The value for the grid cell is then the sum of these population-weighted, county-specific values.

To see a description of any of the variables (both **Available Variables** and **Available County Variables**), simply highlight it - a description will appear below the appropriate window. To insert any of the variables into the **Function** text box at the current cursor position, simply double click it.

9.2 Valuation Functions

BenMAP has a large number of valuation functions to estimate the economic value of adverse effects. In particular, BenMAP has valuation functions for most of the **Endpoint Groups** for which C-R functions exist, and for many of the specific Endpoints as well. See Exhibit 5-1 for a list of the **Endpoint Groups** and **Endpoints** used in BenMAP.

The valuation functions are organized in the same way as the C-R functions (see above) - an **EPA Standard** tree view and a **User** tree view. Again, within each tree view the valuation functions are organized by **Endpoint Group**, then by **Endpoint**. Valuation functions are represented within these groups by their **Qualifier** and age range values.

Valuation functions can be viewed, added, and edited in exactly the same way which C-R functions are viewed, added, and edited (see above). The few differences will be highlighted below.

Valuation functions have different variables than C-R functions. The variables present in the Valuation Functions database are described in Exhibit 9-4.

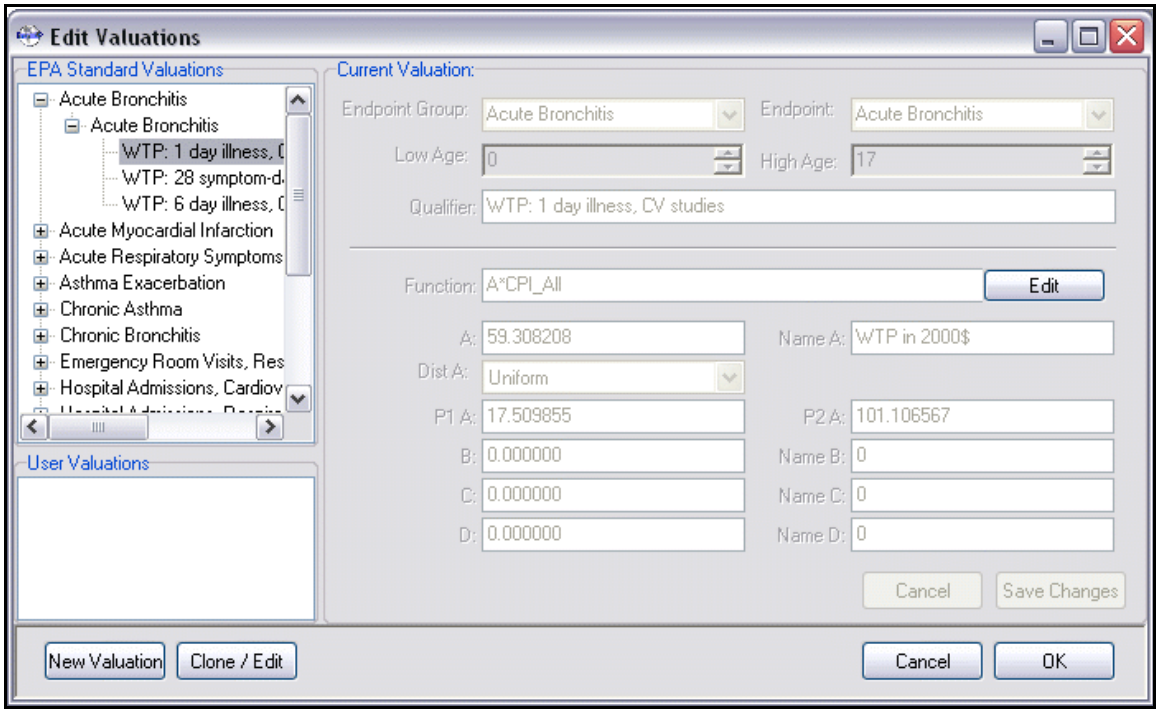
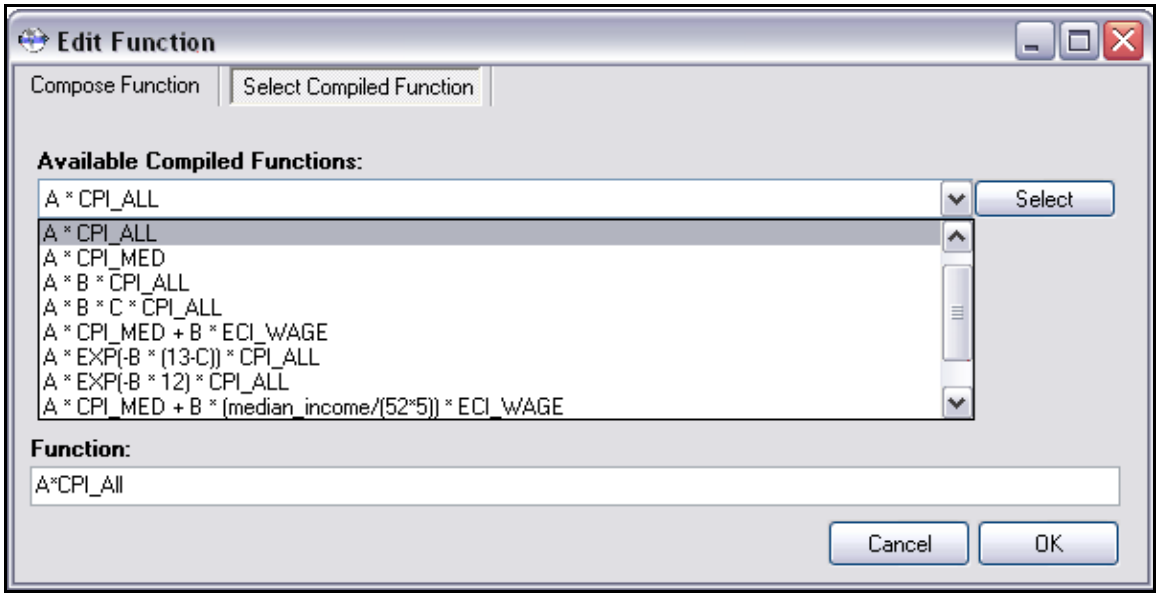


Exhibit 9-4. Selected Variables in the Valuation Database

Variable	Description
Endpoint Group	An endpoint group represents a broad class of adverse health effects, such as premature mortality, chronic bronchitis, and respiratory hospital admissions.
Endpoint	An endpoint represents a relatively small class of adverse health effects, such as hospital admissions for pneumonia. In some cases the endpoint and the endpoint group are the same.
Low Age	Lower bound of the age range included in the valuation function.
High Age	Upper bound of the age range included in the valuation function.
Qualifier	Description that uniquely identifies a valuation function, when combined with the endpoint group, endpoint, and age.
Function	Valuation function.
A	Variable for valuation function.
Name A	Name of varia
Dist A	Distribution for the variable Beta. The distribution may be chosen from a drop-down menu.
P1 A	Parameter used to describe distribution of Beta.
P2 A	Parameter used to describe distribution of Beta.
B	User-defined scalar.
Name B	User-defined description for scalar B.
C	User-defined scalar.
Name C	User-defined description for scalar C.
D	User-defined scalar.
Name D	User-defined description for scalar D.

When editing the **Function** variable of a Valuation Function database entry, an additional tab is available - **Select Compiled Functions**. These functions are built into BenMAP and run much more quickly than custom functions (which are not compiled). To select a compiled function, simply click on the **Select Compiled Functions** tab, select the desired function in the drop down list, click the **Select** button, and click **OK.** If you decide to exit the function editor without saving your changes, click the **Cancel** button instead.



As with the compiled C-R functions, if you make any changes to the compiled valuation functions (adding variables, deleting variables, etc.) your function will no longer be treated as compiled and will not have the speed advantages associated with the compiled functions. Similarly, if you manually type in a function which looks just like a compiled function it will not be treated as compiled.

In the **Compose Function** tab of the function editor, almost everything looks just like the C-R function editor. The only exception is the **Available Variables** window. Exhibit 9-5 presents the variables available for valuation functions and associated descriptions.

Exhibit 9-5. Available Variables for Valuation Functions

Variable	Description
A	The variable A specifies the Valuation Function Database entry variable A, with its associated distribution and parameters. As such, during a point-mode run this variable simply takes on the value specified for A in the Valuation Function Database. During an uncertainty run, however, this variable takes on the various Latin Hypercube values from the distribution specified by Dist A, as described in Chapter 6.
B	The variable B specifies the Valuation Function Database entry variable B.
C	The variable C specifies the Valuation Function Database entry variable C.
D	The variable D specifies the Valuation Function Database entry variable D.
CPI_ALL	The variable CPI_ALL specifies an inflation factor for generic dollar figures - the specific inflation factor used depends on the Dollar Year selected (see Chapter 6 for details).
CPI_MED	The variable CPI_MED specifies an inflation factor for medical dollar figures - the specific value used depends on the Dollar Year selected.
ECI_WAGE	The variable ECI_WAGE specifies an inflation factor for wages - the specific value used depends on the Dollar Year selected.

9.3 Frequently Asked Questions Regarding Viewing and Editing C-R Functions and Valuations

➢Why can I not edit the EPA Standard functions?

BenMAP does not allow you to edit EPA Standard functions, so that you may have confidence when you use them that they are indeed the standard versions, and have not been modified or changed. If you want to use a function that is slightly different than an EPA Standard function, simply clone the function and edit it in the user database.

CHAPTER 10

Tools

In this chapter...

➢ Learn about the Tools menu.

➢ Use the Adjustment Factor Creator to generate adjustment factor files.

➢ Use the Neighbor File Creator to extract information on neighboring monitors from air quality grids.

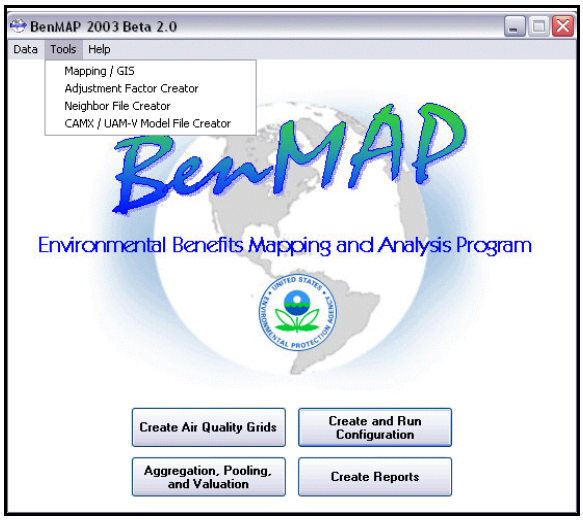
➢ Use the CAMx / UAM-V Model File Creator to create single CAMx/UAM-V model files which can be used by BenMAP.

Chapter Overview

10. Tools

The **Tools** drop-down menu gives you the ability to do tasks that are occasionally needed to perform a standard analysis, or to better understand an analysis that you have already conducted. In particular, the **Tools** menu gives you access to four options: (1) **Mapping / GIS**, (2) **Adjustment Factor Creator**, (3) **Neighbor File Creator**, and (4) **CAMx / UAM-V Model File Creator**.

10.1 Mapping / GIS

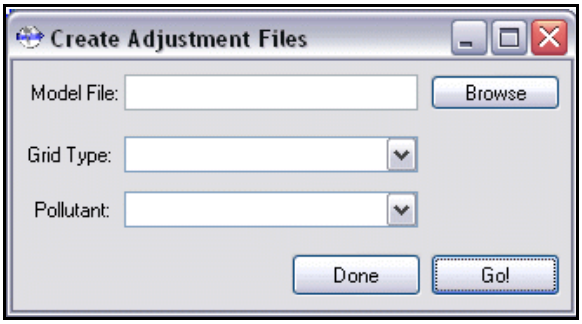


The **Mapping / GIS** option allows you to generate a wide variety of maps, including maps of monitor data, adjustment factors, air quality grids, population data, county incidence rates, and both incidence and valuation results. In addition, you can export the maps you have generated and view them in a shapefile viewer, such as Arcview. Chapter 8 provides additional details on BenMAP's mapping capabilities.

10.2 Adjustment Factor Creator

The **Adjustment Factor Creator** generates the adjustment factor files needed for the *Model and Monitor Relative* air quality grid creation. To use the **Adjustment Factor Creator**, you specify a model file, along with the grid type and pollutant of that file. Supported combinations of grid type and pollutant are:

- **REMSAD, PM_{10} / $PM_{2.5}$ / PMC**
- **CAMx, O3**
- **CMAQ, PM_{10} / $PM_{2.5}$ / PMC**



Note that these are the same combinations supported for *Model Direct* air quality grid creation.

BenMAP includes the **Adjustment Factor Creator** as a standalone tool in case you want to simply generate adjustment factor files, without generating air quality grids, or conducting other parts of a typical analysis. Note that you can also access the **Adjustment Factor Creator** during

Model and Monitor Relative air quality grid creation by clicking the **Create** button for the *Base Year Adjustment File* or *Future Year Adjustment File*. For more information on adjustment factor files, see Chapter 4.

10.3 Neighbor File Creator

The **Neighbor File Creator** can be used to extract information from air quality grids created via the *Monitor Direct* or *Model* and *Monitor Relative* methods. It generates a tab-delimited file containing the neighbors of each grid cell, where "neighbors" refers to the monitors used to generate the air quality metrics for the grid cell. Neighbors are identified by Monitor ID (see Exhibit 1-2 for a description of this value). In addition, if the air quality grid was created using *VNA* or *Kriging* interpolation (see Section 4.2 for more information on interpolation methods in BenMAP) the file will also contain the weights associated with each neighbor.

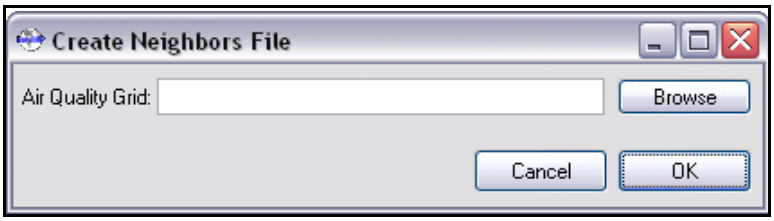
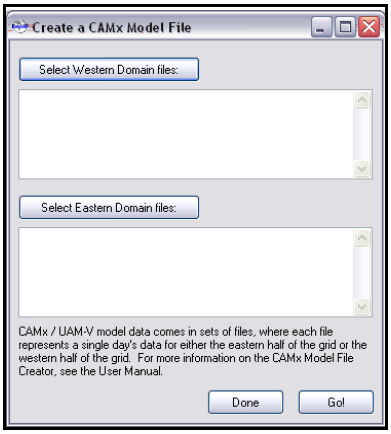


Exhibit 10-1 presents the variables that are included in the files generated by the **Neighbor File Creator.**

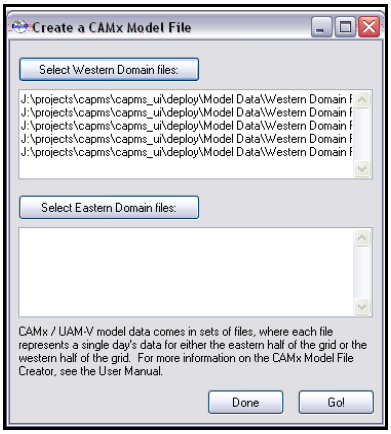
Exhibit 10-1. Variables in Data File Generated by the Neighbor File Creator

Variable	Description
Column	Column of grid cell.
Row	Row of grid cell.
Monitor ID	Monitor identifier.
Weight	Only present for grids created using *VNA* or *Kriging* interpolation. The weight used to generate air quality metrics for the grid cell for this monitor.

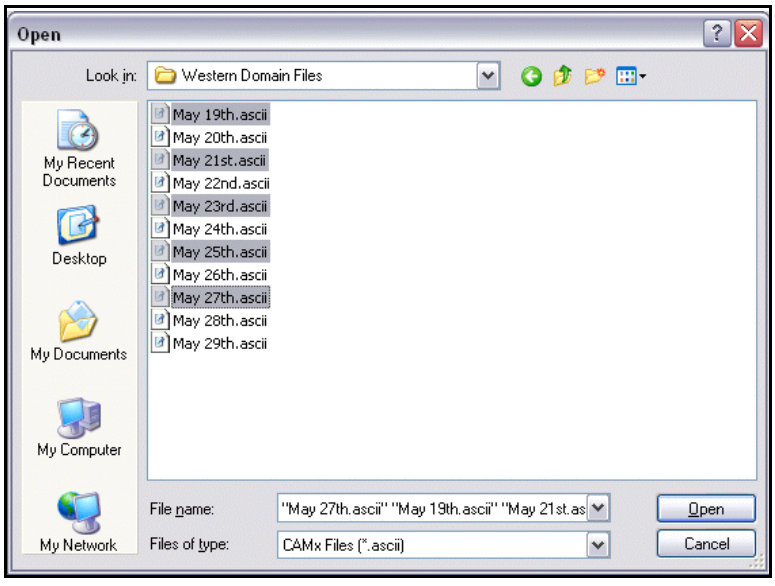
10.4 CAMx / UAM-V Model File Creator



The **CAMx / UAM-V Model File Creator** is used to create single CAMx/UAM-V model files which can be used by BenMAP in *Model Direct* air quality grid creation, or *Adjustment Factor* creation. CAMx and UAM-V data comes in a series of individual files (one file representing one day) for the Eastern United States and a separate series of files for the Western United States. BenMAP cannot read these files directly in *Model Direct* air quality grid creation or in the creation of adjustment Factors. Instead, this tool allows the selection of all Eastern domain files and all Western domain files, and produces a single CAMx/UAM-V model file (*.camx) which BenMAP can read directly.



To use the **CAMx/UAM-V Model File Creator**, you must first place all your Western domain files in a common folder and all your Eastern domain files in a common folder. You can then click the **Select Western Domain Files** button and/or the **Select Eastern Domain Files** button - both of these will bring up **Open File** dialogs. Within this dialog, select all of your Western or Eastern files by clicking on them while holding down the *Ctrl* key on your keyboard. Alternatively, you can click on files while holding down the *Shift* key if you wish to select entire ranges of files.



When you have selected all your files, click the **Open** button. All your selected files should now be listed in the window under the appropriate **Select...Domain Files** button.

At this point, you simply need to hit **Go!** and save your newly create CAMx/UAM-V model file. When you are done using the **CAMx/UAM-V Model File Creator**, hit the **Done** button.

10.5 Questions Regarding Tool Menu

➢Can I use the Neighbor File Creator on an air quality grid just based on model data, without any monitor data?

No. *Model Direct* air quality grids have no neighboring monitors, by definition. BenMAP will produce an Access Violation error if you attempt to create a neighbor file from a *Model Direct* air quality grid. If this happens to you, just click **OK** in the Access Violation dialog and try again with a *Monitor Direct* or *Monitor Model Relative* air quality grid.

Appendix A: Monitoring Data

BenMAP comes supplied with monitoring data for ozone, $PM_{2.5}$, and PM_{10} for the years 1996 through 2002. This Appendix details how you may filter monitor data for use in an analysis. In addition, this Appendix details the rollback procedures that you can perform on monitor data. The rollback procedure is a quick way to determine the monitor levels that would exist under various kinds of changes that you can specify. This includes three basic types of rollbacks: Percentage, Increment, and Rollback to Standard.

Note that the monitor data have a particular format that we simply refer to as the BenMAP format, and are derived from AMP500 and AMP501 files available from the U.S. Environmental Protection Agency. This Appendix describes the BenMAP format, and how we used AMP500 and AMP501 files to generate the BenMAP formatted files that come supplied with BenMAP. If you are interested in using your own monitor data in an analysis, you simply use this same format, and follow the directions for importing data detailed in Chapter 4. We have included the SAS (http://www.sas.com/) code used to manipulate the AMP500 and AMP501 files.

A.1 Monitoring Data Format and Variable Values

Monitoring data are available from the Air Quality System (AQS) maintained by the U.S. Environmental Protection Agency (contact: Virginia Ambrose, email: ambrose.virginia@epa.gov). In order to import the data into BenMAP, we prepare the data with a fixed-field format that applies for both hourly and daily data. Exhibit A-1 lists the names and dates of the files that we received from EPA. Exhibit A-2 lists the variables and a brief description of each, and Exhibit A-3 provides an example of what a $PM_{2.5}$ file would look like.

Exhibit A-1. Format for Air Quality Monitoring Data

Pollutant	File Type	File Name	Date Received from EPA	Description
--	AMP500	AMP500.TXT	9/16/2002	AMP500 data for 1996
		EG_500_SELPARM_1997_2002.TXT	6/19/2003	AMP500 data for 1997-2002
Ozone	AMP501	Eg44201-1996.txt	9/30/2002	AMP501 data for 1996
		Rd_501_44201_1997.txt	12/3/2002	AMP501 data for 1997
		Rd_501_44201_1998.txt	12/3/2002	AMP501 data for 1998
		Eg44201-1999.txt	9/30/2002	AMP501 data for 1999
		Eg44201-2000.txt	9/30/2002	AMP501 data for 2000
		Eg44201-2001.txt	9/30/2002	AMP501 data for 2001
		ozone-reg1-3-2002.txt	3/27/2002	AMP501 data for 2002 – part 1
		ozone-reg4-5-2002.txt	3/27/2002	AMP501 data for 2002 – part 2
		ozone-reg6-8-2002.txt	3/27/2002	AMP501 data for 2002 – part 3
		ozone-reg9-10-25-2002.txt	3/27/2002	AMP501 data for 2002 – part 4
$PM_{2.5}$	AMP501	Eg81104-1996.txt	9/30/2002	AMP501 standard data for 1996
		Eg_501_81104_pm10_nation_1997.txt	6/19/2003	AMP501 standard data for 1997
		Eg_501_81104_pm10_nation_1998.txt	6/19/2003	AMP501 standard data for 1998
		Eg81104-1999.txt	9/30/2002	AMP501 standard data for 1999
		Eg81104-2000.txt	9/30/2002	AMP501 standard data for 2000
		Eg81104-2001.txt	9/30/2002	AMP501 standard data for 2001
		Eg_501_81104_pm10_nation_2002.txt	6/19/2003	AMP501 standard data for 2002
		Eg88101-1996.txt	9/30/2002	AMP501 local data for 1996
		Eg_501_88101_pm10_nation_1997.txt	6/19/2003	AMP501 local data for 1997
		Eg_501_88101_pm10_nation_1998.txt	6/19/2003	AMP501 local data for 1998
		Eg88101-1999.txt	9/30/2002	AMP501 local data for 1999
		Eg88101-2000.txt	9/30/2002	AMP501 local data for 2000
		Eg88101-2001.txt	9/30/2002	AMP501 local data for 2001
		PM25_Continuous_2002.txt	5/5/2003	AMP501 continuous local data for 2002
		pm25_daily2002.txt	5/1/2003	AMP501 daily local data for 2002
PM_{10}	AMP501	Eg81102-1996.txt	9/30/2002	AMP501 standard data for 1996
		Eg_501_81102_pm10_nation_1997.txt	6/19/2003	AMP501 standard data for 1997
		Eg_501_81102_pm10_nation_1998.txt	6/19/2003	AMP501 standard data for 1998
		Eg81102-1999.txt	9/30/2002	AMP501 standard data for 1999
		Eg81102-2000.txt	9/30/2002	AMP501 standard data for 2000
		Eg81102-2001.txt	9/30/2002	AMP501 standard data for 2001
		Eg_501_81102_pm10_nation_2002.txt	6/19/2003	AMP501 standard data for 2002
		Eg85101-1996.txt	9/30/2002	AMP501 local data for 1996
		Eg_501_85101_pm10_nation_1997.txt	6/19/2003	AMP501 local data for 1997
		Eg_501_85101_pm10_nation_1998.txt	6/19/2003	AMP501 local data for 1998
		Eg85101-1999.txt	9/30/2002	AMP501 local data for 1999
		Eg85101-2000.txt	9/30/2002	AMP501 local data for 2000
		Eg85101-2001.txt	9/30/2002	AMP501 local data for 2001
		Eg_501_85101_pm10_nation_2002.txt	6/19/2003	AMP501 local data for 2002

Files obtained from Virginia Ambrose, email: ambrose.virginia@epa.gov.

Exhibit A-2. Format for Air Quality Monitoring Data

Variable [a]	Necessary for BenMAP [b]	Description
year	yes	The *year* is a four-digit variable giving the year of the monitoring data.
monitor ID	yes	The *monitor ID* is a 15 character description of the monitor. It includes a state FIPS code (2 characters), county FIPS code (3 characters), site ID (4 characters), pollutant parameter (5 characters), and POC code (1 character).
latitude	yes	The *latitude* should be in decimal degrees.
longitude	yes	The *longitude* should be in decimal degrees. (Note that the longitude for the United States has a negative sign.)
land use	no	Categorization of the prevalent land use within 1/4 mile of the Monitoring Site.
method	no	The *method* identifies the approach used to collect the monitor data. For example, the Federal Reference Method for $PM_{2.5}$ includes *method* codes 116-120 and 123.
location setting	no	A description of the environmental setting within which the Site is located.
probe location	no	Identification of the location of the sampling Probe
monitor objective	no	Identification of the reason for measuring air quality by the Monitor.
sample frequency	no	Indicates the scheduled elapsed time period between observations.
sample values	yes	Either 365 daily PM values or 8,760 hourly ozone values.

[a] Monitor data available from the EPA AQS (contact: Virginia Ambrose (ambrose.virginia@epa.gov). Each monitor and method represents a unique set of sample values, and occupies one line of data. BenMAP allows you to choose the desired methods, and then averages the data so that a monitor ID occupies a single line of data. In those cases where, for a given monitor ID, there is more than one land use, location setting, probe location, monitor objective, or sample frequency, then we flag the variable with a value of "ZZ." This simply means that there is more than one value for that variable for a particular monitor ID.

[b] The year, monitor ID, latitude, longitude, and sample values are necessary for BenMAP to function. On the other hand, other variables are not strictly necessary. You can filter the data using the *method*, but this variable is not strictly necessary. Some missing *sample values* are allowable, and in fact, missing sample values is a common occurrence. Finally, the *land use, location setting, probe location, monitor objective,* and *sample frequency* are not currently used in BenMAP, and are included now to allow additional flexibility in future versions of BenMAP.

Exhibit A-3. Sample $PM_{2.5}$ File Format for User-Generated Monitor Text Files

Description	Sample Data[a]
List of variables	year, monitor ID, latitude, longitude, land use, method, location setting, probe location, monitor objective, sample frequency, sample values
Sample daily data with some missing sample values.	2002 , 010270001881011 , 33.281111, -85.802222, agricultural, 116, rural, side of building, highest concentration, 3, 15.2, . , . , . , 18.7, . , . , 12.3, . , . , 22.8, . , . , . , . , . , 10, etc.
Sample daily data with some missing sample values, as well as with missing *land use, location setting, probe location, monitor objective,* and *sample frequency* variables.	2002, 010270001881011, 33.281111, -85.802222, , 116, , , , , 3, 15.2, . , . , . , 18.7, . , . , 12.3, . , . , 22.8, . , . , . , . , . , . , 10, etc.

[a] Note that the data files need to be text files. However, the particular extension used for the files (e.g., *.txt, *.dat) is not important. Each line represents a separate monitor observation. In this particular example, there are 18 slots for $PM_{2.5}$ data, with six of the slots occupied by non-missing sample values. An actual file would have 365 slots, with a number of these slots occupied by missing values. Missing values should be delineated with ".".

Data for a given pollutant can include a variety of units. Ozone data may be in parts per million, parts per hundred million and parts per billion and micrograms per meter cubed. Similarly particulate matter can include a variety of measurements. When reading in data, it is converted to common units. Exhibit A-4 lists the measurement units for each of the pollutants and the measurement used in BenMAP, and Exhibit A-5 provides a description of the variables used in BenMAP.

Exhibit A-4. Common Measurement Units Used for Monitoring Data in BenMAP

Pollutant	Common Measurements (AQS Code)	Measurement in BenMAP	Notes
Ozone	1 = micrograms per cubic meter ($\mu g/m^3$) at 25°C 7 = parts per million (ppm) 8 = parts per billion (ppb) 40 = parts per hundred million (pphm)	ppb	1 ppb = 0.51 $\mu g/m^3$ at 25 Celsius and one atmosphere.
$PM_{2.5}$, PM_{10}	1 = $\mu g/m^3$ at 25°C 05 = milligrams per meter cubed (mg/m^3) at 25°C 105 = $\mu g/m^3$ at local conditions.	$\mu g/m^3$	Some data have a units code of 83 which is cubic meters per minute. Not clear what to do with data, so it is dropped from the analysis.

Source: U.S. EPA (U.S. EPA, 2002b). For more information on monitor data, please contact: Virginia Ambrose (ambrose.virginia@epa.gov).

Exhibit A-5. Description of Monitor Variables

Variable Name	Description	Variable Values [a]
method	The *method* identifies the approach used to collect the monitor data. For example, the Federal Reference Method for $PM_{2.5}$ includes *method* codes 116-120 and 123.	There are a large number of method code variables. For $PM_{2.5}$ the federal reference method codes are 116-120, and 123.
land use	Categorization of the prevalent land use within 1/4 mile of the Monitoring Site.	Agricultural; Blighted Areas; Commercial; Desert; Forest; Industrial; Military Reservation; Mobile; Residential
location setting	A description of the environmental setting within which the Site is located.	Rural; Suburban; Urban and Center City
probe location	Identification of the location of the sampling Probe.	Ground Level Support; Pole; Side of Building; Top of Building; Tower; Other
monitor objective	Identification of the reason for measuring air quality by the Monitor.	Extreme Downwind; General/background; Highest Concentration; Invalid Code Test; Maximum Ozone Concentration; Maximum Precursor Emissions Impact Other; Population Exposure; Regional Transport; Source Oriented; Upwind Background; Welfare Related Impacts
sample frequency	Indicates the scheduled elapsed time period between observations.	A = Daily: 24 - 1 Hour Samples B = Daily: 8 - 3 Hour Samples C = Daily: 1 - 3 Hour Samples D = Daily: 1 - 24 Hour Samples E = Daily: 4 - 6 Hour Samples F = Daily: 4 - 3 Hour Samples G = Every 3rd Day: 24 - 1 Hour Samples H = Every 3rd Day: 8 - 3 Hour Samples I = Every 3rd Day: 1 - 3 Hour Samples J = Every 3rd Day: 1 - 24 Hour Samples K = Every 3rd Day: 4 - 6 Hour Samples L = Every 3rd Day: 4 - 3 Hour Samples M = Every 6th Day: 24 - 1 Hour Samples N = Every 6th Day: 8 - 3 Hour Samples O = Every 6th Day: 1 - 3 Hour Samples P = Every 6th Day: 1- 24 Hour Samples Q = Every 6th Day: 4 - 3 Hour Samples S = Seasonal T = 5 out of 7 Days 1 = Every Day 2 = Every Other Day 3 = Every 3rd Day 4 = Every 4th Day 5 = Every 5th Day 6 = Every 6th Day 7 = Every 12th Day 8 = Stratified Random 9 = Random R = Episodic Sampling

[a] Source: U.S. EPA (U.S. EPA, 2002b). For more information on monitor data, please contact: Virginia Ambrose (ambrose.virginia@epa.gov).

A.2 How BenMAP Filters Monitor Data

For a given analysis, we typically choose a subset of the available monitors, using default options typically used in EPA analyses. The **Advanced** button provides access to these default monitor filter option options, and you can change them if you wish.

A.2.1 Filtering Particulate Matter Monitor Data

$PM_{2.5}$ and PM_{10} data are measured under both standard and local conditions. Standard data give readings based on conditions of 25 Celsius and 1 atmosphere of pressure; while local data give the reading based on the temperature and pressure at the monitoring site at the time of the measurement. The pollutant codes for standard and local data differ: local $PM_{2.5}$ = 88101, standard $PM_{2.5}$ = 81104, local PM_{10} = 85101, and standard PM_{10} = 81102.

The monitor selection process can be summarized in the following steps:

(1) Determine whether a monitor is valid. You can specify the minimum number of daily observations necessary for each quarter. Within this same calculation, you can specify:

- states;
- latitude and longitude; and
- POC codes. For PM_{10} the default is to exclude POC codes greater than 4. For $PM_{2.5}$ the default is to exclude POC codes greater than 2. For the POC codes kept in the universe of possible monitors, you also need to specify a preference for POC codes. For PM_{10} the default preference is that POC=1 > POC=2 > POC=3 > POC=4. For $PM_{2.5}$ the default preference is POC=1 > POC=2.

(2) Choose whether to include both local data and standard data. If you choose both types of data, then you need to specify a preference when both types of data are available at the same monitor 9-digit monitor id. (A 9-digit id includes 2-digit state, 3-digit county and 4-digit site identifiers).

Note: to avoid potential confusion, we may not allow you to choose standard $PM_{2.5}$ data. There are relatively few standard $PM_{2.5}$ monitors – on the order of ten or so each year – so the potential loss is small.

(3) Specify whether to have the data used in the model as standard or local data. Standard-to-local conversion factors provided to Don McCubbin from Bryan Hubbell on April 2, 2002 are multiplied with standard data to get local data; alternatively, to go from local to standard data, we divide by these conversion factors. These conversion factors are available at the 9-digit monitor id level. In cases where we do not have a conversion factor, we use a default value of 1.

(4) At the same 9-digit monitor id, there can be more than one monitor, identified by a 10-digit monitor id – the difference being the POC code in the tenth digit. The preference identified in (1) above is used to choose monitors.

A.2.2 Filtering Ozone Monitor Data

The monitor selection process can be summarized in the following steps:

(1) Determine whether a day is valid. You can specify a window of time in each day that will be used to determine if there are sufficient observations for the given day. Within your specified window (e.g., 8:00am to 7:59pm), you can specify the minimum number of non-missing hours (e.g., 9 out of 12 possible). Within this same calculation, you can specify:

- states;
- latitude and longitude; and
- POC codes. The default is to exclude POC codes greater than 4. For the POC codes kept in the universe of possible monitors, you also need to specify a preference for POC codes. The default is that POC=1 > POC=2 > POC=3 > POC=4.

(2) Determine whether a monitor is valid. You can specify a window of time (e.g., May through September) to check to see if there are a sufficient number of valid days. You can to specify the minimum number of observations necessary during this user-specified window.

(3) At the same 9-digit monitor id, there can be more than one monitor, identified by a 10-digit monitor id – the difference being the POC code in the tenth digit. The preference identified in (1) above will be used to choose monitors.

A.3 Monitor Rollbacks

Once a set of monitors has been selected, the user may define one or more non-overlapping rollback **regions**. A region is simply a set of states with an associated set of rollback parameter values. Three rollback **types** are available - *Percentage Rollback*, *Incremental Rollback*, and *Rollback to a Standard*. Each of these rollback types has different rollback parameters associated with it.

A.3.1 Percentage Rollback

Percentage Rollback involves setting only two parameters - a **percentage** and a **background level**. The rollback procedure is similarly straightforward - each observation at each monitor in the region has the portion of its value which is above **background level** reduced by **percentage**.

Example: Background Level: 35; Percentage: 25

Initial Observations at a monitor in rollback region:

20 20 25 59 35 51 83 35 30 67 87 79 63 35 35

If we select the background level of 35, we first calculate the portion of each observation that is above background level, that is, we subtract the **background level** from the initial observation level. Observations below background level are given a value of 0.

Observation portions above **background level**:

0 0 0 24 0 16 48 0 0 32 52 44 28 0 0

When we apply the rollback **percentage**, each observation portion gets reduced by 25%.

Reduced portions above background level:

0 0 0 18 0 12 36 0 0 24 39 33 21 0 0

Then, each reduced portion is added to the background level of 35. Zero values are replaced by the initial observations.

Reduced Observations:

20 20 25 53 35 47 71 35 30 59 74 68 56 35 35

A.3.2 Incremental Rollback

Incremental Rollback similarly involves setting only two parameters - an **increment** and a **background level**. The rollback procedure is quite similar to the percentage rollback procedure - each observation at each monitor in the region has the portion of its value which is above **background level** reduced by **increment**. The reduced values are not allowed to become negative, however - that is, they are truncated at zero.

Example: Background Level: 35; Increment: 25

Initial Observations:

20 20 25 59 35 51 83 35 30 67 87 79 63 35 35

Observation portions above background level:

0 0 0 24 0 16 48 0 0 32 52 44 28 0 0

Reduced portions above background level:

0 0 0 0 0 0 23 0 0 7 27 19 3 0 0

Reduced Observations:

20 20 25 35 35 35 58 35 30 42 62 54 38 35 35

A.3.3 Rollback to a Standard

Rollback to a Standard has two groups of parameters - those associated with the **Attainment Test**, which determines whether a monitor is in attainment (meets the standard), and those associated with the **Rollback Methods**, which are used to bring out of attainment monitors into attainment.

The **Attainment Test** parameters are **Metric**, **Ordinality**, and **Standard**. A monitor is considered in attainment if the n^{th} highest value of the metric specified by **Metric** is at or below the value specified by **Standard**, where *n* is the value specified by **Ordinality**. For example, if **Metric** is TwentyFourHourDailyAverage, **Ordinality** is two, and **Standard** is eighty five, a monitor will be considered in attainment if the second highest value of TwentyFourHourDailyAverage is at or below eighty five.

Supported metrics for pollutants with hourly observations (Ozone) include FiveHourDailyAverage, EightHourDailyAverage, TwelveHourDailyAverage, TwentyFourHourDailyAverage, OneHourDailyMax, and EightHourDailyMax. Supported metrics for pollutants with daily observations (PM10, PM2.5) include TwentyFourHourDailyAverage and AnnualAverage. For Annual Average, **Ordinality** does not apply, since there is only a single metric value to work with.

The **Rollback Method** parameters are **Interday Rollback Method**, **Interday Background Level**, **Intraday Rollback Method**, and **Interday Background Level**. These four parameters determine the rollback procedures used to bring out of attainment monitors into attainment. The **Interday Rollback Method** and **Background Level** are used to generate target values for the metric specified by the **Attainment Test**. The **Intraday Rollback Method** and **Background Level** are used to adjust hourly observations to meet the target metric values generated in the previous step. As such, the **Intraday Rollback Method** and **Background Level** are used only for pollutants with hourly observations (ozone).

Interday Rollback - Generating Target Metric Values

Because standards are defined on metrics, not directly on observations, the first step in rolling back out of attainment monitors is generating target metric values. There are four supported rollback methods for **Interday Rollbacks** - *Percentage*, *Incremental*, *Peak Shaving*, and *Quadratic*. Each of these rollback methods requires some preprocessing of the initial monitor metric values. We will discuss this preprocessing first, and then go through *Percentage*, *Incremental*, and *Peak Shaving* rollbacks in turn. *Quadratic* rollback is more complicated than these first three, and has its own section.

The **Interday Background Level** specifies the portion of each metric value which cannot be affected by human intervention - we call this portion the *non-anthropogenic* portion. Whatever portion is left over after subtracting out the background level is referred to as the *anthropogenic* portion. The *anthropogenic* portion of the initial monitor metric values is the only part which will be affected by the **Interday Rollback Method**.

BenMAP calculates an *out of attainment value* by determining the particular monitor metric value which caused the monitor to be out of attainment - this value is the n^{th} highest value of the metric specified by the **Attainment Test** metric, where *n* is the **Attainment Test** ordinality. BenMAP then calculates an *anthropogenic out of attainment value* by subtracting the **Interday Background Level** from the *out of attainment value*. BenMAP also calculates an *anthropogenic standard* by subtracting the **Interday Background Level** from the **Attainment Test** standard. Finally, BenMAP calculates a set of

anthropogenic metric values and a set of *non-anthropogenic metric values* using the following procedure on each initial monitor metric value:

IF the metric value is less than or equal to the Interday Background Level,
non-anthropogenic metric value = metric value
anthropogenic metric value = 0
ELSE
non-anthropogenic metric value = Interday Background Level
anthropogenic metric value = metric value - Interday Background Level

Interday Rollback - Percentage

To generate target metric values using Percentage rollback, BenMAP calculates the percentage required to reduce the *anthropogenic out of attainment value* to exactly the *anthropogenic standard*. This percentage reduction is then applied to all of the *anthropogenic metric values*. Finally, these *reduced anthropogenic metric values* are added to the *non-anthropogenic metric values* to give the final *target metric values*.

Example:

Initial Metric Values:

30 35 50 100 80 44 67 88 90 70 50 30 55 90 80 85

Attainment Test: Highest value of metric <= 70
Interday Background Level: 40
Out of Attainment Value: 100
Anthropogenic Out of Attainment Value: 60 (= 100 - 40)
Anthropogenic Standard: 30 (= 70 - 40)
Percentage Reduction Required: 50% (=(60-30)/60)

Non-Anthropogenic Metric Values:

30 35 40 40 40 40 40 40 40 40 40 30 40 40 40 40

Anthropogenic Metric Values:

0 0 10 60 40 4 27 48 50 30 10 0 15 50 40 45

Reduced Anthropogenic Metric Values:

0 0 5 30 20 2 14 24 25 15 5 0 8 25 20 23

Target Metric Values:

30 35 45 70 60 42 54 64 65 55 45 30 48 65 60 63

Interday Rollback - Incremental

To generate target metric values using **Incremental Rollback**, BenMAP calculates the increment required to reduce the *anthropogenic out of attainment value* to exactly the *anthropogenic standard*. This incremental reduction is then applied to all of the *anthropogenic metric values* (but - they are not allowed to fall below zero). Finally, these *reduced anthropogenic metric values* are added to the *non-anthropogenic metric values* to give the final *target metric values*.

Example:

Initial Metric Values:

30 35 50 100 80 44 67 88 90 70 50 30 55 90 80 85

Attainment Test: Highest value of metric <= 70
Interday Background Level: 40
Interday Rollback Method: Incremental
Out of Attainment Value: 100
Anthropogenic Out of Attainment Value: 60
Anthropogenic Standard: 30 (=70 - 30)
Incremental Reduction Required: 30

Non-Anthropogenic Metric Values:

30 35 40 40 40 40 40 40 40 40 40 30 40 40 40 40

Anthropogenic Metric Values:

0 0 10 60 40 4 27 48 50 30 10 0 15 50 40 45

Reduced Anthropogenic Metric Values:

0 0 5 30 20 2 14 24 25 15 5 0 8 25 20 23

Target Metric Values:

30 35 45 70 60 42 54 64 65 55 45 30 48 65 60 63

Interday Rollback - Peak Shaving

To generate target metric values using Peak Shaving rollback, BenMAP simply truncates all *anthropogenic metric values* at the *anthropogenic standard*. These *reduced anthropogenic metric values* are added to the *non-anthropogenic metric values* to give the final *target metric values*.

Example:

Initial Metric Values:

30 35 50 100 80 44 67 88 90 70 50 30 55 90 80 85

Attainment Test: Highest value of metric <= 70
Interday Background Level: 40
Interday Rollback Method: Peak Shaving
Anthropogenic Standard: 30

Non-Anthropogenic Metric Values:

30 35 40 40 40 40 40 40 40 40 40 30 40 40 40 40

Anthropogenic Metric Values:

0 0 10 60 40 4 27 48 50 30 10 0 15 50 40 45

Reduced Anthropogenic Metric Values:

0 0 10 30 30 4 27 30 30 30 10 0 15 30 30 30

Target Metric Values:

30 35 50 70 70 44 67 70 70 70 50 30 55 70 70 70

Intraday Rollback - Adjusting Hourly Observations

Once a set of target metric values has been calculated for a pollutant with hourly observations (e.g., Ozone), BenMAP must adjust the hourly observations so that they produce the target metric values. There are three supported rollback methods for **Intraday Rollback** - *Percentage*, *Incremental*, and *Quadratic*. Each of these rollback methods requires some preprocessing of the initial monitor observations, and each can require multiple iterations to hit the target metric values. We will discuss this preprocessing and iteration first, and then go through *Percentage* and *Incremental* rollbacks in turn. *Quadratic* rollback is more complicated than these first two, and has its own section.

For various reasons, each of the **Intraday Rollback** methods can fail to hit the target metric values during a single pass through the rollback procedure (these will be discussed in detail below). As such, each of the rollback methods uses an iterative approach to get within a threshold of each of the target metric values - currently this threshold is 0.05. The iterative approach works as follows:

For each target metric value, BenMAP calculates the current value of the **Attainment Test** metric. For the first iteration, the metric value will be calculated using unadjusted hourly observations. For subsequent iterations, the metric value will be calculated using the current values of the adjusted hourly observations.

If the difference between the metric value and the target metric value is less than or equal to 0.05, the rollback procedure is finished. Otherwise, another iteration is required.

The **Intraday Background Level** specifies the portion of each observation which cannot be affected by human intervention - we call this portion the *non-anthropogenic* portion. Whatever portion is left over after subtracting out the background level is referred to as the *anthropogenic* portion. The *anthropogenic* portion of the initial monitor observations is the only part which will be affected by the **Intraday Rollback Method.**

In a way analogous to the **Interday Rollback** procedure, BenMAP calculates the twenty-four hourly *anthropogenic observations* and the twenty-four hourly *non-anthropogenic observations* using the following procedure for each hourly observation:

IF the current value of the observation is less than or equal to the Intraday Background Level,
    non-anthropogenic observation = observation
    anthropogenic observation = 0
ELSE
    non-anthropogenic observation = Intraday Background Level
    anthropogenic observation = observation - Intraday Background Level

Given (i) an **Attainment Test** Metric (e.g., EightHourDailyMax), (ii) an **Intraday Background Level**, and (iii) a target metric value for the day, BenMAP proceeds to adjust hourly observations in the following steps:

1. Calculate the **Attainment Test** metric (e.g., the 8-hour daily maximum);
2. Identify the "window" – i.e., the set of hours used to calculate the metric (e.g., if the 8-hour daily maximum is achieved in the first 8 hours, then the window is comprised of the first 8 hours);
3. Calculate the *non-anthropogenic hourly observations* (=min(hourly observation, Intraday Background Level));
4. Calculate the *anthropogenic hourly observations* (=hourly observation - Intraday Background Level);
5. Calculate the *non-anthropogenic metric value* (= the metric using the *non-anthropogenic hourly observations* in the "window");
6. Calculate the *anthropogenic metric value* (= the metric using the *anthropogenic hourly observations* in the "window");
7. Calculate the *anthropogenic target metric value* (= the target metric value minus the non-anthropogenic metric value);
8. Calculate the reduction required to get the *anthropogenic metric value* down to the *anthropogenic target metric value*;
9. Adjust all *anthropogenic hourly observations* by the reduction calculated on the previous step;
10. Calculate the adjusted hourly observations (= the adjusted *anthropogenic hourly observation* + the *non-anthropogenic hourly observation*).

Intraday Rollback - Percentage

Below, we present two examples of a percentage-based **Intraday Rollback**. In one example, a single iteration is needed, and in the second example, two iterations are required because a number of the monitor values fall below the assumed background level.

Example: All Hourly Observations Exceed the Intraday Background (Single Iteration)

If all of the hourly observations in a day are greater than the Intraday Background Level, then the above procedure is straightforward and can be accomplished in a single iteration. We illustrate with the following example. Suppose that:

Metric = EightHourDailyMax,
Target metric value for a given day = 85
Intraday Background Level = 40.

And that the hourly observations on that day are:

530 45 50 60 45 45 45 60 70 100 100 100 100 100 100 100 100 60 45 50 45 45 47 47

Based on these observations, we see that the 8-hour daily maximum = 110.

Assuming a background level of 40, then the *Anthropogenic* hourly observations are:

490 5 10 20 5 5 5 20 30 60 60 60 60 60 60 60 60 20 5 10 5 5 7 7

Then, we know:

Anthropogenic metric value = 70.
Non-anthropogenic metric value = 40.
Anthropogenic target metric value = 45.
Percentage reduction required = ((70-45)/70) = 35.7%

All of the hourly anthropogenic observations are reduced by 35.7%. The average of the first 8 values (the window on which the Test metric is based) will be exactly 45, the anthropogenic target metric value. Finally, the adjusted hourly observations are calculated by adding the *non-anthropogenic hourly observation* to the adjusted hourly anthropogenic observations.

Example: Some Hourly Observations are Below the Intraday Background (Multiple Iterations Required)

In the above example, the anthropogenic target metric value was met on a single iteration because all of the hourly observations were greater than the **Intraday Background Level**. In this case, a simple percent reduction of all hourly values will produce an average in the window that is equal to the anthropogenic target metric value. If some of the hourly observations in a day are less than or equal to the **Intraday Background Level**, however, then BenMAP uses an iterative procedure. On each iteration, it adjusts hourly observations using the 10-step method given above. It then compares the new metric value to the target metric value. If the difference is less than or equal to 0.05 ppb, the rollback procedure is finished. Otherwise, another iteration is required. The iterative procedure is illustrated in the following example.

Note that we are presenting an example below with an intraday background of 40 ppb. We only use a non-zero intraday background as a sensitivity analysis in Exhibit 4-5, where we use intraday backgrounds of 10, 20, 30, and 40. For the rest of our results we use an intraday background of 0 ppb.

Suppose that:

Metric = EightHourDailyMax,
Target metric value for a given day = 85
Intraday Background Level = 40.

Suppose also that the hourly observations on that day are:

530 20 25 60 35 35 40 60 70 100 100 100 100 100 100 100 100 60 33 40 30 30 25 20

Then, we know that the 8-hour daily maximum = 100.6.

Non-Anthropogenic Hourly Observations, Iteration One:

40 20 25 40 35 35 40 40 40 40 40 40 40 40 40 40 40 40 33 40 30 30 25 20

Anthropogenic Hourly Observations, Iteration One:

490 0 0 20 0 0 0 20 30 60 60 60 60 60 60 60 60 20 0 0 0 0 0 0

Non-Anthropogenic Metric Value: 34.4 (EightHourDailyMax - calculated over the same eight hour window as the initial metric value was calculated over)
Anthropogenic Metric Value: 66.3
Anthropogenic Target Metric Value: 50.6
Percentage Reduction Required: 23.6%

Reduced Anthropogenic Hourly Observations, Iteration One:

374 0 0 15 0 0 0 15 23 46 46 46 46 46 46 46 46 15 0 0 0 0 0 0

Reduced Hourly Observations, Iteration One:

414 20 25 55 35 35 40 55 63 86 86 86 86 86 86 86 86 55 33 40 30 30 25 20

Reduced Metric Value (EightHourDailyMax): 85.8
Target Metric Value (EightHourDailyMax): 85

Non-Anthropogenic Hourly Observations, Iteration Two:

40 20 25 40 35 35 40 40 40 40 40 40 40 40 40 40 40 40 33 40 30 30 25 20

Anthropogenic Hourly Observations, Iteration Two:

374 0 0 15 0 0 0 15 23 46 46 46 46 46 46 46 46 15 0 0 0 0 0 0

Non-Anthropogenic Metric Value: 40 (EightHourDailyMax - calculated over the same eight hour window the initial metric value was calculated over)
Anthropogenic Metric Value: 45.8
Anthropogenic Target Metric Value: 45
Percentage Reduction Required: 1.9%

Reduced Anthropogenic Hourly Observations, Iteration Two:

368 0 0 15 0 0 0 15 23 45 45 45 45 45 45 45 45 15 0 0 0 0 0 0

Reduced Hourly Observations, Iteration Two:

408 20 25 55 35 35 40 55 63 85 85 85 85 85 85 85 85 55 33 40 30 30 25 20

Reduced Metric Value (EightHourDailyMax): 85

The above example, in addition to illustrating the **Intraday Percentage Rollback**, also illustrates one reason why the iterative procedure can be necessary. When using the EightHourDailyMax metric in the **Attainment Test**, it is possible for the window over which the maximum eight hour average occurs to move after a single pass through the rollback procedure. When this happens, it becomes necessary to go through additional iterations to hit the target metric value.

Intraday Rollback - Incremental

To adjust hourly observations using Incremental rollback, BenMAP calculates the increment required to reduce the *anthropogenic metric value* to exactly the *anthropogenic target metric value*. This incremental reduction is then applied to all of the *anthropogenic observations* (but - they are not allowed to fall below zero). Finally, these *reduced anthropogenic observations* are added to the *non-anthropogenic observations* to give the final *reduced observations*.

Example:

Initial Hourly Observations:

20 20 25 60 35 35 40 70 35 30 65 90 76 65 35 35 54 60 33 40 30 30 25 20

Initial Metric Value (EightHourDailyMax): 60
Target Metric Value (EightHourDailyMax): 55
Intraday Background Level: 40
Intraday Rollback Method: Incremental

Non-Anthropogenic Hourly Observations, Iteration One:

20 20 25 40 35 35 40 40 35 30 40 40 40 40 35 35 40 40 33 40 30 30 25 20

Anthropogenic Hourly Observations, Iteration One:

0 0 0 20 0 0 0 30 0 0 25 50 36 25 0 0 14 20 0 0 0 0 0 0

Non-Anthropogenic Metric Value (EightHourDailyMax): 38.8
Anthropogenic Metric Value (EightHourDailyMax): 21.3
Anthropogenic Target Metric Value (EightHourDailyMax): 16.3
Incremental Reduction Required: 5.0

Reduced Anthropogenic Hourly Observations, Iteration One:

0 0 0 15 0 0 0 25 0 0 20 45 31 20 0 0 9 15 0 0 0 0 0 0

Reduced Hourly Observations, Iteration One:

20 20 25 55 35 35 40 65 35 30 60 85 71 60 35 35 49 55 33 40 30 30 25 20

Reduced Metric Value (EightHourDailyMax): 56.25
Target Metric Value (EightHourDailyMax): 55

Non-Anthropogenic Hourly Observations, Iteration Two:

20 20 25 40 35 35 40 40 35 30 40 40 40 40 35 35 40 40 33 40 30 30 25 20

Anthropogenic Hourly Observations, Iteration Two:

0 0 0 15 0 0 0 25 0 0 20 45 31 20 0 0 9 15 0 0 0 0 0 0

Non-Anthropogenic Metric Value (EightHourDailyMax): 38.8
Anthropogenic Metric Value (EightHourDailyMax): 17.5
Anthropogenic Target Metric Value (EightHourDailyMax): 16.3
Incremental Reduction Required: 1.25

Reduced Anthropogenic Hourly Observations, Iteration Two:

0 0 0 14 0 0 0 24 0 0 19 44 30 19 0 0 8 14 0 0 0 0 0 0

Reduced Hourly Observations, Iteration Two:

20 20 25 54 35 35 40 64 35 30 59 84 70 59 35 35 48 54 33 40 30 30 25 20

Reduced Metric Value (EightHourDailyMax): 55.3
Target Metric Value (EightHourDailyMax): 55

This example should actually continue for one further iteration, with a new Incremental Reduction of 0.3. This illustrates another reason why the iterative procedure can be necessary - for incremental reductions, the prohibition against values becoming negative can cause target metric values to not be met. Incremental reductions thus very often require multiple iterations.

Interday and Intraday Rollback - Quadratic

Quadratic rollback is based on an algorithm developed by Horst and Duff (1995). The idea behind quadratic rollback is to reduce large values proportionally more than small values while just achieving the standard - that is, the out-of-attainment value should be more or less at the standard after the rollback (some small amount of error is involved).

The original quadratic rollback algorithm is designed to roll back hourly observations given a desired peak value. That is, it assumes that the Attainment Test metric is the one-hour average and the Attainment Test ordinality is one. As such, the algorithm was modified slightly to allow for ordinalities other than one to be used.

The basic formula for quadratic rollback is:

$$\textit{Reduced Observation} = [\ 1 - (\ A + B * \textit{Initial Observation}\ )\ ] * \textit{Initial Observation}$$

where:
i ranges over the days being reduced.
$A = 1 - V$
$V = \text{Min}(\ 1, V_i\ )$
$V_i = (\ 2 * \textit{Maximum Observation Value} * \textit{Standard}\ ) / X_i$
$X_i = (\ 2 * \textit{Maximum Observation Value} * \textit{Metrics}_i\ ) - \text{Metrics}_i^2$
$B = \text{Max}(\ 0, [(\ V * \textit{Out of Attainment Value} - \textit{Standard}\ ) / \textit{Out of Attainment Value}^2]\ )$

Quadratic Rollback - Interday

Because Quadratic Rollback was originally designed to adjust hourly observations to meet a daily metric standard, it is slightly complicated to use it to generate target metric values.

First, Quadratic Rollback calculates the *anthropogenic out of attainment value* by subtracting the *Intraday* Background Level from the out of attainment value. Note that this differs from the other interday rollback methods, which subtract the *Interday* Background Level from the out of attainment value. Similarly, the *anthropogenic standard* is calculated by subtracting the Intraday Background Level from the standard.

The *anthropogenic observations* and *non-anthropogenic observations* are then calculated. For pollutants which have daily observations (PM10, PM2.5) the *anthropogenic metric values* are used (see above for their calculation). For pollutants which have hourly observations (Ozone), Quadratic Rollback loops through each metric value and calculates the twenty four corresponding *anthropogenic observations* and *non-anthropogenic observations* as follows:

IF the metric value is at or below the Interday Background Level,
    For each observation,
        non-anthropogenic observation = observation
        anthropogenic observation = 0
ELSE
    For each observation,
        IF the observation is at or below the Intraday Background Level
            non-anthropogenic observation = observation
            anthropogenic observation = 0
        ELSE
            non-anthropogenic observation = Intraday Background Level
            anthropogenic observation = observation - Intraday Background Level

A new set of *anthropogenic metric values* is then calculated by generating the Attainment Test metric from the *anthropogenic observations*. The Quadratic Rollback algorithm is then called, passing in the *anthropogenic metric values* as *Metrics*, *anthropogenic observations* as *Observations*, *anthropogenic standard* as *Standard*, and *anthropogenic out of attainment value* as *Out of Attainment Value*. The result is a set of *reduced anthropogenic observations*. These are then added together with the *non-anthropogenic observations* to give a final set of *reduced observations*.

Then, if Quadratic Rollback was also selected as the Intraday Rollback method, these observations are used as the final reduced observations for the monitor. Otherwise, metric targets are generated from these hourly observations, and the observations themselves are discarded.

Quadratic Rollback - Intraday

Quadratic Rollback can also be used to adjust hourly observations to meet metric targets generated via a different rollback method. In this case, the algorithm is used to adjust each set of twenty four hourly observations to meet the corresponding metric target. Intraday Quadratic Rollback uses the normal set of *anthropogenic observations* as *Observations*, a single normal *anthropogenic metric value* as *Metrics*, and the normal *anthropogenic metric target* as *Standard*. Intraday Quadratic Rollback tends to always slightly miss its metric target, so it is not run in an iterative fashion as the other Intraday Rollback Methods are (doing so would sometimes result in an infinite loop).

A.4 SAS Code Used to Prepare Monitor Data for BenMAP

In this sub-section, we present the SAS code that we used to process the data for BenMAP. In preparing your own monitor data for BenMAP, you may want to use some of the algorithms presented here.

To ease the need of differentiating between leap and non-leap years, BenMAP assumes a year with 365 days. When incorporating monitoring data from a leap year (e.g., 2000), we average the values for February 28 and 29.

A.4.1 SAS Code for Processing Raw Particulate Matter Monitor Data for Input to BenMAP

```
title "file: amp500_501_pmdata_v3.sas";

/*****************************************************************************
02/12/03.
Etienne Gabel. Abt Associates.

This SAS code reads in from raw AMP500 file which contains monitor description data,
and from raw AMP501 file which contains formatted hourly pm2.5 and pm10 monitor sampling
data.
The code prepares a data set which contains the daily average data within a single array of
365 days. Each observation (row) corresponds to a unique monitor (including
poc code) and method pair.
For 1996 (leap year), February 28th values are set as the average of February 28th and 29th
sample values, and February 29th data are then dropped since CAPMS is hardwired to accept 365
days of pm data.

7/15/03 - Ed Al-Hussainy -      changed code to read-in AMP500 data for 1997-2001 and merge it with the older 1996 data added
miami-dade county correction.

7/16/03 - Ed Al-Hussainy -     changed macro to accomodate file format of the 2002 PM 2.5 (88101) source data file
(PM25_Continuous_2002.txt).  Note that samplefreq is not available in 2002 PM 2.5 (88101) file. setting it = to 'blank' for
compatibility with rest of code.


*****************************************************************************/

options missing = '.';

filename yr366 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\year_366_days.csv";

filename infile1 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG81104-1996.TXT";
filename infile2 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG_501_81104_NATION_1997.TXT";
filename infile3 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG_501_81104_NATION_1998.TXT";
filename infile4 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG81104-1999.TXT";
filename infile5 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG81104-2000.TXT";
filename infile6 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG81104-2001.TXT";
filename infile7 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG_501_81104_NATION_2002.TXT";
filename infile8 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG88101-1996.TXT";
filename infile9
"\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\RD_501_88101_PM25_NATION_1997.TXT";
filename infile10
"\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\RD_501_88101_PM25_NATION_1998.TXT";
filename infile11 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG88101-1999.TXT";
filename infile12 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG88101-2000.TXT";
filename infile13 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\EG88101-2001.TXT";
filename infile14 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\PM25_Continuous_2002.txt";
filename infile15 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG81102-1996.TXT";
filename infile16
"\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG_501_81102_PM10_NATION_1997.TXT";
filename infile17
"\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG_501_81102_PM10_NATION_1998.TXT";
filename infile18 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG81102-1999.TXT";
filename infile19 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG81102-2000.TXT";
filename infile20 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG81102-2001.TXT";
filename infile21
"\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG_501_81102_PM10_NATION_2002.TXT";
filename infile22 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG85101-1996.TXT";
filename infile23
"\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG_501_85101_NATION_1997.TXT";
```

```
filename infile24
"\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG_501_85101_NATION_1998.TXT";
filename infile25 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG85101-1999.TXT";
filename infile26 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG85101-2000.TXT";
filename infile27 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG85101-2001.TXT";
filename infile28
"\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM10\EG_501_85101_NATION_2002.TXT";
* additional 2002 pm2.5 monitor data file for methods <701;
filename infile29 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\data\PM2.5\pm25_daily2002.txt";

filename outf1 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std96.TXT";
filename outf2 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std97.TXT";
filename outf3 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std98.TXT";
filename outf4 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std99.TXT";
filename outf5 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std00.TXT";
filename outf6 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std01.TXT";
filename outf7 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std02.TXT";
filename outf8 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl96.TXT";
filename outf9 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl97.TXT";
filename outf10 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl98.TXT";
filename outf11 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl99.TXT";
filename outf12 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl00.TXT";
filename outf13 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl01.TXT";
filename outf14 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl02.TXT";
filename outf15 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std96.TXT";
filename outf16 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std97.TXT";
filename outf17 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std98.TXT";
filename outf18 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std99.TXT";
filename outf19 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std00.TXT";
filename outf20 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std01.TXT";
filename outf21 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std02.TXT";
filename outf22 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl96.TXT";
filename outf23 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl97.TXT";
filename outf24 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl98.TXT";
filename outf25 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl99.TXT";
filename outf26 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl00.TXT";
filename outf27 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl01.TXT";
filename outf28 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl02.TXT";

filename outjk1 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std96junk.TXT";
filename outjk2 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std97junk.TXT";
filename outjk3 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std98junk.TXT";
filename outjk4 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std99junk.TXT";
filename outjk5 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std00junk.TXT";
filename outjk6 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std01junk.TXT";
filename outjk7 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25std02junk.TXT";
filename outjk8 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl96junk.TXT";
filename outjk9 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl97junk.TXT";
filename outjk10 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl98junk.TXT";
filename outjk11 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl99junk.TXT";
filename outjk12 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl00junk.TXT";
filename outjk13 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl01junk.TXT";
filename outjk14 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM25lcl02junk.TXT";
filename outjk15 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std96junk.TXT";
filename outjk16 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std97junk.TXT";
filename outjk17 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std98junk.TXT";
filename outjk18 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std99junk.TXT";
filename outjk19 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std00junk.TXT";
filename outjk20 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std01junk.TXT";
filename outjk21 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10std02junk.TXT";
filename outjk22 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl96junk.TXT";
filename outjk23 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl97junk.TXT";
```

```
filename outjk24 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl98junk.TXT";
filename outjk25 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl99junk.TXT";
filename outjk26 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl00junk.TXT";
filename outjk27 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl01junk.TXT";
filename outjk28 "\\BE5832\C$\capms_revision\data_bases\air_monitor\sas_work\output\PM10lcl02junk.TXT";

/*************************************************************************************
READ IN AMP500 FILE. CREATE MONITOR ID VARIABLE WITH STATE, COUNTY, AND SITE ID.
TRANSACTION CODE 'AA' INCLUDES MONID, LAND USE, LOCATION SETTING, LATITUDE AND LONGITUDE
VARIABLES.  TRANSACTION CODE 'MA' INCLUDES MONID, POC AND PROBE LOCATION VARIABLES.
TRANSACTION CODE 'ME' INCLUDES MONID, POC AND MONITOR OBJECTIVE VARIABLES.
NOTE: PROBE LOCATION AND MONITOR OBJECTIVE ARE EQUIVALENT FOR PM25 STANDARD AND LOCAL, AND
ARE EQUIVALENT FOR PM10 STANDARD AND LOCAL.
*************************************************************************************/

* start of Ed's edit;

/* AMP 500 data */
%let netdata=H:\ENVR\MCCUBBIN\capms_revision\data_bases\air_monitor\sas_work\data\AMP500 file -- contains monitor
description info;

filename amp500_1 "&netdata\EG_500_SELPARM_1997_2002.TXT";
filename amp500_2 "&netdata\AMP500.TXT";

* read-in new and old AMP500 data;

%macro read_amp500(param);

data tempAA_&param;
  infile amp500_&param firstobs=1 delimiter='|' missover lrecl = 6000 dsd;
  length transactioncode $ 2 state $ 2 county $ 3 siteid $ 4 poc $ 1 x1-x14 $ 5 landuse $ 20 locsetting $ 50
       probelocation $ 20 monobjective $ 50 monid $ 9;
  input transactioncode $ @;

  if transactioncode="AA" then do;
    input actioncode $ state $ county $ siteid $ latitude longitude x1-x14 $ landuse $
      locsetting $;
  end;

            if transactioncode="MA" then do;
    input actioncode $ state $ county $ siteid $ parameter $ poc $ x1-x4 $ probelocation $;
              if probelocation="" then delete;
            end;

            if transactioncode="ME" then do;
    input actioncode $ state $ county $ siteid $ parameter $ poc $ monobjective $;
              if monobjective="" then delete;
  end;

  drop x1-x14 actioncode;

  * Miami-Dade county correction;
  if state="12" and county="025" then county="086";

  monid = state || county || siteid;

  if transactioncode in ("AA" "MA" "ME");
run;

proc sort data=tempAA_&param;
            by monid;
run;
```

```
%mend read_amp500;

%read_amp500(1);
%read_amp500(2);

* merge new and old AMP500 data, separate data based on transactioncode and parameter;
data tempAA(keep=monid state county siteid latitude longitude landuse locsetting)
             tempMA25(keep=monid state county siteid parameter poc probelocation)
    tempMA10(keep=monid state county siteid parameter poc probelocation)
    tempME25(keep=monid state county siteid parameter poc monobjective)
    tempME10(keep=monid state county siteid parameter poc monobjective);

            * note the merge order - tempAA_1 overwrites tempAA_2 in case of duplicate entries;
            merge  tempAA_2(in=dum2) tempAA_1(in=dum1);
            by monid;

            if transactioncode="AA" then output tempAA;

            else if transactioncode="MA" then do;
                        if parameter="88101" or parameter="81104" then output tempMA25;
                   if parameter="85101" or parameter="81102" then output tempMA10;
            end;

            else if transactioncode="ME" then do;
            if parameter="88101" or parameter="81104" then output tempME25;
            if parameter="85101" or parameter="81102" then output tempME10;
            end;
run;

* end of Ed's edit;

proc sort data=tempMA25;
  by monid poc;
proc sort data=tempME25;
  by monid poc;
proc sort data=tempMA10;
  by monid poc;
proc sort data=tempME10;
  by monid poc;

/** ENTER FLAG "ZZ" FOR MONITORS IN AMP500 FILE WITH MORE THAN ONE LAND USE CODE. **/
proc sort data=tempAA out=tempAA1 nodupkey;
  by monid landuse;
proc summary data=tempAA1;
  by monid;
  output out=tempAA2;
data tempAA3;
  merge tempAA2 tempAA;
  by monid;
  if _freq_ > 1 then landuse="ZZ";
  keep monid latitude longitude landuse locsetting;
proc sort data=tempAA3 out=tempAA4 nodupkey;
  by monid;
proc print data=tempAA4(obs=50);
  where landuse="ZZ";
  title2 "MORE THAN ONE LAND USE FOR A MONITOR IN AMP500 FILE, tempAA4";

/** ENTER FLAG "ZZ" FOR MONITORS IN AMP500 FILE WITH MORE THAN ONE LOCATION SETTING CODE. **/
proc sort data=tempAA out=tempAA5 nodupkey;
  by monid locsetting;
proc summary data=tempAA5;
  by monid;
```

```
  output out=tempAA6;
data tempAA7;
  merge tempAA6(in=check1) tempAA;
  by monid;
  if check1=1 and _freq_ > 1 then locsetting="ZZ";
  keep monid latitude longitude landuse locsetting;
proc sort data=tempAA7 out=tempAA8 nodupkey;
  by monid;
proc print data=tempAA8(obs=50);
  where locsetting="ZZ";
  title2 "MORE THAN ONE LOCATION SETTING FOR A MONITOR IN AMP500 FILE, tempAA8";

/** MAKE SINGLE FILE WITH TRANSACTION AA DATA, WITH FLAG FOR REPEATING LAND USE OR LOCATION
SETTING CODES. **/
data tempAA9;
  merge tempAA4(drop=locsetting) tempAA8(drop=landuse);
  by monid;
  keep monid latitude longitude landuse locsetting;

/** ENTER FLAG "ZZ" FOR PM2.5 MONITORS IN AMP500 FILE WITH MORE THAN ONE PROBE LOCATION CODE. **/
proc sort data=tempMA25 out=tempMA25a nodupkey;
  by monid poc probelocation;
proc summary data=tempMA25a;
  by monid poc;
  output out=tempMA25b;
data tempMA25c;
  merge tempMA25b tempMA25;
  by monid poc;
  if _freq_ > 1 then probelocation="ZZ";
  keep monid poc probelocation;
proc sort data=tempMA25c out=tempMA25d nodupkey;
  by monid poc;
proc print data=tempMA25d(obs=50);
  where probelocation="ZZ";
  title2 "MORE THAN ONE PROBE LOCATION FOR A PM2.5 MONITOR IN AMP500 FILE, tempMA25d";

/** ENTER FLAG "ZZ" FOR PM10 MONITORS IN AMP500 FILE WITH MORE THAN ONE PROBE LOCATION CODE. **/
proc sort data=tempMA10 out=tempMA10a nodupkey;
  by monid poc probelocation;
proc summary data=tempMA10a;
  by monid poc;
  output out=tempMA10b;
data tempMA10c;
  merge tempMA10b tempMA10;
  by monid poc;
  if _freq_ > 1 then probelocation="ZZ";
  keep monid poc probelocation;
proc sort data=tempMA10c out=tempMA10d nodupkey;
  by monid poc;
proc print data=tempMA10d(obs=50);
  where probelocation="ZZ";
  title2 "MORE THAN ONE PROBE LOCATION FOR A PM10 MONITOR IN AMP500 FILE, tempMA10d";

/** ENTER FLAG "ZZ" FOR PM2.5 MONITORS IN AMP500 FILE WITH MORE THAN ONE MONITOR OBJECTIVE CODE.
**/
proc sort data=tempME25 out=tempME25a nodupkey;
  by monid poc monobjective;
proc summary data=tempME25a;
  by monid poc;
  output out=tempME25b;
data tempME25c;
  merge tempME25b tempME25;
```

```
  by monid poc;
  if _freq_ > 1 then monobjective="ZZ";
  keep monid poc monobjective;
proc sort data=tempME25c out=tempME25d nodupkey;
  by monid poc;
proc print data=tempME25d(obs=50);
  where monobjective="ZZ";
  title2 "MORE THAN ONE MONITOR OBJECTIVE FOR A MONITOR IN AMP500 FILE, tempME25d";

/** ENTER FLAG "ZZ" FOR PM10 MONITORS IN AMP500 FILE WITH MORE THAN ONE MONITOR OBJECTIVE CODE.
**/
proc sort data=tempME10 out=tempME10a nodupkey;
  by monid poc monobjective;
proc summary data=tempME10a;
  by monid poc;
  output out=tempME10b;
data tempME10c;
  merge tempME10b tempME10;
  by monid poc;
  if _freq_ > 1 then monobjective="ZZ";
  keep monid poc monobjective;
proc sort data=tempME10c out=tempME10d nodupkey;
  by monid poc;
proc print data=tempME10d(obs=50);
  where monobjective="ZZ";
  title2 "MORE THAN ONE MONITOR OBJECTIVE FOR A MONITOR IN AMP500 FILE, tempME10d";


/*****************************************************************************
READ IN COMPLETE YEAR TEMPLATE FOR PM (366 DAYS -- INCLUDES FEB 29TH FOR LEAP YEAR).
NOTE: CAPMS IS HARDWIRED TO ACCEPT 365 DAYS, SO FEB29 VALUES ARE AVERAGED WITH FEB28
VALUES FOR PM.
*****************************************************************************/
data temp0;
  length day $ 4 monid $ 9;
  infile yr366 firstobs=2 delimiter=',' missover;
  input monid $ day $;
  /** ADD ZERO TO MONID TO MAKE IT APPEAR FIRST IN THE MONITOR LIST AFTER THE MERGE BELOW. **/
  monid = "0" || monid;
  if length(day)=3 then day= "0" || day;
  keep monid day;
proc sort;
  by monid;
proc print data=temp0(obs=5);
  title2 "TEMPLATE FOR FULL YEAR, temp0";


/*****************************************************************************
MACRO TO LOOP THROUGH THE DATA SETS.
*****************************************************************************/
%macro getdata;

        %do i=1 %to 28;

/*****************************************************************************
READ IN AMP501 FILES. NOTE THAT FOR SAMPLE DURATION: 1=HOURLY DATA, AND 7=DAILY
AVERAGE DATA. WITHIN A GIVEN SAMPLE DURATION, AVERAGE OBSERVATIONS OVER THE COURSE
OF A DAY.  THAT IS, IF WE HAVE TWO 24-HOUR AVERAGES FOR A GIVEN DAY, JUST TAKE THE
AVERAGE OF THESE TWO VALUES.
*****************************************************************************/

* start of Ed's edit;
```

```
* note:
The format for the 2002 PM 2.5 (88101) comma delimited files is:
State Code, County Code, Site ID, Parameter, POC, Method, Unit, Sample Duration,
Date (YYYYMMDD), Start Time (Hour), Sample Value, Qualifier-1, Qualifier-2, Qualifier-3,
Qualifier-4, Qualifier-5, Qualifier-6, Qualifier-7, Qualifier-8, Qualifier-9, Qualifier-10 ;

* for 2002 PM2.5 lcl data ( in loop), read in two source files (infile 14 and 29) and merge them;
%if &i=14 %then %do;
  %do j=14 %to 29 %by 15;
    data temp1_t&j(keep=monid state county siteid poc parameter sampleduration unit method year
                  day samplefreq samplevalue)
      junk1_t&j(keep=monid poc parameter method sampleduration day samplevalue);
      infile infile&j firstobs=1 missover lrecl=6000 dsd;
      length state $ 2 county $ 3 siteid $ 4 parameter $ 5
           poc $ 1 unit $ 3 date $ 8 year $ 4 starttime $ 5
           samplefreq $ 8 monid $ 9 x1-x10 $ 2;
      input  state $ county $ siteid $ parameter $ poc $ method unit $ sampleduration
           date $ starttime $ samplevalue x1-x10 $;
      * note: samplefreq not available in input file. setting it = to 'blank' for compatibility with rest of code - ed;
      samplefreq='';

      /* KEEP DATA FOR RELEVANT PARAMETER. */
                if parameter="88101";

              * Miami-Dade county correction;
       if state="12" and county="025" then county="086";

       /* CREATE MONITOR ID. DO NOT INCLUDE POC CODE. */
       monid = state || county || siteid;

            /* DROP MISSING OBSERVATIONS. */
            if samplevalue=. then delete;

            /* CALCULATE YEAR AND DAY, AND KEEP DATA FOR RELEVANT YEAR. */
            year = substr(date,1,4);
            day = substr(date,5,4);

            if year="2002";

            /* OUTPUT IF SAMPLE DURATION IS HOURLY (1) OR DAILY (7), WHERE DAILY IS NONNEGATIVE.*/
      if sampleduration=1 or (sampleduration=7 and samplevalue >= 0) then output temp1_t&j;
            else output junk1_t&j;
    run;
  %end;

  data temp1_&i;
    set temp1_t14 temp1_t29;
  run;
  data junk1_&i;
    set junk1_t14 junk1_t29;
  run;

%end;

* read-in and process all other data;
%else %do;
data temp1_&i(keep=monid state county siteid poc parameter sampleduration unit method year
              day samplefreq samplevalue)
   junk1_&i(keep=monid poc parameter method sampleduration day samplevalue);

  infile infile&i firstobs=1 delimiter='|' missover lrecl = 6000 dsd;
  length transactioncode $ 2 actioncode $ 1 state $ 2 county $ 3 siteid $ 4 parameter $ 5
```

```
      poc $ 1 unit $ 3 date $ 8 year $ 4 starttime $ 5
      samplefreq $ 8 monid $ 9;
  input transactioncode $ @;
  if transactioncode="RD" then do;
    input actioncode $ state $ county $ siteid $ parameter $ poc $ sampleduration unit $
        method date $ starttime $ samplevalue x1 $ samplefreq $;
  end;

    /* KEEP DATA FOR RELEVANT PARAMETER. */
            if &i < 8 then do;
              if parameter="81104";
            end;
            if 8 <= &i < 15 then do;
              if parameter="88101";
            end;
            if 15 <= &i < 22 then do;
              if parameter="81102";
            end;
            if 22 <= &i then do;
              if parameter="85101";
            end;

             * Miami-Dade county correction;
    if state="12" and county="025" then county="086";

    /* CREATE MONITOR ID. DO NOT INCLUDE POC CODE. */
    monid = state || county || siteid;

            /* DROP MISSING OBSERVATIONS. */
            if samplevalue=. then delete;

            /* CALCULATE YEAR AND DAY, AND KEEP DATA FOR RELEVANT YEAR. */
            year = substr(date,1,4);
            day = substr(date,5,4);
            if &i in(1,8,15,22) then do;
              if year="1996";
            end;
            if &i in(2,9,16,23) then do;
              if year="1997";
            end;
            if &i in(3,10,17,24) then do;
              if year="1998";
            end;
            if &i in(4,11,18,25) then do;
              if year="1999";
            end;
            if &i in(5,12,19,26) then do;
              if year="2000";
            end;
            if &i in(6,13,20,27) then do;
              if year="2001";
            end;
            if &i in(7,14,21,28) then do;
              if year="2002";
            end;

/* OUTPUT IF SAMPLE DURATION IS HOURLY (1) OR DAILY (7), WHERE DAILY IS NONNEGATIVE.*/
    if sampleduration=1 or (sampleduration=7 and samplevalue >= 0) then output temp1_&i;
            else output junk1_&i;
run;
%end;
```

```
* end of Ed's edit;

/* CHECK VARIOUS VARIABLES: SAMPLE DURATION, UNIT, METHOD, STATE, AND SAMPLE VALUES.*/
proc freq data=temp1_&i;
  tables sampleduration unit method state;
  title2 "CHECK VARIOUS VARIABLES, temp1";
proc means data=temp1_&i;
  var samplevalue;
  title2 "SAMPLEVALUE DISTRIBUTION, temp1";

/* CHECK IF MORE THAN ONE SAMPLE DURATION ON A GIVEN DAY FOR A GIVEN MONITOR+METHOD PAIR.*/
proc sort data=temp1_&i;
  by monid poc day method sampleduration;
proc summary data=temp1_&i;
  by monid poc day method sampleduration;
  var samplevalue;
  output out=new1_&i mean=;
proc summary data=new1_&i;
  by monid poc day method;
  var samplevalue;
  output out=new1b_&i mean=;
proc print data=temp1_&i(obs=5);
  title2 "RAW DATA WITH BOTH HOURLY AND DAILY AVERAGE DATA, temp1";
proc print data=new1_&i(obs=5);
  title2 "DAILY AVERAGE DATA, new1";
proc print data=new1b_&i(obs=5);
  by _freq_;
  title2 "CHECK IF MORE THAN ONE SAMPLE DURATION ON A GIVEN DAY, new1b";
run;

/*******************************************************************************
PUT DATA IN CORRECT UNITS (ug/m3) AND CHECK FOR REALLY LARGE OR NEGATIVE VALUES.
*******************************************************************************/
data temp2_&i(keep=monid poc method day sampleduration samplevalue)
    junk2_&i(keep=monid poc method unit day samplevalue);
  set new1_&i;
  /* UNIT=105: mg/m3 (milligram) AT 25 C */
  if unit = "005" then samplevalue=samplevalue*1000;
  if unit in("001","005","105") then output temp2_&i;
  else output junk2_&i;
proc print data=junk2_&i(obs=10);
  title2 "DATA WITH STRANGE UNITS, SHOULD NOT PRINT, junk2";

data temp2b_&i;
  set temp2_&i;
  if samplevalue > 1000;
proc print data=temp2b_&i(obs=50);
  title2 "SAMPLE OF LARGE VALUES, temp2b";


/*******************************************************************************
CHECK FOR MORE THAN ONE METHOD FOR EACH MONITOR, CHECK FOR MORE THAN ONE SAMPLE
FREQUENCY FOR EACH MONITOR+METHOD PAIR, AND CHECK FOR MORE THAN ONE SAMPLE DURATION FOR
EACH MONITOR+METHOD PAIR.
*******************************************************************************/
/** CHECK METHOD VARIABLE. **/
proc sort data=temp2_&i nodupkey out=temp2c_&i;
  by monid poc method;
proc summary data=temp2c_&i;
  by monid poc;
  output out=temp2cc_&i;
data temp2ccc_&i;
```

```
  set temp2cc_&i;
  if _freq_ > 1;
proc print data=temp2ccc_&i(obs=50);
  title2 "MORE THAN ONE METHOD FOR A MONITOR, temp2ccc";

/** CHECK SAMPLE FREQUENCY VARIABLE. **/
proc sort data=temp2_&i nodupkey out=temp2d_&i;
  by monid poc method samplefreq;
proc summary data=temp2d_&i;
  by monid poc method;
  output out=temp2dd_&i;
run;
* a redundant sort procedure to eliminate merge error;
proc sort data=temp2_&i;
by monid poc method;
run;

data temp2ddd_&i;
  merge temp2dd_&i temp2_&i;
  by monid poc method;
  if _freq_ > 1 then samplefreq="ZZ";
proc sort data=temp2ddd_&i nodupkey;
  by monid poc method;
proc print data=temp2ddd_&i(obs=50);
  where samplefreq="ZZ";
  title2 "MORE THAN ONE SAMPLE FREQUENCY FOR A MONITOR+METHOD PAIR, temp2ddd";

/** CHECK SAMPLE DURATION VARIABLE. **/
proc sort data=temp2_&i;
  by monid poc method sampleduration;
proc summary data=temp2_&i;
  by monid poc method sampleduration;
  var samplevalue;
  output out=temp2e_&i mean=;
proc summary data=temp2e_&i;
  by monid poc method;
  var samplevalue;
  output out=temp2ee_&i n=;
data temp2eee_&i;
  set temp2ee_&i;
  if _freq_ > 1;
proc print data=temp2eee_&i;
  title2 "MORE THAN ONE SAMPLE DURATION FOR A MONITOR, temp2eee";


/*******************************************************************************
IF WE HAVE DAILY VALUES BASED ON MORE THAN ONE SAMPLE DURATION, CHOOSE SAMPLEDURATION=7
IN PREFERENCE TO SAMPLEDURATION=1.
*******************************************************************************/
/** SEPERATE DAILY VALUES. **/
data temp3_&i;
  set temp2_&i;
  if sampleduration=7;
  keep monid poc day sampleduration method samplevalue;
/** MAKE DATA SET OF HOURLY VALUES FOR DAYS AND MONITORS WHERE NO DAILY VALUES EXIST. **/
data temp4_&i;
  merge temp3_&i(in=check1) temp2_&i(in=check2);
  by monid poc day method;
  if check1=0 and check2=1;
  keep monid poc day sampleduration method samplevalue;
/** MERGE THE TWO. **/
data temp5_&i;
```

```
  merge temp3_&i temp4_&i;
  by monid poc day method;
  keep monid poc day sampleduration method samplevalue;


/*******************************************************************************
PUT IN FAKE MONITOR TO FILL OUT THE 366 DAYS OF DATA (FOR TRANSPOSE).
TRANSPOSE THE DATA. RESULTING FILE HAS VARIABLES MONID, POC, METHOD, AND DAYS 1 THROUGH 366.
DELETE DUMMY MONITOR.
CALCULATE FEBRUARY 28TH DATA AS AVERAGE OF FEBRUARY 28TH AND 29TH SAMPLING DATA, AND THEN
DELETE FEBRUARY 29TH DATA.
*******************************************************************************/
data temp6_&i;
  merge temp5_&i temp0;
  by monid;
  keep monid poc method samplevalue day;
proc sort data=temp6_&i;
  by monid poc method;
proc transpose data=temp6_&i out=new6_&i name=link;
  by monid poc method;
  id day;
data temp6b_&i(drop=_0229) junk6b_&i(keep=monid poc method _0229);
  set new6_&i;
  if monid = "010000000" then delete;
  if _0228 ne . or _0229 ne . then _0228=mean(_0228, _0229);


/*******************************************************************************
MERGE IN LATITUDE AND LONGITUDE. CHECK FOR MISSING AND ZERO LATITUDE AND LONGITUDE VALUES.
*******************************************************************************/
data temp7_&i
    junk7_&i (keep=monid poc latitude longitude method);
  merge temp6b_&i(in=chek1) tempAA9(in=chek2 keep=monid latitude longitude);
  by monid;
  /** CHECK IF MISSING OR HAVE ZEROS FOR LATITUDE AND LONGITUDE DATA. **/
  if (chek1=1 and chek2=0) or (chek1=1 and chek2=1 and (latitude=0 or longitude=0 or latitude=. or longitude=.)) then output
junk7_&i;
  /** OUTPUT DATA OF INTEREST. **/
  else if chek1=1 then output temp7_&i;
proc means data=temp7_&i;
  var latitude longitude;
  title2 "LATITUDE AND LONGITUDE FOR MONITORS, temp7";
proc print data=junk7_&i;
  var monid poc latitude longitude;
  title2 "MISSING OR ZERO LATITUDE AND LONGITUDE, SHOULD NOT PRINT, junk7";


/*******************************************************************************
MERGE IN PROBE LOCATION, MONITOR OBJECTIVE, LAND USE, LOCATION SETTING, SAMPLING FREQUENCY.
MONITOR+METHOD PAIRS WITH MORE THAN ONE PROBE LOCATION, MONITOR OBJECTIVE, LAND USE,
LOCATION SETTING, AND/OR SAMPLING FREQUENCY HAVE 'ZZ' FLAG IN CORRESPONDING FIELD.
*******************************************************************************/
/** ADD IN PROBE LOCATION. **/
data temp8a_&i;
  if &i<9 then do;
    merge temp7_&i(in=check1) tempMA25d(keep=monid poc probelocation);
    by monid poc;
    if check1=1;
  end;
  if &i>8 then do;
    merge temp7_&i(in=check1) tempMA10d(keep=monid poc probelocation);
    by monid poc;
```

```
    if check1=1;
  end;
/** ADD IN MONITOR OBJECTIVE. **/
data temp8b_&i;
  if &i<15 then do;
    merge temp8a_&i(in=check1) tempME25d(keep=monid poc monobjective);
    by monid poc;
    if check1=1;
  end;
  if &i>=15 then do;
    merge temp8a_&i(in=check1) tempME10d(keep=monid poc monobjective);
    by monid poc;
    if check1=1;
  end;
/** ADD IN LAND USE AND LOCATION SETTING. **/
data temp8c_&i;
  merge temp8b_&i(in=check1) tempAA9(keep=monid landuse locsetting);
  by monid;
  if check1=1;
/** ADD IN SAMPLING FREQUENCY.
   CREATE FINAL DATA FILE. PLACE BACK PARAMETER AND YEAR VARIABLES, AND CREATE SINGLE
   VARIABLE MONID15 WITH STATE CODE + COUNTY CODE + SITE ID + PARAMETER + POC. **/
data temp9_&i;
  merge temp8c_&i(in=check1) temp2ddd_&i(keep=monid poc method samplefreq);
  by monid poc method;
  if &i < 8 then parameter="81104";
  if 8 <= &i < 15 then parameter="88101";
  if 15 <= &i < 22 then parameter="81102";
  if 22 <= &i then parameter="85101";
  if &i in(1,8,15,22) then year="1996";
  if &i in(2,9,16,23) then year="1997";
  if &i in(3,10,17,24) then year="1998";
  if &i in(4,11,18,25) then year="1999";
  if &i in(5,12,19,26) then year="2000";
  if &i in(6,13,20,27) then year="2001";
  if &i in(7,14,21,28) then year="2002";

  monid15 = monid || parameter || poc;
  if check1=1;
  keep year monid15 latitude longitude method landuse locsetting probelocation monobjective
      samplefreq _0101-_0131 _0201-_0228 _0301-_0331 _0401-_0430 _0501-_0531 _0601-_0630
      _0701-_0731 _0801-_0831 _0901-_0930 _1001-_1031 _1101-_1130 _1201-_1231;


/*****************************************************************************
EXPORT FINAL DATA.
*****************************************************************************/
data _null_;
  set temp9_&i;
  file outf&i lrecl=10000;
  heading="year, monid, latitude, longitude, method, landuse, locationsetting, probelocation, monitorobjective, samplefrequency,
365dailyvalues, ";
  if _n_=1 then put heading;
  put year ',' monid15 ',' latitude ',' longitude ',' method ','
      landuse ',' locsetting ',' probelocation ',' monobjective ',' samplefreq (_0101-_0131) (',') (_0201-_0228) (',')
      (_0301-_0331) (',') (_0401-_0430) (',') (_0501-_0531) (',') (_0601-_0630) (',') (_0701-_0731) (',') (_0801-_0831) (',')
      (_0901-_0930) (',') (_1001-_1031) (',') (_1101-_1130) (',') (_1201-_1231) (',');


/*****************************************************************************
EXPORT JUNK DATA.
*****************************************************************************/
```

```
/** PUT ALL JUNK FILES IN SAME FORMAT WITH SAME VARIABLES: MONID, POC, LATITUDE, LONGITUDE,
        METHOD, SAMPLE DURATION, UNIT, DAY AND SAMPLE VALUE. **/
data junk6b_&i;
  set junk6b_&i;
  if _0229 ne .;
  day="0229";
data junk6b_&i;
  set junk6b_&i;
  sampleduration=.;
  unit="";
  latitude=.;
  longitude=.;
  keep monid poc latitude longitude method sampleduration unit day;
proc sort data=junk7_&i nodupkey out=junk7_&i;
  by monid poc method;
data _null_;
  set junk6b_&i junk1_&i junk2_&i junk7_&i;
  file outjk&i lrecl=5000;
  heading="monid, poc, latitude, longitude, method, sampleduration, unit, day, samplevalue";
  if _n_=1 then put heading;
  put monid ',' poc ',' latitude ',' longitude ',' method ',' sampleduration ',' unit ','
     day ',' samplevalue;

run;

%end;

%mend;

%getdata;
```

A.4.2 SAS Code for Processing Raw Ozone Monitor Data for Input to BenMAP

```
title "file: amp500_501_o3data.sas";

/*****************************************************************************
02/12/03.
Etienne Gabel.

This SAS code reads in from raw AMP500 file which contains monitor description data,
and from raw AMP501 file which contains formatted hourly ozone monitor sampling data.
The code prepares a data set which contains the hourly data within a single array of
24*365 = 8760 hours. Each observation (row) corresponds to a unique monitor (including
poc code) and method pair.
February 29th data from leap year are dropped, since CAPMS is hardwired to accept 8760
hours of ozone data.

7/16/03 - ed al-hussainy - added 1997, 1998 and 2002 data
                  note that 2002 input data is fragmented into 4 files
                  re-organized junk export macro
7/28/03 - ed - modified macro loop to prevent resource overload (too many temp files)
*****************************************************************************/

libname out "J:\Ozone\Output\";

filename yr8760 "J:\Ozone\Data\year_8760_hours.csv";

filename infile1 "J:\Ozone\Data\EG44201-1996.TXT";
```

```
filename infile2 "J:\Ozone\Data\EG44201-1999.TXT";
filename infile3 "J:\Ozone\Data\EG44201-2000.TXT";
filename infile4 "J:\Ozone\Data\EG44201-2001.TXT";
filename infile5 "J:\Ozone\Data\RD_501_44201_1997.TXT";
filename infile6 "J:\Ozone\Data\RD_501_44201_1998.TXT";
filename infile7 "J:\Ozone\Data\ozone-reg1-3-2002.txt";
filename infile8 "J:\Ozone\Data\ozone-reg4-5-2002.txt";
filename infile9 "J:\Ozone\Data\ozone-reg6-8-2002.txt";
filename infile10 "J:\Ozone\Data\ozone-reg9-10-25-2002.txt";

filename outf1 "J:\Ozone\Output\Ozone96.TXT";
filename outf2 "J:\Ozone\Output\Ozone99.TXT";
filename outf3 "J:\Ozone\Output\Ozone00.TXT";
filename outf4 "J:\Ozone\Output\Ozone01.TXT";
filename outf5 "J:\Ozone\Output\Ozone97.TXT";
filename outf6 "J:\Ozone\Output\Ozone98.TXT";
filename outf7 "J:\Ozone\Output\Ozone02.TXT";

filename outjk1 "J:\Ozone\Output\Ozone96junk.TXT";
filename outjk2 "J:\Ozone\Output\Ozone99junk.TXT";
filename outjk3 "J:\Ozone\Output\Ozone00junk.TXT";
filename outjk4 "J:\Ozone\Output\Ozone01junk.TXT";
filename outjk5 "J:\Ozone\Output\Ozone97junk.TXT";
filename outjk6 "J:\Ozone\Output\Ozone98junk.TXT";
filename outjk7 "J:\Ozone\Output\Ozone02junk.TXT";

options missing = '.';

/*************************************************************************************
READ IN AMP500 FILE. CREATE MONITOR ID VARIABLE WITH STATE, COUNTY, AND SITE ID.
TRANSACTION CODE 'AA' INCLUDES MONID, LAND USE, LOCATION SETTING, LATITUDE AND LONGITUDE
VARIABLES.  TRANSACTION CODE 'MA' INCLUDES MONID, POC AND PROBE LOCATION VARIABLES.
TRANSACTION CODE 'ME' INCLUDES MONID, POC AND MONITOR OBJECTIVE VARIABLES.
*************************************************************************************/

* start of Ed's edit;

/* AMP 500 data */
%let netdata=H:\ENVR\MCCUBBIN\capms_revision\data_bases\air_monitor\sas_work\data\AMP500 file -- contains monitor
description info;

filename amp500_1 "&netdata\EG_500_SELPARM_1997_2002.TXT";
filename amp500_2 "&netdata\AMP500.TXT";

* read-in new and old AMP500 data;

%macro read_amp500(param);

data tempAA_&param;
  infile amp500_&param firstobs=1 delimiter='|' missover lrecl = 6000 dsd;
  length transactioncode $ 2 state $ 2 county $ 3 siteid $ 4 poc $ 1 x1-x14 $ 5 landuse $ 20 locsetting $ 50
       probelocation $ 20 monobjective $ 50 monid $ 9;
  input transactioncode $ @;

  if transactioncode="AA" then do;
    input actioncode $ state $ county $ siteid $ latitude longitude x1-x14 $ landuse $
      locsetting $;
  end;

            if transactioncode="MA" then do;
    input actioncode $ state $ county $ siteid $ parameter $ poc $ x1-x4 $ probelocation $;
              if probelocation="" then delete;
```

```
            end;

            if transactioncode="ME" then do;
    input actioncode $ state $ county $ siteid $ parameter $ poc $ monobjective $;
              if monobjective="" then delete;
  end;

  drop x1-x14 actioncode;

  * Miami-Dade county correction;
  if state="12" and county="025" then county="086";

  monid = state || county || siteid;

  if transactioncode in ("AA" "MA" "ME");
run;

proc sort data=tempAA_&param;
            by monid;
run;

%mend read_amp500;

%read_amp500(1);
%read_amp500(2);

* merge new and old AMP500 data, separate data based on transactioncode and parameter;
data tempAA(keep=monid state county siteid latitude longitude landuse locsetting)
    tempMA(keep=monid state county siteid parameter poc probelocation)
    tempME(keep=monid state county siteid parameter poc monobjective);

            * note the merge order - tempAA_1 overwrites tempAA_2 in case of duplicate entries;
            merge  tempAA_2(in=dum2) tempAA_1(in=dum1);
            by monid;

            if transactioncode="AA" then output tempAA;

            else if transactioncode="MA" then do;
    if parameter="44201" then output tempMA;
            end;

            else if transactioncode="ME" then do;
    if parameter="44201" then output tempME;
            end;
run;


* end of Ed's edit;


proc sort data=tempAA;
  by monid;
proc sort data=tempMA;
  by monid poc;
proc sort data=tempME;
  by monid poc;

/** ENTER FLAG "ZZ" FOR MONITORS IN AMP500 FILE WITH MORE THAN ONE LAND USE CODE. **/
proc sort data=tempAA out=tempAA1 nodupkey;
  by monid landuse;
proc summary data=tempAA1;
  by monid;
```

```
  output out=tempAA2;
data tempAA3;
  merge tempAA2 tempAA;
  by monid;
  if _freq_ > 1 then landuse="ZZ";
  keep monid latitude longitude landuse locsetting;
proc sort data=tempAA3 out=tempAA4 nodupkey;
  by monid;
proc print data=tempAA4(obs=50);
  where landuse="ZZ";
  title2 "MORE THAN ONE LAND USE FOR A MONITOR IN AMP500 FILE, tempAA4";

/** ENTER FLAG "ZZ" FOR MONITORS IN AMP500 FILE WITH MORE THAN ONE LOCATION SETTING CODE. **/
proc sort data=tempAA out=tempAA5 nodupkey;
  by monid locsetting;
proc summary data=tempAA5;
  by monid;
  output out=tempAA6;
data tempAA7;
  merge tempAA6(in=check1) tempAA;
  by monid;
  if check1=1 and _freq_ > 1 then locsetting="ZZ";
  keep monid latitude longitude landuse locsetting;
proc sort data=tempAA7 out=tempAA8 nodupkey;
  by monid;
proc print data=tempAA8(obs=50);
  where locsetting="ZZ";
  title2 "MORE THAN ONE LOCATION SETTING FOR A MONITOR IN AMP500 FILE, tempAA8";

/** MAKE SINGLE FILE WITH TRANSACTION AA DATA, WITH FLAG FOR REPEATING LAND USE OR LOCATION
SETTING CODES. **/
data tempAA9;
  merge tempAA4(drop=locsetting) tempAA8(drop=landuse);
  by monid;
  keep monid latitude longitude landuse locsetting;

/** ENTER FLAG "ZZ" FOR MONITORS IN AMP500 FILE WITH MORE THAN ONE PROBE LOCATION CODE. **/
proc sort data=tempMA out=tempMA1 nodupkey;
  by monid poc probelocation;
proc summary data=tempMA1;
  by monid poc;
  output out=tempMA2;
data tempMA3;
  merge tempMA2 tempMA;
  by monid poc;
  if _freq_ > 1 then probelocation="ZZ";
  keep monid poc probelocation;
proc sort data=tempMA3 out=tempMA4 nodupkey;
  by monid poc;
proc print data=tempMA4(obs=50);
  where probelocation="ZZ";
  title2 "MORE THAN ONE PROBE LOCATION FOR A MONITOR IN AMP500 FILE, tempMA4";

/** ENTER FLAG "ZZ" FOR MONITORS IN AMP500 FILE WITH MORE THAN ONE MONITOR OBJECTIVE CODE. **/
proc sort data=tempME out=tempME1 nodupkey;
  by monid poc monobjective;
proc summary data=tempME1;
  by monid poc;
  output out=tempME2;
data tempME3;
  merge tempME2 tempME;
  by monid poc;
```

```
  if _freq_ > 1 then monobjective="ZZ";
  keep monid poc monobjective;
proc sort data=tempME3 out=tempME4 nodupkey;
  by monid poc;
proc print data=tempME4(obs=50);
  where monobjective="ZZ";
  title2 "MORE THAN ONE MONITOR OBJECTIVE FOR A MONITOR IN AMP500 FILE, tempME4";
run;

/*****************************************************************************
READ IN COMPLETE YEAR TEMPLATE FOR OZONE (365*24 = 8760 HOURS). TEMPLATE EXCLUDES
FEBRUARY 29TH FOR LEAP YEARS, SINCE CAPMS IS HARDWIRED TO ACCEPT 8760 HOURS.
*****************************************************************************/
data temp0;
  length monid $ 9 hour $ 8 sampletime $ 9;
  infile yr8760 firstobs=2 delimiter=',' missover;
  input monid $ day $ time $ sampletime $ hour $;
  /** ADD ZERO TO MONID TO MAKE IT APPEAR FIRST IN THE MONITOR LIST AFTER THE MERGE BELOW. **/
  monid = "0" || monid;
  keep monid sampletime hour;
proc print data=temp0(obs=5);
  title2 "TEMPLATE FOR FULL YEAR, temp0";
run;

* start of Ed's edit;

/*****************************************************************************
MACRO TO LOOP THROUGH DATA SETS.
*****************************************************************************/
%macro getdata;
%do i=1 %to 7;

/*****************************************************************************
READ IN AMP501 FILES. NOTE: DATA ONLY HAS SAMPLE DURATION = 1 (=HOURLY DATA).
*****************************************************************************/

%* for 2002 data ( in loop), read in 4 source files (infile 7-10) and merge them;
%if &i=7 %then %do;
  %do j=7 %to 10; %*read-in 2002 data (4 files);
    data temp1_&j(keep=monid state county siteid poc parameter sampleduration unit method year
                  samplefreq sampletime samplevalue)
      junk1_&j(keep=monid poc parameter method sampleduration sampletime samplevalue);
      infile infile&j firstobs=1 delimiter='|' missover lrecl = 6000 dsd;
      length transactioncode $ 2 actioncode $ 1 state $ 2 county $ 3 siteid $ 4 parameter $ 5
           poc $ 1 sampleduration unit $ 3 method 8 date $ 8 year $ 4 starttime $ 5
           samplefreq $ 8 monid $ 9;
      input transactioncode $ @;
      if transactioncode="RD" then do;
        input actioncode $ state $ county $ siteid $ parameter $ poc $ sampleduration unit $
            method date $ starttime $ samplevalue x1 $ samplefreq $;
        if parameter="44201";

            * Miami-Dade county correction;
        if state="12" and county="025" then county="086";

            /* CREATE MONITOR ID. MONID VARIABLE EXCLUDES POC CODE. */
        monid = state || county || siteid;

            /* DROP MISSING OBSERVATIONS. */
            if samplevalue=. then delete;

            /* CALCULATE YEAR AND KEEP DATA FOR RELEVANT YEAR. */
```

```
            year = substr(date,1,4);
                        if year="2002";

            /* MAKE SAMPLE TIME A SINGLE VARIABLE (FOR TRANSPOSE BELOW) */
          sampletime = substr(date,5,4) || starttime;

            /* OUTPUT IF SAMPLE DURATION IS HOURLY (1).*/
          if sampleduration=1 then output temp1_&j;
          else output junk1_&j;
        end;
      run;
    %end;
    data temp1;
      set temp1_7 temp1_8 temp1_9 temp1_10;
    run;
    data junk1;
      set junk1_7 junk1_8 junk1_9 junk1_10;
    run;
%end;

* read-in and process all other data;
%if &i ^= 7 %then %do;
  data temp1(keep=monid state county siteid poc parameter sampleduration unit method year
                  samplefreq sampletime samplevalue)
       junk1(keep=monid poc parameter method sampleduration sampletime samplevalue);
    infile infile&i firstobs=1 delimiter='|' missover lrecl = 6000 dsd;
    length transactioncode $ 2 actioncode $ 1 state $ 2 county $ 3 siteid $ 4 parameter $ 5
          poc $ 1 sampleduration unit $ 3 method 8 date $ 8 year $ 4 starttime $ 5
          samplefreq $ 8 monid $ 9;
    input transactioncode $ @;
    if transactioncode="RD" then do;
      input actioncode $ state $ county $ siteid $ parameter $ poc $ sampleduration unit $
            method date $ starttime $ samplevalue x1 $ samplefreq $;
      if parameter="44201";

             %* Miami-Dade county correction;
      if state="12" and county="025" then county="086";

            /* CREATE MONITOR ID. MONID VARIABLE EXCLUDES POC CODE. */
      monid = state || county || siteid;

            /* DROP MISSING OBSERVATIONS. */
            if samplevalue=. then delete;

            /* CALCULATE YEAR AND KEEP DATA FOR RELEVANT YEAR. */
            year = substr(date,1,4);
            if &i=1 then do;
              if year="1996";
            end;
            if &i=2 then do;
              if year="1999";
            end;
            if &i=3 then do;
              if year="2000";
            end;
            if &i=4 then do;
              if year="2001";
            end;
       if &i=5 then do;
              if year="1997";
            end;
       if &i=6 then do;
```

```
            if year="1998";
          end;

          /* MAKE SAMPLE TIME A SINGLE VARIABLE (FOR TRANSPOSE BELOW) */
      sampletime = substr(date,5,4) || starttime;

          /* OUTPUT IF SAMPLE DURATION IS HOURLY (1).*/
      if sampleduration=1 then output temp1;
      else output junk1;
    end;
  run;
%end;



/** CHECK FOR MORE THAN ONE SAMPLE VALUE FOR A GIVEN MONITOR, DAY AND HOUR. **/
proc sort data=temp1;
  by monid poc method sampleduration unit samplefreq sampletime;
proc summary data=temp1;
  by monid poc method sampleduration unit samplefreq sampletime;
  var samplevalue;
  output out=new1 mean=;
  /*
proc print data=new1_&i;
  where _freq_ > 1;
  title2 "MORE THAN ONE SAMPLE VALUE FOR SAME DAY AND HOUR, SHOULD NOT PRINT, new1";
  */

/*****************************************************************************
PUT DATA IN CORRECT UNITS (PPB) AND CHECK FOR REALLY LARGE OR NEGATIVE VALUES.
*****************************************************************************/
data temp2(keep=monid poc method sampletime samplefreq samplevalue)
    junk2(keep=monid poc method unit sampletime samplevalue);
  set new1;

  /* UNIT=1: ug/m3 AT 25 C */
  if unit = "001" then samplevalue=samplevalue*.51;
  /* UNIT=40: PARTS PER HUNDRED MILLION */
  if unit = "040" then samplevalue=samplevalue*10;
  /* UNIT=007: PARTS PER MILLION */
  if unit = "007" then samplevalue=samplevalue*1000;

  if unit in("001","007","008","040") then output temp2;
  else output junk2;

data temp2b;
  set temp2;
  if samplevalue > 1000;
  /*
proc print data=temp2b_&i(obs=50);
  title2 "SAMPLE OF LARGE VALUES, temp2b";
  */

/*****************************************************************************
CHECK FOR MORE THAN ONE METHOD FOR EACH MONITOR, AND CHECK FOR MORE THAN ONE SAMPLE
FREQUENCY FOR EACH MONITOR+METHOD PAIR.
*****************************************************************************/
/** CHECK METHOD VARIABLE. **/
proc sort data=temp2 nodupkey out=temp2c;
  by monid poc method;
proc summary data=temp2c;
  by monid poc;
```

```
  output out=temp2cc;
data temp2ccc;
  set temp2cc;
  if _freq_ > 1;
  /*
proc print data=temp2ccc_&i(obs=50);
  title2 "MORE THAN ONE METHOD FOR A MONITOR, temp2ccc";
  */
/** CHECK SAMPLE FREQUENCY VARIABLE. **/
proc sort data=temp2 nodupkey out=temp2d;
  by monid poc method samplefreq;
proc summary data=temp2d;
  by monid poc method;
  output out=temp2dd;
data temp2ddd;
  merge temp2dd temp2;
  by monid poc method;
  if _freq_ > 1 then samplefreq="ZZ";
proc sort data=temp2ddd nodupkey;
  by monid poc method;
  /*
proc print data=temp2ddd_&i(obs=50);
  where samplefreq="ZZ";
  title2 "MORE THAN ONE SAMPLE FREQUENCY FOR A MONITOR+METHOD PAIR, temp2ddd";
  */

/*****************************************************************************
MERGE WITH HOUR TEMPLATE. 365 DAYS*24 HOURS = 8760 HOURS. REPLACE SAMPLETIME VARIABLE (example
010100:00 for Jan 1rst at midnight) WITH HOUR VARIABLE (example hour1 for Jan 1rst at midnight).
*****************************************************************************/
proc sort data=temp2 sortsize=8M;
  by sampletime;
proc sort data=temp0;
  by sampletime;
data temp3;
  merge temp2(in=check1) temp0(keep=sampletime hour);
  by sampletime;
  if check1=1;
  keep monid poc method sampletime hour samplevalue;


/*****************************************************************************
PUT IN FAKE MONITOR TO FILL OUT THE 8760 HOURS OF DATA (FOR TRANSPOSE). REMOVED LATER.
REMOVE DATA FROM FEBRUARY 29TH FOR LEAP YEAR.
*****************************************************************************/
proc sort data=temp3 sortsize=8M;
  by monid poc method;
data temp3b;
  set temp0(keep=monid hour) temp3;
  keep monid poc method sampletime hour samplevalue;
data temp3c(keep=monid poc method hour samplevalue)
    junk3(keep=monid poc method sampletime samplevalue);
  set temp3b;
  /** DROP LEAP YEAR FEBRUARY 29TH VALUES. **/
  if hour="" then output junk3;
  else output temp3c;


/*****************************************************************************
TRANSPOSE THE DATA. RESULTING FILE HAS VARIABLES MONID, POC, METHOD, AND HOUR 1 THROUGH 8760.
DELETE DUMMY MONITOR.
*****************************************************************************/
```

```
/** TRANSPOSE DATA **/
proc transpose data=temp3c out=new3 name=link;
  by monid poc method;
  id hour;
data temp3d;
  set new3;
  /** DELETE DUMMY MONITOR **/
  if monid = "010000000" then delete;
  keep monid poc method hour1-hour8760;


/*****************************************************************************
MERGE IN LATITUDE AND LONGITUDE. CHECK FOR MISSING AND ZERO LATITUDE AND LONGITUDE VALUES.
*****************************************************************************/
data temp4 (keep=monid poc latitude longitude method hour1-hour8760)
    junk4 (keep=monid poc latitude longitude method);
  merge temp3d(in=chek1) tempAA9(in=chek2 keep=monid latitude longitude);
  by monid;
  /** CHECK IF MISSING OR HAVE ZEROS FOR LATITUDE AND LONGITUDE DATA. **/
  if (chek1=1 and chek2=0) or (chek1=1 and chek2=1 and (latitude=0 or longitude=0 or latitude=. or longitude=.)) then output junk4;
  /** OUTPUT DATA OF INTEREST. **/
  else if chek1=1 then output temp4;
proc univariate data=temp4;
  var latitude longitude;
  title2 "LATITUDE AND LONGITUDE FOR MONITORS, temp4";
proc print data=junk4_&i;
  var monid poc latitude longitude;
  title2 "MISSING OR ZERO LATITUDE AND LONGITUDE, SHOULD NOT PRINT, junk4";


/*****************************************************************************
MERGE IN PROBE LOCATION, MONITOR OBJECTIVE, LAND USE, LOCATION SETTING, SAMPLING FREQUENCY.
MONITOR+METHOD PAIRS WITH MORE THAN ONE PROBE LOCATION, MONITOR OBJECTIVE, LAND USE,
LOCATION SETTING, AND/OR SAMPLING FREQUENCY HAVE 'ZZ' FLAG IN CORRESPONDING FIELD.
*****************************************************************************/
/** ADD IN PROBE LOCATION. **/
data temp5a;
  merge temp4(in=check1) tempMA4(keep=monid poc probelocation);
  by monid poc;
  if check1=1;
/** ADD IN MONITOR OBJECTIVE. **/
data temp5b;
  merge temp5a(in=check1) tempME4(keep=monid poc monobjective);
  by monid poc;
  if check1=1;
/** ADD IN LAND USE AND LOCATION SETTING. **/
data temp5c;
  merge temp5b(in=check1) tempAA9(keep=monid landuse locsetting);
  by monid;
  if check1=1;
/** ADD IN SAMPLING FREQUENCY.
   CREATE FINAL DATA FILE. PLACE BACK PARAMETER AND YEAR VARIABLES, AND CREATE SINGLE
   VARIABLE MONID15 WITH STATE CODE + COUNTY CODE + SITE ID + PARAMETER + POC. **/
data temp6;
  merge temp5c(in=check1) temp2ddd(keep=monid poc method samplefreq);
  by monid poc method;
  parameter = "44201";
  if &i=1 then year="1996";
  if &i=2 then year="1999";
  if &i=3 then year="2000";
  if &i=4 then year="2001";
  if &i=5 then year="1997";
```

```
  if &i=6 then year="1998";
  if &i=7 then year="2002";
  if check1=1;
  monid15 = monid || parameter || poc;
  keep year monid15 latitude longitude method landuse locsetting probelocation monobjective
      samplefreq hour1-hour8760;


/*******************************************************************************
EXPORT FINAL DATA.
*******************************************************************************/
data _null_;
  set temp6;
  file outf&i lrecl=500000;
  heading="year, monid, latitude, longitude, method, landuse, locationsetting, probelocation, monitorobjective, samplefrequency,
8760hourlyvalues, ";
  if _n_=1 then put heading;
  put year ',' monid15 ',' latitude ',' longitude ',' method ',' landuse ',' locsetting ','
     probelocation ',' monobjective ',' samplefreq (hour1-hour8760) (',');


/*******************************************************************************
EXPORT JUNK DATA.
*******************************************************************************/
/** PUT ALL JUNK FILES IN SAME FORMAT WITH SAME VARIABLES: MONID, POC, LATITUDE, LONGITUDE,
          METHOD, SAMPLE DURATION, UNIT, SAMPLE TIME AND SAMPLE VALUE. **/
proc sort data=junk4 nodupkey out=junk4;
  by monid poc method;
data junk4;
  set junk4;
  sampleduration="";
  unit="";
  sampletime="";
  samplevalue=.;
  keep monid poc latitude longitude method sampleduration unit sampletime samplevalue;
run;

%end;
%mend getdata;

%macro dojunk;
  %do j=1 %to 7;
    data _null_;
      set junk&j._1 junk&j._2 junk&j._3 junk&j._4;
      file outjk&j lrecl=5000;
      heading="monid, poc, latitude, longitude, method, sampleduration, unit, sampletime, samplevalue";
      if _n_=1 then put heading;
      put monid ',' poc ',' latitude ',' longitude ',' method ',' sampleduration ',' unit ','
         sampletime ',' samplevalue;
    run;
  %end;
%mend dojunk;


%getdata;
%dojunk;
```

Appendix B: Population Data

BenMAP calculates health impacts at the level of U.S. counties as well as for a variety of grid structures used in air quality modeling (i.e., CMAQ 36 km, REMSAD 36km, REMSAD 12 km, and UAM-V/CAMX 12km). In this description, we use the term "population grid-cells" to refer to counties or the cells within an air quality modeling grid. The foundation for calculating the population level in the population grid-cells is the 1990 and 2000 Census block data.[1] A separate application developed by Abt Associates, called "PopGrid," combines the Census block data with any user-specified set of population grid-cells, so long as they are defined by a GIS shape file. Unfortunately, PopGrid relies on extremely large census files that are too large to include with BenMAP. If you need a custom population grid please contact Bryan Hubbell.[2]

If the geographic center of a Census block falls within a population grid-cell, PopGrid assigns the block population to this particular population grid-cell. Note that the grid-cells in air quality model, such as CMAQ, may cross multiple county boundaries. PopGrid keeps track of the total number of people by county within a particular population grid-cell. Of course, when the population grid-cell is for U.S. counties, then there is only a single county associated with the population grid-cell. With air quality models, there can clearly be multiple counties in a population grid-cell. Keeping track of the total number of people in a county is useful in the estimation of adverse health effects, where the calculation of premature mortality depends on county-level mortality rates. It is also useful in the presentation of health benefits, should you want state- and county-level estimates, as opposed to national estimates.

Within any given population grid-cell, BenMAP has 256 demographic variables, including 180 unique racial-gender-age groups: 19 age groups by gender by 5 racial groups (19*2*5=180). In addition there is an Hispanic ethnicity variable, which includes a number of different racial groups, as well as a number of variables that aggregate the population by race and gender. Exhibit B-1 presents the 256 population variables available in BenMAP. As discussed below, these variables are available for use in developing age estimates in whatever grouping desired by you.

Exhibit B-1 Demographic Groups and Variables Available in BenMAP

Racial/Ethnic Group	Gender	Age	# Variables
White, African American, Asian, American Indian, Other, Hispanic	Female, Male	<1, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+	228
All	–	<1, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+	19
All	Female, Male	–	2
White, African American, Asian, American Indian, Other, Hispanic	–	–	6
All	–	–	1

[1]Geolytics (2001a; 2002) provided the 1990 and 2000 census data.

[2] Bryan Hubbell: C339-01, USEPA Mailroom, Research Triangle Park, NC 27711. Email: hubbell.bryan@epa.gov. Telephone: 919-541-0621.

B .1 Census Data 1990

In developing the 1990 Census data, we use block-level data in conjunction with detailed tract-level data. The block data has 10 racial/ethnic groups, each divided between persons 17 and under and 18 and older, to give a total of 20 variables (Exhibit B-2). The tract-level data has essentially all of the age groupings of interest; some combining of variables is necessary to obtain the final set of variables used in BenMAP (Exhibit B-3).

Exhibit B-2 Race, Ethnicity and Age Variables in 1990 Census Block Data

Race	Ethnicity	Age
White	Hispanic / Non-Hispanic	0-17, 18+
African American	Hispanic / Non-Hispanic	0-17, 18+
Asian & Pacific Islander	Hispanic / Non-Hispanic	0-17, 18+
Native American	Hispanic / Non-Hispanic	0-17, 18+
Other	Hispanic / Non-Hispanic	0-17, 18+

Source: Geolytics (2001a).

Exhibit B-3 Race, Ethnicity and Age Variables in 1990 Census Tract Data

Type	Race / Ethnicity	Gender	Age
Initial Variables	White, African American, Asian, American Indian, Other, Hispanic	Female, Male	<1, 1-2, 3-4, 5, 6, 7-9, 10-11, 12-13, 14, **15, 16, 17, 18, 19**, 20, 21, 22-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-61, 62-64, 65-69, 70-74, 75-79, 80-84, 85+
Final Variables	White, African American, Asian, American Indian, Other, Hispanic	Female, Male	<1, 1-4, 5-9, 10-14, **15-19**, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+

Source: Geolytics (2001b).

There is a trade-off between the geographic detail and the number of variables. To protect the confidentiality of the respondents to the Census, the block data, which are the most detailed geographically, have only two age groups. The Census groupings that cover a larger geographic area, such as the blockgroup and tract, have many more demographic variables. For reasons discussed below, we do not use the blockgroup data, and instead use the tract data to estimate the population distribution in each block. In combining the information at the tract level with that at the block level, we assume that the age distribution in a particular census tract is similar to all of the blocks comprising that tract.

The first step in the process of using the tract data is to combine the "initial" set of tract variables, to form a set of "intermediate" variables that better matches the "final" set of variables desired and the characteristics of the block data. Table B-3 lists the initial and final variables.

The goal is to estimate the 38 age-gender groups for each of the five racial groups and for the Hispanic ethnic group. The approach for the five racial groups is similar, so we simply give a few examples

for the treatment of the Other group. Since Hispanics cross multiple racial groups, we give examples specific for Hispanics.

B .1.1 Other Group Estimation with 1990 Census

The first step is to get a total estimate for the Other category at the block level:

$$Other_{block,\,0-17} = (Other\ Hisp\ age_{block,\,0-17} + Other\ nonHisp\ age_{block,\,0-17})$$

$$Other_{block,\,18+} = (Other\ Hisp\ age_{block,\,18+} + Other\ nonHisp\ age_{block,\,18+})$$

We then use the intermediate variables from the tract data to apportion the two age groups in the Other racial group. For the case of children ages 1 to 4, we calculate:

$$Other_{block,\,1-4} = Other_{block,\,0-17} \cdot \frac{Other_{tract,\,1-4}}{Other_{tract,\,0-17}}$$

For the case of children ages 15-19, we use both block variables:

$$Other_{block,\,15-19} = Other_{block,\,0-17} \cdot \frac{Other_{tract,\,15-17}}{Other_{tract,\,0-17}} + Other_{block,\,18+} \cdot \frac{Other_{tract,\,18-19}}{Other_{tract,\,18+}}$$

The calculation of additional age and gender groupings is similar. However, it should be noted that in some cases there are data available at the block data, but the larger tract data indicate that no one for a particular demographic group is living there. In turn this leads to dividing by zero. This curiosity arises because the tract data is based on a 10 percent sample of the population that fills out the long form, while the block data are based on the 100 percent sample of the population that fills out the short form. In those instances, where the tract data have zero population, we use county-level data:

$$Other_{block,\,1-4} = Other_{block,\,0-17} \cdot \frac{Other_{county,\,1-4}}{Other_{county,\,0-17}}$$

Finally, we note that in a few exceptional cases we used state-level data.

B .1.2 Hispanic Group Estimation with 1990 Census

We use a similar process to estimate the Hispanic groups. We start by estimating the population at the block level for ages 0 to 17 and for ages 18 and older:

$$Hisp\ age_{block,\,0-17} = \sum_{i=1}^{5} Hisp\ age_{block,\,0-17,\,i}$$

$$Hisp\ age_{block,\,18+} = \sum_{i=1}^{5} Hisp\ age_{block,\,18+,\,i}$$

where i=1 to 5 refers to the five racial groups: White, African American, Asian, American Indian and Other.

To then estimate the 38 age-gender groupings for Hispanics, we then perform the same calculations as when estimating the 38 age-gender groups for the five racial groups. For example, in calculating the number of Hispanics ages 1 to 4:

$$Hisp_{block,\,1-4} = Hisp_{block,\,0-17} \cdot \frac{Hisp_{tract,\,1-4}}{Hisp_{tract,\,0-17}}$$

As with the five racial groups, the Hispanic ethnic group has instances where the tract data indicate zero Hispanics while the block data has a positive number. To avoid dividing by zero, we then turn to the county-level to provide information to apportion the block data between different age groups.

B .2 Census Data 2000

The 2000 Census allows respondents to choose more than one racial category, unlike the 1990 Census, which allowed only one choice. As a result there are seven racial categories in the 2000 Census versus five in the 1990 Census. To make the 2000 Census data consistent with the 1990 Census, we reduced the seven racial groups to the five used in the 1990 Census.

The initial data set at the block level includes 368 demographic groups: seven racial groups and Hispanic ethnicity, by 23 pre-defined age groups by gender (Table B-4). As discussed below, we generated for the 2000 Census the same 256 demographic variables that we generated for the 1990 Census.

Exhibit B-4 Race, Ethnicity and Age Variables in 2000 Census Block Data

Type	Race / Ethnicity	Gender	Age
Initial Variables	White Alone, Black Alone, Native American Alone, Asian Alone, Pacific Islander / Hawaiian Alone, Other Alone, Two or More Alone, Hispanic (Non-Exclusive)	Male, Female	0-5, 5-10, 10-14, 15-17, 18-19, 20, 21, 22-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-61, 62-64, 65-66, 67-69, 70-74, 75-79, 80-84 85+
Final Variables (identical to 1990 variables)	White, African American, Asian & Pacific Islander, American Indian, Other, Hispanic	Female, Male	<1, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+

Source: Geolytics (2002). Note: Some population values were errors in the original Census data (e.g., values of a billion or more). Following personal communication with Geolytics, these were set to zero.

Because the 2000 Census includes somewhat different age groupings than that for the final set generated for the 1990 Census. Age variables 15-17 and 18-19 are combined, 20, 21, and 22-24 are combined, 60-61 and 62-24 are combined, and 65-66 and 67-69 are combined at the block level. One variable, under 5 years, must be split into two variables (Under 1 and 1-4 years). Using the previous assumption that population is uniformly distributed within age groups, we apply a factor of 1/5 to create the <1 age group and 4/5 to create the 1-4 age group.

B .2.1 Matching Racial Categories in the 1990 and 2000 Censuses

Unlike the 1990 Census, respondents in the 2000 Census respondents could check more than one box for race, so the reported results included a grouping of individuals that had checked two or more racial categories. In addition, the 2000 Census separately reported the categories "Pacific Islander / Hawaiian Along" and "Asian Alone." To make the racial groupings comparable with the 1990 Census, we first combined Pacific Islander / Hawaiian Alone with the Asian Alone category to form the category Asian and Pacific Islander category. Then we divided the category Two-or-More between the remaining five racial categories.

Exhibit B-3 presents the estimated percentage of the national population by five racial groups: (1) American Indian or Alaska Native, (2) Asian or Pacific Islander, (3) Black, (4) White, and (5) Other, as well as for four combinations: (1) American Indian or Alaska Native (AIAN)/White, (2) Asian or Pacific Islander (API)/White, (3) Black/White, and (4) Other combinations. Slightly over 98 percent of individuals chose a single racial category, with 1.45 percent choosing three AIAN/White, API/White, and Black/White, and 0.30 choosing other combinations (e.g., Black/Asian). Exhibit B-3 also presents the estimated primary racial affiliation of individuals in these subcategories if they were to choose a single racial affiliation.

Exhibit B-5. Distribution of Racial Groups

Racial Category	% of Total U.S. Population [a]	% of Population in Sub-Groups by Primary Racial Affiliation [b]					
		AIAN	API	Black	White	Other	All
American Indian or Alaska Native (AIAN)	0.85	100	–	–	–	–	100
Asian or Pacific Islander (API)	3.35	–	100	–	–	–	100
Black	12.07	–	–	100	–	–	100
White	79.72	–	–	–	100	–	100
Other race	2.25	–	–	–	–	100	100
AIAN/White	0.89	12.4	–	–	80.9	6.7	100
API/White	0.30	–	34.6	–	46.9	18.4	100
Black/White	0.26	–	–	48.2	25.2	26.6	100
Other combinations [c]	0.30	–	–	–	–	100.0	100
Two-or-More Sub-Total [d]	1.75	6.3	5.9	7.2	52.9	27.7	100

[a] All percentages weighted to be nationally representative. Percentages taken from Parker and Makuc (2001, Table 2), who cited the National Health Interview Survey 1993-1995, APPENDIX: Percent Distribution (Standard Error) of Primary Racial Identification for Selected Detailed Race Groups.

[b] Primary racial affiliation based on survey results from Parker and Makuc (2001, Appendix).

[c] Parker and Makuc (2001) did not provide an estimate of the primary racial affiliation for "Other combinations, so we assume that it belongs to the "Other" category. Note that they did provide the primary racial affiliation for a fourth group "Black/AIAN:" 85.4% Black, 7.0% AIAN, and 7.6% Other. However, we do not have an estimate of the relative abundance of Black/AIAN in the general population, so we have dropped it from further consideration.

[d] As described in the text below, we calculated the percentages in this row from the percentages in the previous four rows for AIAN/White, API/White, Black/White, and Other combinations.

To estimate how to assign a single racial group for individuals that chose two or more racial groups, we used the results of Exhibit B-3 for the three main categories for which we an estimate of the primary racial affiliation: AIAN/White, API/White, and Black/White. To account for the 0.30 percent of the population in other combinations, we For each Census block, we assume that .89 / (.89+.30+.26+.30) = 50.8% of respondents in the Two or More category will fall into the AIAN / White category, and of these, 80.9% would primarily identify themselves as White if they were to choose a single racial category, 12.4% would primarily identify themselves as American Indian or Alaska Native, and 6.7% would primarily identify themselves as Other. Thus 0.508 * .809 = 41% of Two or More we will call White, 10% we identify as Native American, and 5% as Other.

We did not attempt to predict what respondents in the 'Other Combinations' category would have selected if they were to choose a single racial category, so we assume they are part of the "Other" category. To estimate the number of individuals in each of the five races, we performed the following calculations:

$$NativeAmerican = NativeAmericanAlone_{Pop} + TwoorMore_{Pop} \cdot \left(\frac{AIAN/White_{Nat\%}}{MultipleRace_{Nat\%}} \cdot AIAN\%_{AIAN/White} \right)$$

$$Asian = AsianAlone_{Pop} + PacificIslander/Hawaiian_{Pop} + TwoorMore_{Pop} \cdot \left(\frac{API/White_{Nat\%}}{MultipleRace_{Nat\%}} \cdot API\%_{API/White} \right)$$

$$Black = BlackAlone_{Pop} + TwoorMore_{Pop} \cdot \left(\frac{Black/White_{Nat\%}}{MultipleRace_{Nat\%}} \cdot Black\%_{Black/White} \right)$$

$$White = White\ Alone_{Pop} + Two\ \ or\ More_{Pop} \cdot \left(\begin{array}{l} \frac{AIAN/White_{Nat\%}}{MultipleRace_{Nat\%}} \cdot White\%_{AIAN/White} + \\ \frac{API/White_{Nat\%}}{MultipleRace_{Nat\%}} \cdot White\%_{API/White} + \\ \frac{Black/White_{Nat\%}}{MultipleRace_{Nat\%}} \cdot White\%_{Black/White} \end{array} \right)$$

$$Other = Other\ Alone_{Pop} + Two\ \ or\ More_{Pop} \cdot \left(\begin{array}{l} \frac{AIAN/White_{Nat\%}}{MultipleRace_{Nat\%}} \cdot Other\%_{AIAN/White} + \\ \frac{API/White_{Nat\%}}{MultipleRace_{Nat\%}} \cdot Other\%_{API/White} + \\ \frac{Black/White_{Nat\%}}{MultipleRace_{Nat\%}} \cdot Other\%_{Black/White} + \\ \frac{Other\ \ Combinations_{Nat\%}}{MultipleRace_{Nat\%}} \end{array} \right)$$

This then reduces to:

$$Native\ American_{Pop} = Native\ American\ Alone_{Pop} + (0.063)Two\ or\ More_{Pop}$$

$$Asian_{Pop} = Asian\ Alone_{Pop} + Pacific\ Islander\ /\ Hawaiian_{Pop} + (0.059)Two\ or\ More_{Pop}$$

$$Black_{Pop} = Black\ Alone_{Pop} + (0.072)Two\ or\ More_{Pop}$$

$$White_{Pop} = White\ Alone_{Pop} + (0.530)Two\ or\ More_{Pop}$$

$$Other_{Pop} = White\ Alone_{Pop} + (0.276)Two\ or\ More_{Pop} \quad .$$

B .3 Estimating Population Levels in Alternative Age Groups

In calculating the population in age groups that may include a portion of one of the pre-specified demographic groups in Exhibit B-1, BenMAP assumes the population is uniformly distributed in the age group. For example, to calculate the number of children ages 3 through 12, BenMAP calculates:

$$age_{3-12} = \frac{1}{2} \cdot age_{1-4} + age_{5-9} + \frac{3}{5} \cdot age_{10-14} \quad .$$

B .4 Estimating Population Levels in Non-Census Years

To estimate population levels in non-Census years, BenMAP uses two basic approaches. To estimate population between 1990 and 2000, BenMAP linearly interpolates between the two Census years of 1990 and 2000. To forecast population levels beyond 2000, BenMAP scales the 2000 Census value with the ratio of the county-level forecast for the future-year of interest and the county-level population in 2000. Woods & Poole (2001) provides the county-level population forecasts used to calculate the scaling ratios. Below we give examples with each approach.

B .4.1 Estimating Population Levels in 1991-1999

To estimate population levels between 1990 and 2000, BenMAP linearly interpolates between age groupings of interest.

$$population_i = \frac{10 - i}{10} \cdot population_{1990} + \frac{i}{10} \cdot population_{2000}$$

where i is an index running from 0 (for 1990) through 10 (for 2000).

In the case where there is a single age group, such as determining the number of people ages 20 through 24 in 1996, BenMAP calculates:

$$age_{20-24,1996} = \frac{4}{10} \cdot age_{20-24,1990} + \frac{6}{10} \cdot age_{20-24,2000}$$

The steps are essentially similar when more than one age goup is involved, such as determining the number of people ages 20-27 in 1996. BenMAP first calculates the number of people ages 20-27 in 1990 and 2000, and then interpolates between the 1990 and 2000 values:

$$age_{20-27,1990} = age_{20-24,1990} + \frac{3}{5} \cdot age_{25-29,1990}$$

$$age_{20-27,2000} = age_{20-24,2000} + \frac{3}{5} \cdot age_{25-29,2000}$$

$$age_{20-27,1996} = \frac{4}{10} \cdot age_{20-27,1990} + \frac{6}{10} \cdot age_{20-27,2000} \quad .$$

B .4.2 Forecasting Population Levels after 2000

To estimate population levels for the years after the last Census in 2000, BenMAP scales the 2000 Census-based estimate with the ratio of the county-level forecast for the future year of interest over the 2000 county-level population level. Woods & Poole (2001) provides the county-level population forecasts used to calculate the scaling ratios; these data are discussed in detail in section B.5.

In the simplest case, where one is forecasting a single population variable, say, children ages 4 to 9 in the year 2010, CAMPS calculates:

$$age_{4-9,g,2010} = age_{4-9,g,2000} \cdot \frac{age_{4-9,county,2010}}{age_{4-9,county,2000}}$$

where the g^{th} population grid-cell is wholly located within a given county.

In the case, where the g^{th} grid-cell includes "n" counties in its boundary, the situation is somewhat more complicated. BenMAP first estimates the fraction of individuals in a given age group (e.g., ages 4 to 9) that reside in the part of each county within the g^{th} grid-cell. BenMAP calculates this fraction by simply dividing the population all ages of a given county within the g^{th} grid-cell by the total population in the g^{th} grid-cell:

$$fraction\ of\ age_{4-9,g\ in\ county_c} = \frac{age_{all,g\ in\ county_c}}{age_{all,g}}$$

Multiplying this fraction with the number of individuals ages 4 to 9 in the year 2000 gives an estimate of the number of individuals ages 4 to 9 that reside in the fraction of the county within the g^{th} grid-cell in the year 2000:

$$age_{4-9,g\ in\ county_c,2000} = age_{4-9,g,2000} \cdot fraction\ age_{4-9,g\ in\ county_c}$$

To then forecast the population in 2010, we scale the 2000 estimate with the ratio of the county projection for 2010 to the county projection for 2000:

$$age_{4-9,\ g\ in\ county_c,\ 2010} = age_{4-9,\ g\ in\ county_c,\ 2000} \cdot \frac{age_{4-9,\ county_c,\ 2010}}{age_{4-9,\ county_c,\ 2000}} \quad .$$

Combining all these steps for "n" counties within the gth grid-cell, we forecast the population of persons ages 4 to 9 in the year 2010 as follows:

$$age_{4-9,\ g,\ 2010} = \sum_{c=1}^{n} age_{4-9,\ g,\ 2000} \cdot \frac{total\ pop_{g\ in\ county_c}}{total\ pop_g} \cdot \frac{age_{4-9,\ county_c,\ 2010}}{age_{4-9,\ county_c,\ 2000}}$$

In the case where there are multiple age groups and multiple counties, BenMAP first calculates the forecasted population level for individual age groups, and then combines the forecasted age groups. In calculating the number of children ages 4 to 12, BenMAP calculates:

$$age_{4-9,\ g,\ 2010} = \sum_{c=1}^{n} age_{4-9,\ g,\ 2000} \cdot \frac{total\ pop_{g\ in\ county_c}}{total\ pop_g} \cdot \frac{age_{4-9,\ county_c,\ 2010}}{age_{4-9,\ county_c,\ 2000}}$$

$$age_{10-14,\ g,\ 2010} = \sum_{c=1}^{n} age_{10-14,\ g,\ 2000} \cdot \frac{total\ pop_{g\ in\ county_c}}{total\ pop_g} \cdot \frac{age_{10-14,\ county_c,\ 2010}}{age_{10-14,\ county_c,\ 2000}}$$

$$age_{4-12,\ g,\ 2010} = age_{4-9,\ g,\ 2010} + \frac{3}{5} \cdot age_{10-14,\ g,\ 2010} \quad .$$

Since the Woods and Poole (2001) projections only extend through 2025, we used the existing projections and constant growth factors to provide additional projections. To estimate population levels beyond 2025, CAPMS linearly extrapolates from the final two years of data. For example, to forecast population in 2030, CAPMS calculates:

$$age_{4-9,\ 2030} = age_{4-9,\ 2025} + (2030 - 2025) \cdot \left(age_{4-9,\ 2025} - age_{4-9,\ 2024} \right) \quad .$$

Exhibit B-6 summarizes the forecasted age-stratified, state-level populations for 2020 and 2030. In addition, to provide a point of comparison, it includes population levels for year 2000.

Exhibit B-6. State-Level Population Estimates by Age Group

	2000			2020			2030		
State	***0-18***	***18-64***	***65+***	***0-18***	***18-64***	***65+***	***0-18***	***18-64***	***65+***
AL	1,126,337	2,740,965	579,798	1,212,702	3,057,540	909,065	1,284,828	3,069,518	1,195,921
AZ	1,371,099	3,091,693	667,839	1,832,737	4,351,254	1,309,943	2,117,749	4,708,843	1,812,966
AR	681,003	1,618,378	374,019	769,208	1,821,876	557,233	832,285	1,861,954	698,259
CA	9,254,212	21,021,768	3,595,658	10,105,474	25,308,061	5,717,329	10,895,067	26,084,324	7,681,464
CO	1,101,772	2,783,415	416,073	1,321,930	3,359,731	954,691	1,498,688	3,453,808	1,370,557
CT	839,051	2,096,330	470,183	821,773	2,110,504	588,222	845,210	1,991,695	728,973
DE	195,997	485,877	101,726	213,375	571,224	149,015	227,024	574,497	204,731
DC	120,659	381,502	69,898	95,389	337,146	103,401	93,833	305,565	123,728
FL	3,643,004	9,531,774	2,807,597	4,466,384	12,098,406	4,472,647	5,026,785	12,483,019	5,933,620
GA	2,176,259	5,224,918	785,275	2,600,100	6,296,967	1,328,722	2,851,139	6,620,751	1,778,194
ID	369,522	778,515	145,916	451,473	980,346	282,616	507,776	1,045,592	374,826
IL	3,247,904	7,671,362	1,500,025	3,286,653	8,148,579	2,069,429	3,425,612	7,962,757	2,669,430
IN	1,581,993	3,745,661	752,831	1,691,800	4,113,510	1,087,932	1,800,717	4,096,828	1,421,006
IA	737,415	1,752,696	436,213	734,433	1,820,333	593,034	766,374	1,750,358	755,945
KS	714,371	1,617,818	356,229	760,573	1,795,227	499,065	814,382	1,778,859	653,139
KY	998,042	2,538,933	504,793	1,077,101	2,762,379	801,696	1,154,120	2,750,564	1,052,988
LA	1,221,651	2,730,396	516,929	1,247,161	2,952,038	850,018	1,318,748	2,917,899	1,116,293
ME	299,691	791,830	183,402	284,880	852,466	289,399	297,507	807,626	394,873
MD	1,350,517	3,346,661	599,307	1,453,726	3,868,715	926,465	1,559,338	3,877,266	1,256,566
MA	1,508,818	3,980,116	860,162	1,533,618	4,071,543	1,144,857	1,604,543	3,871,104	1,469,089
MI	2,596,118	6,123,307	1,219,018	2,587,563	6,590,540	1,798,905	2,703,858	6,417,627	2,366,125
MN	1,285,100	3,040,113	594,266	1,413,120	3,525,458	948,035	1,547,597	3,524,435	1,298,319
MS	779,939	1,721,196	343,523	826,142	1,912,067	513,412	867,469	1,926,497	667,632
MO	1,428,853	3,410,978	755,379	1,526,846	3,830,433	1,062,471	1,631,969	3,798,554	1,404,065
MT	228,916	552,330	120,949	234,129	612,736	241,971	258,376	604,179	322,696
NE	450,372	1,028,696	232,195	483,340	1,148,129	329,112	522,703	1,142,368	426,556
NV	510,633	1,268,694	218,929	757,488	1,921,749	477,249	904,840	2,150,250	654,589

Appendix B. Population Data

	2000			2020			2030		
State	***0-18***	***18-64***	***65+***	***0-18***	***18-64***	***65+***	***0-18***	***18-64***	***65+***
NH	309,490	778,326	147,970	321,958	906,776	236,489	346,320	897,774	329,510
NJ	2,074,020	5,227,192	1,113,136	2,126,538	5,560,594	1,507,553	2,222,228	5,402,892	1,949,786
NM	506,558	1,100,262	212,225	583,389	1,372,898	412,394	644,935	1,420,580	572,907
NY	4,696,232	11,831,869	2,448,352	4,487,417	11,815,310	3,179,326	4,540,245	11,246,710	3,953,934
NC	1,977,387	5,102,877	969,048	2,399,345	6,081,807	1,588,246	2,646,039	6,337,401	2,079,430
ND	162,017	385,705	94,478	152,979	407,052	152,185	160,056	387,072	203,240
OH	2,889,207	6,956,175	1,507,757	2,894,902	7,316,549	2,031,922	3,021,197	7,084,923	2,597,112
OK	894,531	2,100,173	455,950	968,204	2,255,616	685,395	1,037,634	2,249,445	865,166
OR	846,233	2,136,988	438,177	1,028,841	2,593,792	824,166	1,161,142	2,731,400	1,065,084
PA	2,930,189	7,431,699	1,919,165	2,807,320	7,589,422	2,473,482	2,879,828	7,112,827	3,152,928
RI	252,438	643,479	152,402	248,650	664,840	185,270	253,697	626,436	236,588
SC	1,017,627	2,509,052	485,333	1,125,147	2,936,359	879,310	1,217,702	2,989,589	1,188,398
SD	202,496	444,217	108,131	209,379	498,258	159,468	222,092	487,168	215,460
TN	1,402,958	3,583,013	703,311	1,614,405	4,118,556	1,147,546	1,759,007	4,213,846	1,513,183
TX	5,891,741	12,887,542	2,072,532	7,108,830	15,994,222	3,802,007	7,929,363	16,840,990	5,236,651
UT	724,466	1,318,481	190,222	989,440	1,826,327	368,454	1,120,100	2,046,412	497,421
VT	147,949	383,368	77,510	137,590	414,505	138,315	144,053	397,442	190,941
VA	1,743,459	4,542,721	792,333	1,955,331	5,201,333	1,308,689	2,132,729	5,295,036	1,734,954
WA	1,511,831	3,720,140	662,148	1,785,937	4,644,371	1,174,213	2,006,978	4,857,761	1,581,410
WV	404,484	1,126,965	276,895	388,379	1,094,529	403,851	404,280	1,038,496	488,364
WI	1,369,215	3,291,907	702,553	1,413,693	3,680,062	1,070,942	1,511,982	3,612,218	1,450,806
WY	128,585	307,504	57,693	124,005	314,574	117,862	132,595	301,356	154,460

B .5 County Population Forecasts 2000-2025

Woods & Poole (2001) developed county-level forecasts for each year from 2000 through 2025, for three racial groups "Black," "White," and "Other," and by age and by gender. For the Hispanic ethnic group, Woods and Poole developed forecasts just for the total population, and not by age and gender. As discussed in the section on population forecasts, CAPMS uses these forecasts to simply scale the 2000 Census block data, in order to estimate the population in the population grid-cells for any given year after 2000.

B .5.1 Aligning Woods & Poole FIPS Codes with BenMAP FIPS Codes

The county geographic boundaries used by Woods & Poole are somewhat more aggregated than the county definitions used in the 2000 Census (and BenMAP), and the FIPS codes used by Woods and Poole are not always the standard codes used in the Census. To make the Woods and Poole data consistent with the county definitions in BenMAP, we disaggregated the Woods and Poole data and changed some of the FIPS codes. Exhibit B-7 lists the discrepancies in the county definitions between Woods & Poole and those used in BenMAP.

To assign the population in the more aggregated Woods & Poole county definitions to the more disaggregated definitions used in BenMAP (and the U.S. Census), we used the total county population from the 2000 U.S. Census. We then assumed that the age and racial groups were distributed uniformly across the BenMAP counties contained within a Woods & Poole county definition. For example, in estimating the population of children ages 4-9 in county "c" contained within a more broadly defined Woods & Poole county, we would do the following:

$$age_{4-9,\,county_c} = age_{4-9,\,W\&P\ county} \cdot \frac{age_{all,\,county_c}}{age_{all,\,W\&P\ county}}$$

After this factor was applied, we rounded the estimates to the nearest integer so as to avoid having data with "partial people."

Exhibit B-7. Linkage Between Woods & Poole County Definitions and BenMAP County Definitions

Woods and Poole Counties (FIPS)	Counties in BenMAP (FIPS)
Northwest Arctic Borough, AK (02188)	Kobuk, AK (02140)
Remainder of Alaska, AK (02999)	Aleutian Islands, AK (02010), Aleutian Islands East Borough, AK (02013), Aleutian Islands West Census Area, AK (02016), Bethel Census Area, AK (02050), Denali Borough, AK (02068), Dillingham Census Area, AK (02070), Haines Borough, AK (02100), Kenai Peninsula Borough, AK (02122), Lake and Peninsula Borough, AK (02164), North Slope Borough, AK (02185), Prince of Wales-Outer Ketchikan, AK (02201), Sitka Borough, AK (02220), Skagway-Yukatat-Angoon, AK (02231), Skagway-Hoonah-Angoon Census Area, AK (02232), Southeast Fairbanks Census Area, AK (02240), Valdez-Cordova Census Area, AK (02261), Wrangell-Petersburg Census Area, AK (02280), Yakutat Borough, AK (02282), Yukon-Koyukuk, AK (02290)
Yuma + La Paz, AZ (04027)	La Paz, AZ (04012), Yuma, AZ (04027)

Woods and Poole Counties (FIPS)	Counties in BenMAP (FIPS)
Miami-Dade, FL (12086)	Dade, FL (12025)
Maui + Kalawao, HI (15901)	Kalawao, HI (15005), Maui, HI (15009)
Fremont, ID (16043)	Fremont, ID (16043), Yellowstone Park, ID
Park, MT (30067)	Park, MT (30067), Yellowstone Park, MT (30113)
Valencia + Cibola, NM (35061)	Cibola, NM (35006), Valencia, NM (35061)
Halifax, VA (51083)	Halifax, VA (51083), South Boston City, VA (51780)
Albemarle + Charlottesville, VA (51901)	Albemarle, VA (51003), Charlottesville City, VA (51540)
Alleghany + Clifton Forge + Covington, VA (51903)	Alleghany, VA (51005), Clifton Forge City, VA (51560), Covington City, VA (51580)
Augusta + Staunton + Waynesboro, VA (51907)	Augusta, VA (51015), Staunton City, VA (51790), Waynesboro City, VA (51820)
Bedford + Bedford City, VA (51909)	Bedford, VA (51019), Bedford City, VA (51515)
Campbell + Lynchburg, VA (51911)	Campbell, VA (51031), Lynchburg City, VA (51680)
Carroll + Galax, VA (51913)	Carroll, VA (51035), Galax City, VA (51640)
Dinwiddie + Colonial Heights + Petersburg, VA (51918)	Dinwiddie, VA (51053), Colonial Heights City, VA (51570), Petersburg City, VA (51730)
Fairfax + Fairfax City + Falls Church City, VA (51919)	Fairfax, VA (51059), Fairfax City, VA (51600), Falls Church City, VA (51610)
Frederick + Winchester, VA (51921)	Frederick, VA (51069), Winchester City, VA (51840)
Greensville + Emporia, VA (51923)	Greensville, VA (51081), Emporia City, VA (51595)
Henry + Martinsville, VA (51929)	Henry, VA (51089), Martinsville City, VA (51690)
James City + Williamsburg, VA (51931)	James City County, VA (51095), Williamsburg City, VA (51830)
Montgomery + Radford, VA (51933)	Montgomery, VA (51121), Radford City, VA (51750)
Pittsylvania + Danville, VA (51939)	Pittsylvania, VA (51143), Danville City, VA (51590)
Prince George + Hopewell, VA (51941)	Prince George, VA (51149), Hopewell City, VA (51670)
Prince William + Manassas + Manassas Park, VA (51942)	Prince William, VA (51153), Manassas City, VA (51683), Manassas Park City, VA (51685)
Roanoke + Salem, VA (51944)	Roanoke, VA (51161), Salem City, VA (51775)
Rockbridge + Buena Vista + Lexington, VA (51945)	Rockbridge, VA (51163), Buena Vista City, VA (51530), Lexington City, VA (51678)
Rockingham + Harrisonburg, VA (51947)	Rockingham, VA (51165), Harrisonburg City, VA (51660)
Southampton + Franklin, VA (51949)	Southampton, VA (51175), Franklin City, VA (51620)
Spotsylvania + Fredericksburg, VA (51951)	Spotsylvania, VA (51177), Fredericksburg City, VA (51630)
Washington + Bristol, VA (51953)	Washington, VA (51191), Bristol City, VA (51520)
Wise + Norton, VA (51955)	Wise, VA (51195), Norton City, VA (51720)
York + Poquoson, VA (51958)	York, VA (51199), Poquoson City, VA (51735)
Shawano (includes Menominee), WI (55901)	Menominee, WI (55078), Shawano, WI (55115)

B .5.2 Age, Gender, Race, and Ethnicity

We generated the same 38 age and gender categories developed from the 1990 and 2000 Census data. Since these projections are available for every year of age, it is a simple matter to sum the individual years to get the same age categories used by BenMAP.

However, the only racial categories available are "White," "Black," and "Other." Since we do not have an Asian or Native American group, or an Other group which is consistent with the definition used by the 1990 and 2000 Census data, we assume that the projection data's Other category is representative of all 3 groups, and that they move together over time.

The county projections only forecast the Hispanic population of all ages, and does not have separate gender and age forecasts. Lacking further information, we use the ratio of future-year all age population to the year 2000 all age population when forecasting any particular age group of Hispanics. In effect, we assume for all forecast years the same distribution of age and gender as found in the 2000 Census.

B .5.3 Creating Growth Ratios from Absolute Population Values

For each year from 2000 through 2025 and for each of the 256 demographic groups listed in Exhibit B-1, BenMAP stores the ratio of the future-year to year 2000 county-level population projections. As described below, these ratios are used to forecast population levels in the population grid-cells used by BenMAP to health effects.

Note that there are a small number of cases were the 2000 county population for a specific demographic group is zero, so the ratio of any future year to the year 2000 data is undefined. In these relatively rare cases, we set the year 2000 ratio and all subsequent ratios to 1, assuming no growth.

There are an even smaller number of cases where a total population variable dwindles from some non-zero number to zero, creating ratios of zero. Variables which represent a subpopulation of the first variable may not be zero, however. In these cases, we set all subset population variables for that year to zero.

For instance, if a county only had one person in it for the year 2000 - a 79 year old black male - we set all variables (excluding total variables and BlackMale75to79) to a ratio of 1, because their 2000 values of 0 produce undefined ratios. If the man dies at age 82, the total black population variable for years 2003 and beyond is calculated as 0/1 = 0. Thus for each of those years where the total black population is listed as zero, we go back and set all black population variables to zero, to reflect the knowledge that the block is empty. For all variables except the BlackMale75to79 age group (already zero), 1 becomes 0.

Appendix C: Air Pollution Exposure Estimation Algorithms

BenMAP has grouped individuals into what we refer to as "population grid-cells," where the grid-cells typically conform to some type of grid used in an air quality model, such as the REMSAD air quality model, or just the counties of the United States. For each type of grid, the population is built in each grid-cell by aggregating census block data. In the next step, BenMAP estimates the air pollution exposure for each grid-cell, with the assumption that people living within a particular grid-cell experience the same air pollution levels.

You have a variety of approaches to estimate the exposure to air pollution for the people living within a given population grid-cell. Perhaps the simplest approach is to use model data directly, and to assume that the people living within a particular model grid-cell experience the level estimated by the model. An alternative approach is to use air pollution monitoring data, where you may choose the closest monitor data to the center of a grid-cell, take an average of nearby monitors, or use kriging. In a third general approach, you may combine both modeling and monitoring data to estimate exposure.

When combining modeling and monitoring data, BenMAP scales or adjusts the monitoring data with modeling data. The advantage of modeling data is that they can provide predictions for years in which monitoring data are not available, as well as to provide predictions in areas of the country for which monitoring data are not available. And the advantage of monitor data is that they are based on actual observations. Combining both sources of information, allows BenMAP to make more informed predictions.

The goal of estimating exposure is to provide the necessary input for concentration-response functions, so that BenMAP can estimate the impact of air pollution on adverse health effects. Exhibit C-1 lists the types of metrics commonly used in concentration-response functions. In the case of air pollution metrics calculated on a daily basis, such as the one-hour maximum and the 24-hour average, it is often the case that there are missing days of data. Air quality modeling is often conducted on a subset of the days in the year, and air quality monitors often miss a number of observations through out the year. To account for missing days, BenMAP represents the distribution of daily metrics with a certain number points or "bins," where each bin represents a certain range of the distribution, with the underlying assumption that missing days have the same distribution as the available data. For ozone, we use 153 bins to represent the ozone season from May through September, and for particulate matter we use 365 bins to represent the year. In addition to being able to account for incomplete or missing data, and using bins to represent the distribution provides a uniform approach that allows for easy comparison of different monitors.

Exhibit C-1. Metrics Typically Used in Concentration-Response Functions for Criteria Air Pollutants

Measurement Frequency	Metric Name	Metric Description
Daily (e.g., $PM_{2.5}$, PM_{10})	Daily Average	Daily average
	Annual Average	Average of four quarterly averages. The four quarters are defined as: Jan-Mar, April-June, Jul-Sep, Oct-Dec.
	Annual Median	Median of values through out the year.
Hourly (e.g., Ozone)	1-hour Daily Max	Highest hourly value from 12:00 A.M. through 11:59 P.M.
	5-hour Daily Average	Average of hourly values from 10:00 A.M. through 2:59 P.M.
	8-hour Daily Average	Average of hourly values from 9:00 A.M. through 4:59 P.M.
	12-hour Daily Average	Average of hourly values from 8:00 A.M. through 7:59 P.M.
	24-hour Daily Average	Average of hours from 12:00 A.M. through 11:59 P.M.
	SUM06	SUM06 index is the sum of the ozone concentrations (measured in ppm) that exceed 0.06 ppm between 8:00 am and 7:59 pm.

Note that the 8-hour daily average differs from the maximum 8-hour moving average described in the Federal Register (6 FR / Vol. 62, No. 138 / Friday, July 18, 1997 / Prepublication).

C .1 Direct Modeling

When using direct modeling data to estimate exposure, BenMAP assumes that the people living within a particular air pollution model grid-cell experience the same air pollution levels. BenMAP then estimates the air pollution metrics of interest. For pollutants measured hourly, such as ozone, these include the one-hour maximum and 24-hour average, and for pollutants measured daily, such as particulate matter, these include the annual mean and annual median.

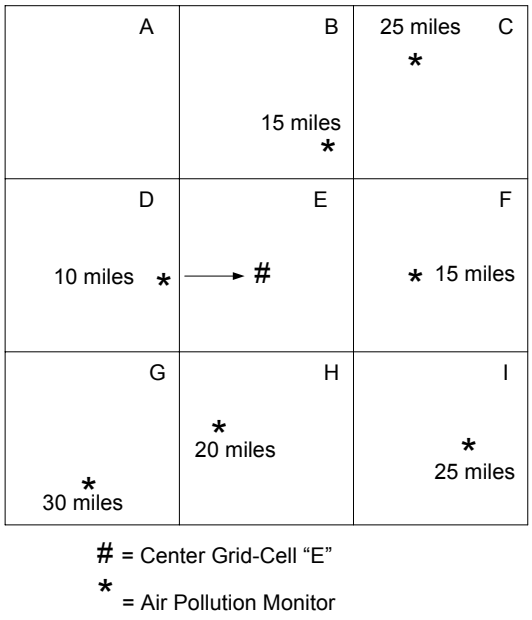
Generally modeling data providing hourly observations are complete for any given day. However, both hourly and daily often have missing days of observations. Given the estimated metrics, BenMAP then represents the distribution of daily metrics with 153 ozone bins and 365 particulate matter bins. By calculating bins with the available days, BenMAP assumes that the distribution of missing days is similar to the distribution of available monitoring.

C .2 Closest Monitor

When using the closet monitor to represent air pollution levels at a population grid-cell, BenMAP identifies the center of the population grid-cell, and then chooses the monitor that is closest to the center. In the simplest case, BenMAP assigns the closest monitor to a population grid-cell, uses the monitoring data to calculate the annual and daily air pollution metrics, and then calculates the bins that represent the distribution of the daily metrics. The annual metrics and the binned daily metrics are then used in the calculation of health effects.

The figure below presents nine population grid-cells and three monitors, with the focus on identifying the monitor closest to grid-cell "E." In this example, the closest monitor happens to be 10 miles

away from the center of grid-cell E, and the data from this monitor would be used to estimate air pollution levels for the population in this grid-cell. An analogous procedure would be used to estimate air pollution levels in the other grid-cells (A, B, C, D, F, G, H, and I).



To capture some of the information generated by air pollution models, such as REMSAD, UAM-V and others, BenMAP can also scale the data from the closest monitor with air pollution modeling data. BenMAP includes two types of scaling – "temporal" and "spatial" scaling. We discuss each below.

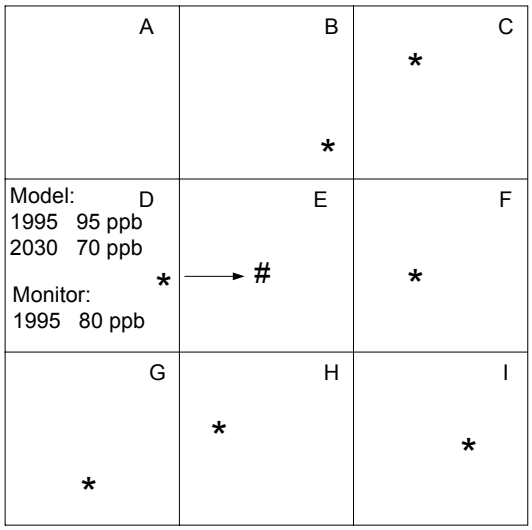
C .2.1 Closest Monitor – Temporal Scaling

With temporal scaling, BenMAP scales monitoring data with the ratio of the future-year to base-year modeling data, where the modeling data is from the modeling grid-cell containing the monitor. In the case of pollutants typically measured hourly, such as ozone, BenMAP scales the hourly monitor values, calculates the annual and daily metrics of interest, and then bins the daily metrics. In the case of pollutants typically measured daily, BenMAP scales the daily values, calculates the annual metrics of interest, and then bins the daily metric.

Consider the case in the figure below. To forecast air pollution levels for 2030, BenMAP would multiply the 1995 monitor value (80 ppb) by the ratio of the 2030 model value (70 ppb) to the 1995 model value (95 ppb):

$$\text{Forecast}_{2030} = \text{Monitor Value}_{1995} * (\text{Model Value}_{D,\,2030} / \text{Model Value}_{D,\,1995})$$

$$\text{Forecast}_{2030} = 80 \text{ ppb} * (70 \text{ ppb} / 95 \text{ ppb}) = 58.9 \text{ ppb}.$$



In this example, we have examined the adjustment of a single monitor value with the ratio of single model values. The approach is essentially the same when there are multiple monitor values and multiple model values. When there are multiple monitor values,

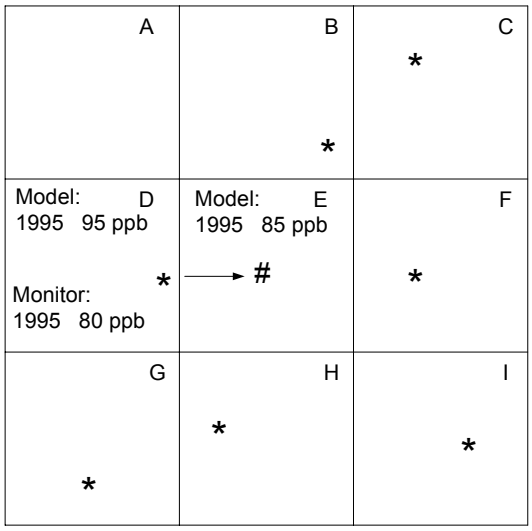
C .2.2 Closest Monitor – Spatial Scaling

With spatial scaling, we are estimating a monitor value for the center of each population grid-cell. We start by choosing the closest monitor to the center of each population grid-cell, and then we scale this closest monitor with modeling data. In particular, BenMAP multiplies the monitoring data with the ratio of the base-year modeling data for the destination grid-cell to the base-year modeling data for grid-cell containing the monitor. The spatial scaling occurs in the same fashion as with temporal scaling. In the case of pollutants typically measured hourly, such as ozone, BenMAP scales the hourly monitor values, calculates the annual and daily metrics of interest, and then bins the daily metrics. In the case of pollutants typically measured daily, BenMAP scales the daily values, calculates the annual metrics of interest, and then bins the daily metric.

To estimate air pollution levels for 1995 in grid-cell "E" below, BenMAP would multiply the 1995 closest monitor value (80 ppb) by the ratio of the 1995 model value for grid-cell "E" (70 ppb) to the 1995 model value for grid-cell "D" (95 ppb):

$$\text{Forecast}_{1995} = \text{Monitor Value}_{1995} * (\text{Model Value}_{E,\,1995} / \text{Model Value}_{D,\,1995})$$

$$\text{Forecast}_{1995} = 80 \text{ ppb} * (85 \text{ ppb} / 95 \text{ ppb}) = 71.6 \text{ ppb}.$$



C .2.3 Closest Monitor – Temporal and Spatial Scaling

Combining both temporal and spatial scaling, BenMAP first multiplies monitoring data with both the ratio of the future-year to base-year modeling data, where the modeling data is from the modeling grid-cell containing the monitor. This gives a temporary forecast for 2030. BenMAP then multiplies this temporary forecast with the ratio of the future-year modeling data for the destination grid-cell to the future-year modeling data for grid-cell containing the monitor. As seen below, this simplifies to multiplying monitoring data with both the ratio of future-year modeling data from the destination grid-cell to the base-year modeling data from the grid-cell containing the monitor. Again, as described for temporal and spatial scaling, BenMAP first scales the hourly and daily values, generates the metrics of interest and then bins the daily metrics.

To forecast air pollution levels for 2030 in the figure below, BenMAP would multiply the 1995 monitor value (80 ppb) by the ratio of the 2030 model value (70 ppb) to the 1995 model value (95 ppb):

$$\text{Temporary Forecast}_{2030} = \text{Monitor Value}_{1995} * (\text{Model Value}_{D, 2030} / \text{Model Value}_{D, 1995})$$

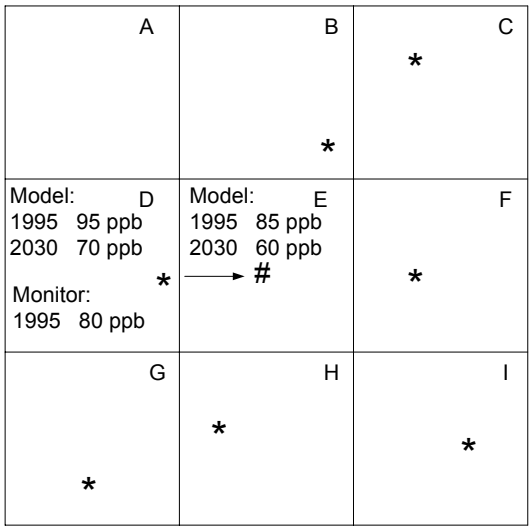
$$\text{Temporary Forecast}_{2030} = 80 \text{ ppb} * (70 \text{ ppb} / 95 \text{ ppb}) = 58.9 \text{ ppb}.$$

$$\text{Forecast}_{2030} = \text{Temporary Forecast}_{2030} * (\text{Model Value}_{E, 2030} / \text{Model Value}_{D, 2030})$$

$$\text{Forecast}_{2030} = 58.9 \text{ ppb} * (60 \text{ ppb} / 70 \text{ ppb}) = 50.5 \text{ ppb}.$$

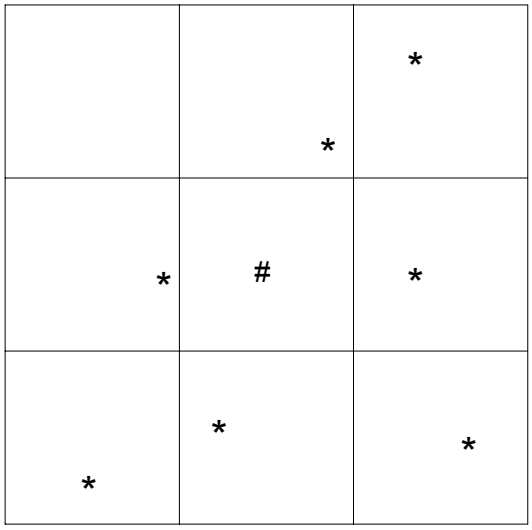
Note that through cancellation, this equation simplifies to:

$$\text{Forecast}_{2030} = \text{Monitor Value}_{1995} * (\text{Model Value}_{E, 2030} / \text{Model Value}_{D, 1995})$$

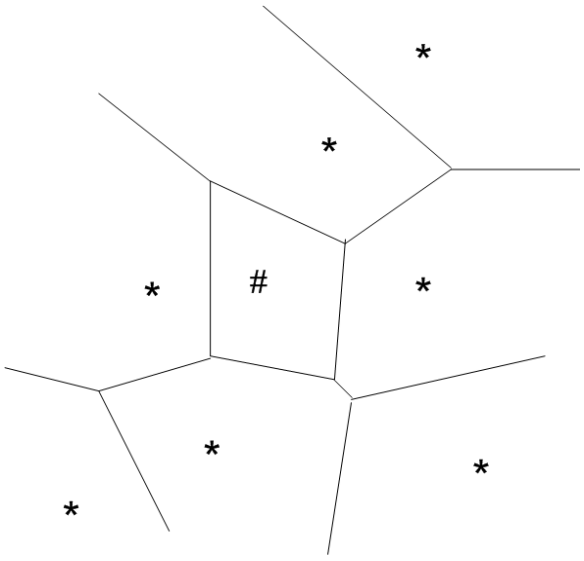


C .3 Voronoi Neighbor Averaging (VNA)

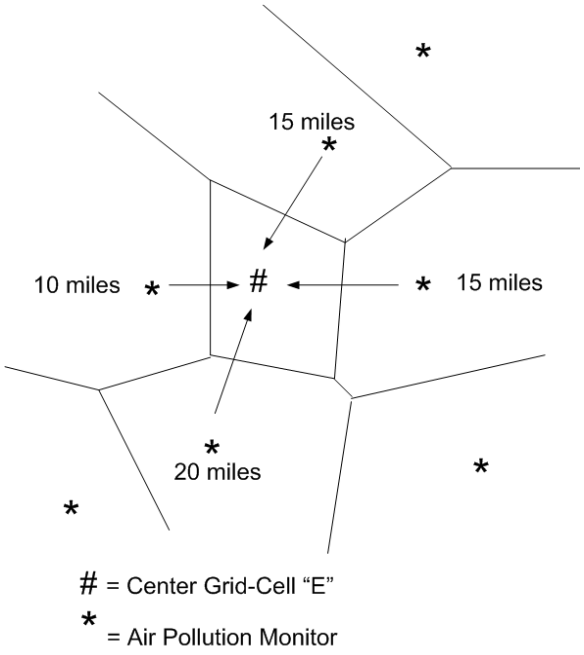
Like the closest monitor approach, the Voronoi Neighbor Averaging (VNA) algorithm uses monitor data directly or in combination with modeling data. However, instead of using the single closest monitor to estimate exposure at a population grid-cell, the VNA algorithm interpolates air quality at every population grid cell by first identifying the set of monitors that best "surround" the center of the population grid-cell.



In particular, BenMAP identifies the nearest monitors, or "neighbors," by drawing a polygon, or "Voronoi" cell, around the center of each BenMAP grid cell. The polygons have the special property that the boundaries are the same distance from the two closest points.



We then choose those monitors that share a boundary with the center of grid-cell "E." These are the nearest neighbors, we use these monitors to estimate the air pollution level for this grid-cell.



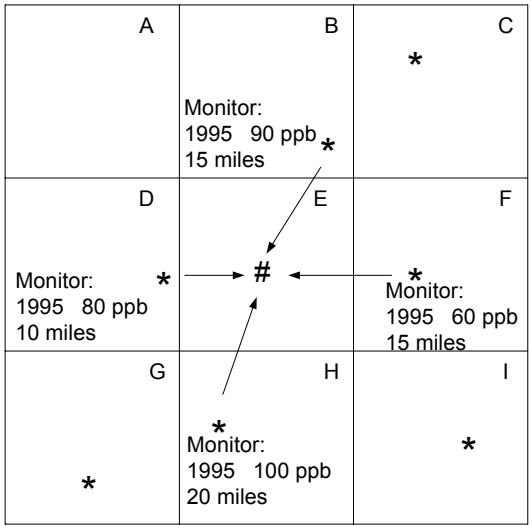
To estimate the air pollution level in each grid-cell, BenMAP calculates the annual and the binned daily metrics for each of the neighboring monitors, and then calculates an inverse-distance weighted average of the metrics. The further the monitor is from the BenMAP grid-cell, the smaller the weight.

In the figure below, the weight for the monitor 10 miles from the center of grid-cell E is calculated as follows:

$$weight_{10\,miles} = \frac{\frac{1}{10}}{\left(\frac{1}{10}+\frac{1}{15}+\frac{1}{15}+\frac{1}{20}\right)} = 0.35 \ .$$

The weights for the other monitors would be calculated in a similar fashion. BenMAP would then calculate an inverse-distance weighted average for 1995 air pollution levels in grid-cell E as follows:

$$\text{Forecast}_{1995} = 0.35*80 \text{ ppb} + 0.24*90 \text{ ppb} + 0.24*60 \text{ ppb} + 0.18*100 \text{ ppb} = 81.2 \text{ ppb} \ .$$



Note that BenMAP is calculating an inverse-distance weighted average of the annual metrics and the binned daily metrics. Alternatively, BenMAP could calculate an inverse-distance weighted average of the hourly and daily observations, calculated the annual and daily metrics, and then binned the daily metrics.

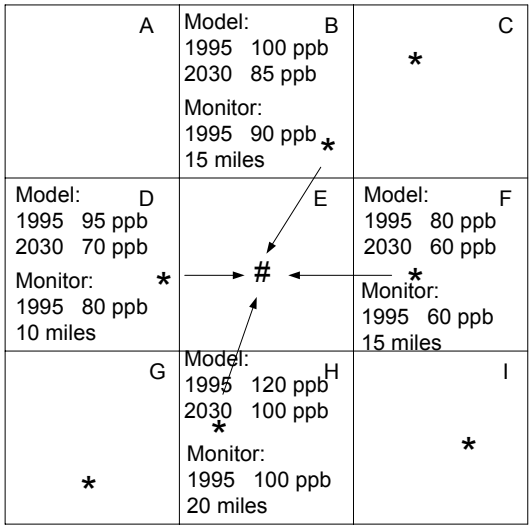
C .3.1 Voronoi Neighbor Averaging (VNA) – Temporal Scaling

As with forecasting air pollution levels by temporally scaling the closest monitor, BenMAP can combine VNA with temporal scaling. BenMAP temporally scales all of the neighboring monitors, calculates the metrics of interest, and then calculates an inverse distance-weighted average of the metrics.

Consider the example in the figure below. To forecast air pollution levels for 2030, BenMAP would multiply the 1995 monitor value by the ratio of the 2030 model value to the 1995 model value:

$$Forecast_{2030} = \sum_{i=1}^{4} Weight_i * Monitor_i * \frac{Model_{i,2030}}{Model_{i,1995}}$$

$$Forecast_{2030} = \left(0.35 * 80 * \frac{70}{95}\right) + \left(0.24 * 90 * \frac{100}{85}\right) + \left(0.24 * 60 * \frac{80}{60}\right) + \left(0.18 * 100 * \frac{120}{100}\right) = 64.1 ppb$$



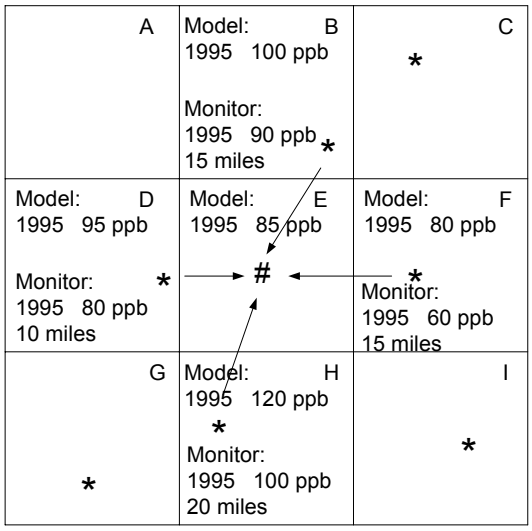
C .3.2 Voronoi Neighbor Averaging (VNA) – Spatial Scaling

BenMAP can also combine VNA with spatial scaling. For each of the neighbor monitors, BenMAP multiplies the monitoring data with the ratio of the base-year modeling data for the destination grid-cell to the base-year modeling data for grid-cell containing the monitor. The spatial scaling occurs in the same fashion as with temporal scaling. In the case of pollutants typically measured hourly, such as ozone, BenMAP scales the hourly monitor values, calculates the annual and daily metrics of interest, and then bins the daily metrics. In the case of pollutants typically measured daily, BenMAP scales the daily values, calculates the annual metrics of interest, and then bins the daily metric.

Consider the example in the figure below. To forecast air pollution levels for 1995, BenMAP would multiply the 1995 monitor value by the ratio of the 1995 model value to the 1995 model value:

$$Forecast_{1995} = \sum_{i=1}^{4} Weight_i * Monitor_i * \frac{Model_{E,1995}}{Model_{i,1995}}$$

$$Forecast_{1995} = \left(0.35 * 80 * \frac{85}{95}\right) + \left(0.24 * 90 * \frac{85}{100}\right) + \left(0.24 * 60 * \frac{85}{80}\right) + \left(0.18 * 100 * \frac{85}{120}\right) = 70.8\,ppb$$



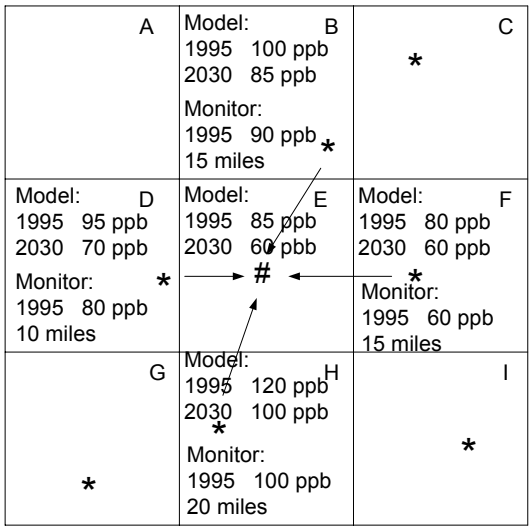
C .3.3 Voronoi Neighbor Averaging (VNA) – Temporal & Spatial Scaling

Combining both temporal and spatial scaling, BenMAP multiplies monitoring data with the ratio of the future-year to base-year modeling data, where the future-year modeling data are from the destination grid-cell and the base-year modeling data are from the grid-cell containing the monitor. One the hourly and daily monitoring data are scaled, BenMAP generates the metrics of interest, bins the daily metrics, and then uses the metrics to estimate adverse health effects in the population grid-cell.

The figure below gives an example of combining temporal and spatial scaling.

$$Forecast_{2030} = \sum_{i=1}^{4} Weight_i * Monitor_i * \frac{Model_{E,2030}}{Model_{i,1995}}$$

$$Forecast_{2030} = \left(0.35 * 80 * \frac{60}{95}\right) + \left(0.24 * 90 * \frac{60}{100}\right) + \left(0.24 * 60 * \frac{60}{80}\right) + \left(0.18 * 100 * \frac{60}{120}\right) = 50.0$$



C .4 Kriging

Following the same approach as the closest monitor approach and VNA, the kriging algorithm can use monitor data directly or in combination with modeling data. BenMAP includes two types: Ordinary Kriging and Block Kriging. Below, we discuss each type, and then briefly discuss their use in BenMAP.

C .4.1 Ordinary Kriging

Ordinary Kriging estimates an unknown value using a linear combination of available sample data. It is a "best linear unbiased estimator". Unbiased, because it tries to have the residual mean equal to 0, best because it minimizes the variance σ_R^2 of the errors.

The estimate $\hat{v}$ can be expressed as linear combination of true values.

$$\hat{v} = \sum_{j=1}^{n} w_j \cdot v$$

Defining the error r as the difference between the estimated value and the true value allows us to write the error of the i-th estimate as

$$r_i = \hat{v}_i - v_i$$

.

The average error or residual mean can therefore be written as

$$m_r = \frac{1}{k}\sum_{i=1}^{k} r_i = \frac{1}{k}\sum_{i=1}^{k} \hat{v}_i - v_i$$

Since the estimate is unbiased and a stationary random model function is assumed, the following holds true[3]:

$$\sum_{i=1}^{n} w_i = 1$$

At this point we can write the error variance σ_R^2 of a set of k estimates as:

$$\sigma_R^2 = \frac{1}{k}\sum_{i=1}^{k} (r_i - m_R)^2$$

.

Assuming a mean error of 0 simplifies this equation to:

$$\sigma_R^2 = \frac{1}{k}\sum_{i=1}^{k} [\hat{v}_i - v_i]^2$$

.

Unfortunately, the true values v_i are unknown. However, applying a stationary random function model allows us to rewrite (6) as

$$\sigma_R^2 = \tilde{\sigma}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \tilde{C}_{ij} - 2\sum_{i-1}^{n} w_i \tilde{C}_{i0}$$

with $\tilde{\sigma}^2$ being the variance and $\tilde{C}_{ij}$ being the covariances.

In order to solve this equation an additional term is introduced into the equation.

$$\sigma_R^2 = \tilde{\sigma}^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \tilde{C}_{ij} - 2\sum_{i=1}^{n} w_i \tilde{C}_{i0} + 2\mu(\sum_{i=1}^{n} w_i - 1)$$

[3]Please review additional literature for a detailed explanation.

Due to the unbiasedness assumption the last term in (8) equals 0. Furthermore, setting the first partial derivative with respect to μ to 0 reproduces the unbiasedness condition.

Since we assumed that the error variance will be minimal, the partial derivatives with respect to w_i will be calculated and set to 0.

This approach leads us to:

$$\sum_{j=1}^{n} w_j \tilde{C}_{ij} + \mu = \tilde{C}_{i0} \qquad \forall\, i = 1,\ldots,n$$

and

$$\sum_{i=1}^{n} w_i = 1$$

This system of equations can be written in matrix notation (referred to as the Ordinary Kriging System)

$$C \cdot w = D$$

or

$$\begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix} = \begin{bmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \\ 1 \end{bmatrix}$$

In order to solve this equation for the weights, we multiply this equation with the inverse of the covariance matrix C (assuming that C is positive definite):

$$w = C^{-1} \cdot D$$

C .4.2 Block Kriging

Block Kriging is a simple extension to the Ordinary Kriging system. It allows the computation of a mean value over a local area. Since any linear combination of random variables is also a random variable, the mean value over an area can be described as follows:

$$V_A = \frac{1}{|A|} \sum_{j \in A} V_j$$

Introducing a point to block covariance:

$$\overline{\tilde{C}}_{iA} = \frac{1}{|A|}\sum_{j \in A} \tilde{C}_{ij}$$

we can rewrite the ordinary kriging equation:

$$\begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix} = \begin{bmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \\ 1 \end{bmatrix}$$

as follows:

$$\begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix} = \begin{bmatrix} \overline{\tilde{C}}_{1A} \\ \vdots \\ \overline{\tilde{C}}_{nA} \\ 1 \end{bmatrix}$$

C .4.3 Kriging in BenMAP

BenMAP allows you to enter the empirically determined values for the "nugget" and the covariance function that is to be used during the interpolation process. At the current stage these need to be computed using external programs. After entering these values the program requires you to hit the validate button before allowing you to commit the newly entered values. This ensures that the entered covariance function can be interpreted at runtime.

Block Kriging is similar to Ordinary Kriging and introduces only one new set of parameters. In the Block Kriging options you can enter the matrix parameters for the support points that are to be overlaid on top of each grid cell. These support points are used to compute the point to block covariances introduced in

$$\overline{\tilde{C}}_{iA} = \frac{1}{|A|}\sum_{j \in A} \tilde{C}_{ij}$$

Note for shapefile grids that the support points will be computed by superimposing the bounding box of the shape element with the matrix of supporting cells. Supporting points that fall outside the shape element are deleted. Should all supporting points be eliminated, the system will default to the shape's centerpoint for the covariance calculation.

BenMAP also allows the user to save the current system settings (saved user default settings will be restored the next time BenMap is started). Note that all changes to the Kriging settings will persist for the entire work session with BenMap. Should it become necessary, the user can restore the original system settings using the **Restore Original Default** button.

Finally, as noted above, you may use kriging with modeling to scale the monitor data, in the same that you can do scaling with the closest monitor and VNA interpolation options.

C .5 Temporal and Spatial Scaling Adjustment Factors

As presented in the preceding examples of temporal and spatial scaling, both closest monitor and VNA use model data to scale monitor observations. In the examples, we scaled single monitor values with the ratio of single model values. In fact, however, the scaling involves multiple monitor values and multiple model values.

To proceed with the scaling, BenMAP takes the modeling values and splits them into groups. For ozone, we use 10 adjustment factors for the ozone season, where the first group represents the first 10 percent of the model observations; the second group represents the observations between the 10th and 20th percentile; and so on through the tenth group, which represents the observations between the 90th and 100th percentiles. BenMAP then averages the values in each group. For particulate matter, there are five adjustment factors for each of the four seasons in the year, where the first group in each season represents the first 20 percent of the model observations; the second group represents the observations between the 20th and 40th percentiles; and so on. Then, as for ozone model values, BenMAP averages the particulate matter model values in each group.

BenMAP treats the monitor values in a similar way. It sorts the monitor values from low to high, and divides them into the same number groups as there are scaling factors. Exhibit C-2 summarizes some of the types of analyses that have been conducted recently using scaling factors.

Exhibit C-2. Types of Analyses Using Scaling Factors

Pollutant	Daily Time Period	Season	# Scaling Factors Each Season
Ozone	Daylight hours: 9:00 am – 8:59 pm	May 1 – September 30	10
Particulate Matter	–	Four seasons: January-March, April-June, July-September, October-December	5

C .5.1 Calculation of Scaling Factors

In developing scaling factors, BenMAP sorts the modeling data into either 10 groups or 20 groups, depending on the pollutant (10 for ozone and for particulate five for each of the four seasons). Given the number of groups, then BenMAP determines how to assign the model values. In determining to which group a value belongs, BenMAP assigns a two-digit "percentile" to each value. With values in a given grid-cell sorted from low to high, the percentile for each value will equal: (the observation rank number minus 0.5) divided by (the total number of values) multiplied by (100). If there are 250 hourly values, the

first hourly value will have a percentile = (1-0.5)/(250)*(100) = 0.20%; the 27th value will have a percentile = (27-0.5)/(250)*(100) = 10.60%; and so on.

Each data group is represented by "group-lo" and "group-hi" values. These are the minimum and the maximum percentiles in each group, where group-lo equals: (group rank minus 1) multiplied by (100) divided by (the number of groups); and group-hi equals: (group rank) multiplied by (100) divided by (the number of groups) minus 0.001. If there are ten groups: the first group will have: group-lo = (1-1)/100*10 = 0.000%, and group-hi = (1/100*10)-0.001 = 9.999% ; the second group will have: group-lo = (2-1)/100*10 = 10.000%, and group-hi = (2/100*10)-0.001 = 19.999% ; and so on to the tenth group, which will have: group-lo = (10-1)/100*10 = 90.000%, and group-hi = (10/100*10)-0.001 = 99.999%. BenMAP assigns each observation to a particular group with the following algorithm: if "group-lo" <"percentile" < "group-hi", then assign the observation to that data group.

C .5.2 How BenMAP Scales PM and Ozone Monitor Data

Below we give the equations that BenMAP uses when scaling particulate matter and ozone monitor values.

Scaling Particulate Matter Monitor Values

After preparing the model and monitor data, BenMAP calculates the following:

$$adjusted\ monitor_{i,j,future} = monitor_{i,j,base} \cdot \frac{REMSAD_{j,k,future}}{REMSAD_{j,l,base}}$$

where:

adjusted monitor	= predicted daily $PM_{2.5}$ level, after adjustment by model data ($\mu g/m^3$)
monitor	= observed daily $PM_{2.5}$ monitor level ($\mu g/m^3$)
i	= day identifier
j	= model season/quintile group (1 to 20)
k	= grid cell identifier for population grid cell
l	= grid cell identifier for grid cell containing monitor
base	= base-year (e.g., 2000)
future	= future-year (e.g., 2020)
REMSAD	= representative model season/quintile value ($\mu g/m^3$)

After adjusting the monitor values to reflect air quality modeling, BenMAP calculates for each monitor the $PM_{2.5}$ metrics needed to estimate adverse health effects. In the case of Voronoi Neighbor Averaging, BenMAP then calculates an inverse-distance weighted average of the neighbors identified for each population grid cell:

$$population\ grid\ cell_{future} = \sum_{m=1}^{n} adjusted\ monitor_{m,future} \cdot weight_m \cdot$$

where:

population grid cell	= inverse distance-weighted $PM_{2.5}$ metric at population grid cell ($\mu g/m^3$)
adjusted monitor	= predicted $PM_{2.5}$ metric, after adjustment by model data ($\mu g/m^3$)
m	= monitor identifier
future	= future-year (e.g., 2020)
weight	= inverse-distance weight for monitor

After generating the bins for both the baseline and control scenarios, BenMAP can use these to calculate the change in air quality needed in most C-R functions to calculate the change in adverse health effects. To calculate the change in air quality, BenMAP subtracts the baseline value in the first bin from the control value in the first bin, and so on for each of the 365 bins created for the daily $PM_{2.5}$ average.

Scaling Ozone Monitor Values

After preparing the model and monitor data, BenMAP calculates the following:

$$adjusted\ monitor_{i,j,future} = monitor_{i,j,base} \cdot \frac{CAMX_{j,k,future}}{CAMX_{j,l,base}}$$

where:

adjusted monitor	= predicted hourly ozone level, after adjustment by model data (ppb)
monitor	= observed hourly ozone monitor level (ppb)
i	= hour identifier
j	= model decile group (1 to 10)
k	= grid cell identifier for population grid cell
l	= grid cell identifier for grid cell containing monitor
base	= base-year (e.g., 1996)
future	= future-year (e.g., 2030)
CAMX	= representative model decile value (ppb)

After adjusting the monitor values to reflect air quality modeling, BenMAP calculates for each monitor the ozone metrics needed to estimate adverse health effects. In the case of Voronoi Neighbor Averaging, BenMAP then calculates an inverse-distance weighted average of the neighbors identified for each population grid cell:

$$population\ grid\ cell_{future} = \sum_{m=1}^{n} adjusted\ monitor_{m,future} \cdot weight_m \cdot$$

where:

population grid cell	= inverse distance-weighted ozone metric at population grid cell (ppb)
adjusted monitor	= predicted ozone metric, after adjustment by model data (ppb)
m	= monitor identifier
future	= future-year (2020, 2030)
weight	= inverse-distance weight for monitor

After generating the bins for both the baseline and control scenarios, BenMAP can use these to calculate the change in air quality needed in most C-R functions to calculate the change in adverse health effects. To calculate the change in air quality, BenMAP subtracts the baseline value in the first bin from the control value in the first bin, and so on for each of the 153 bins created for the daily ozone metrics.

C .6 Binned Metrics

When estimating air pollution exposure, BenMAP calculates both daily metrics, such as the 24-hour daily average, and annual metrics, such as the annual mean. Because daily metrics are often not available for the entire year, BenMAP calculates representative values or bins with the available daily metrics, under the assumption that the missing days have a similar distribution. Each bin represents a day. In the case where there are 365 bins, the set of bins represents the entire year.

When combining air pollution metrics from multiple monitors, BenMAP first calculates the bins for the daily metrics, and then combines the bins, such as with some form of VNA. Once BenMAP has calculated binned exposure measures for both a baseline and a control scenario, BenMAP then takes the difference between the two scenarios for each bin – taking the difference between the baseline value in the first bin and the control value in the first bin, and so on for each of the bins.

Appendix D: Types of Concentration-Response Functions & Issues in the Estimation of Adverse Health Effects

This Appendix provides of an overview regarding the concentration-response (C-R) functions that BenMAP uses to estimate the impact of a change in air pollution on adverse health effects. It provides a description of the particular types of C-R functions that BenMAP uses. And then summarizes the approach used to choose the C-R functions included in BenMAP, and presents some issues associated with the use of C-R functions.

D .1 Overview

The relationship between the concentration of a pollutant, x, and the population response, y, is called the concentration-response (C-R) function. For example, the concentration of the pollutant may be fine particulate matter ($PM_{2.5}$) in $\mu g/m^3$ per day, and the population response may be the number of premature deaths per 100,000 population per day. C-R functions are estimated in epidemiological studies. A functional form is chosen by the researcher, and the parameters of the function are estimated using data on the pollutant (e.g., daily levels of $PM_{2.5}$) and the health response (e.g., daily mortality counts). There are several different functional forms, discussed below, that have been used for C-R functions. The one most commonly used is the log-linear form, in which the natural logarithm of the health response is a linear function of the pollutant concentration.

For the purposes of estimating benefits, we are not interested in the C-R function itself, however, but the relationship between the *change* in concentration of the pollutant, Δx, and the corresponding change in the population health response, Δy. We want to know, for example, if the concentration of $PM_{2.5}$ is reduced by 10 $\mu g/m^3$, how many premature deaths will be avoided? The relationship between Δx and Δy can be derived from the C-R function, as described below.

Many epidemiological studies, however, do not report the C-R function, but instead report some measure of the change in the population health response associated with a *specific* change in the pollutant concentration. The most common measure reported is the relative risk associated with a given change in the pollutant concentration. A general relationship between Δx and Δy can, however, be derived from the relative risk. The relative risk and similar measures reported in epidemiological studies are discussed in the sections below. The derivation of the relationship of interest for BenMAP – the relationship between Δx and Δy – is discussed in the subsequent sections.

D .1.1 Review Relative Risk and Odds Ratio

The terms relative risk and odds ratio are related but distinct. Exhibit D-1 provides the basis for demonstrating their relationship.

Exhibit D-1. Relative Risk and Odds Ratio Notation

Exposure	Fraction of Population		Adverse Effect Measure	
	Affected	*Not Affected*	*Relative Risk*	*Odds*
Baseline Pollutant Exposure	y_0	$1-y_0$	y_0/y_c	$y_0/(1-y_c)$
Control Pollutant Exposure	y_c	$1-y_c$		$y_c/(1-y_c)$

The "risk" that people with baseline pollutant exposure will be adversely affected (e.g., develop chronic bronchitis) is equal to y_0, while people with control pollutant exposure face a risk, y_c, of being adversely affected. The relative risk (RR) is simply:

$$RR = \frac{y_0}{y_c}.$$

The odds that an individual facing high exposure will be adversely affected is:

$$Odds = \frac{y_0}{1-y_0}.$$

The odds ratio is then:

$$Odds\ Ratio = \frac{\left(\frac{y_0}{1-y_0}\right)}{\left(\frac{y_c}{1-y_c}\right)}.$$

This can be rearranged as follows:

$$Odds\ Ratio = \frac{y_0}{y_c}\cdot\left(\frac{1-y_c}{1-y_0}\right) = RR\cdot\left(\frac{1-y_c}{1-y_0}\right).$$

As the risk associated with the specified change in pollutant exposure gets small (i.e., both y_0 and y_c approach zero), the ratio of $(1-y_c)$ to $(1-y_0)$ approaches one, and the odds ratio approaches the relative risk. This relationship can be used to calculate the pollutant coefficient in the C-R function from which the reported odds ratio or relative risk is derived, as described below.

D .2 The estimation of health effect incidence change

The functional form of the relationship between the change in pollutant concentration, Δx, and the change in population health response (usually an incidence rate), Δy depends on the functional form of the C-R function from which it is derived, and this depends on the underlying relationship assumed in the epidemiological study chosen to estimate a given effect. For expository simplicity, the following subsections refer simply to a generic adverse health effect, "y" and uses particulate matter (PM) as the pollutant – that is, Δx = ΔPM – to illustrate how the relationship between Δx and Δy is derived from each of several different C-R functions.

Estimating the relationship between ΔPM and Δy can be thought of as consisting of three steps:

(1) choosing a functional form of the relationship between PM and y (the C-R function),
(2) estimating the values of the parameters in the C-R function assumed, and
(3) deriving the relationship between ΔPM and Δy from the relationship between PM and y (the C-R function).

Epidemiological studies have used a variety of functional forms for C-R functions. Some studies have assumed that the relationship between adverse health and pollution is best described by a linear form, where the relationship between y and PM is estimated by a linear regression in which y is the dependent variable and PM is one of several independent variables. Log-linear regression[4] and logistic regression are other common forms.

D .2.1 Linear Model

A linear relationship between the rate of adverse health effects (incidence rate) and various explanatory variables is of the form:

$$y = \alpha + \beta \cdot PM$$

where α incorporates all the other independent variables in the regression (evaluated at their mean values, for example) times their respective coefficients. The relationship between the change in the rate of the adverse health effect from the baseline rate (y_0) to the rate after control (y_c) associated with a change from PM_0 to PM_c is then:

$$\Delta y = y_c - y_0 = \beta(PM_c - PM_0) = \beta \Delta PM \ .$$

[4]The log-linear form used in the epidemiological literature on ozone- and PM-related health effects is often referred to as "Poisson regression" because the underlying dependent variable is a count (e.g., number of deaths), believed to be Poisson distributed. The model may be estimated by regression techniques but is often estimated by maximum likelihood techniques. The form of the model, however, is still log-linear.

For example, Ostro et al. (1991, Table 5) reported a $PM_{2.5}$ coefficient of 0.0006 (with a standard error of 0.0003) for a linear relationship between asthma and $PM_{2.5}$ exposure.[5]

The lower and upper bound estimates for the $PM_{2.5}$ coefficient are calculated as follows:

$$\beta_{lower\ bound} = \beta - (1.96 \cdot \sigma_\beta) = 0.0006 - (1.96 \cdot 0.0003) = 1.2 \cdot 10^{-5}$$
$$\beta_{upper\ bound} = \beta + (1.96 \cdot \sigma_\beta) = 0.0006 + (1.96 \cdot 0.0003) = 0.00119$$

It is then straightforward to calculate lower and upper bound estimates of the change in asthma.

D .2.2 Log-linear Model

The log-linear relationship defines the incidence rate (y) as:

$$y = B \cdot e^{\beta \cdot PM}$$

or, equivalently,

$$\ln(y) = \alpha + \beta \cdot PM\ ,$$

where the parameter B is the incidence rate of y when the concentration of PM is zero, the parameter β is the coefficient of PM, ln(y) is the natural logarithm of y, and $\alpha = \ln(B)$.[6]

The relationship between ΔPM and Δy is:

$$\Delta y = y_c - y_0 = Be^{\beta \cdot PM_c} - Be^{\beta \cdot PM_0}\ .$$

This may be rewritten as:

$$\Delta y = e^{\beta \cdot PM_0}\left(e^{\beta \cdot (PM_c - PM_0)} - 1\right) = y_0 \cdot \left(e^{\beta \cdot \Delta PM} - 1\right)\ ,$$

where y_0 is the baseline incidence rate of the health effect (i.e., the incidence rate before the change in PM).

[5]Ostro et al. (1991) happen to use the natural logarithm of $PM_{2.5}$.

[6] Other covariates besides pollution clearly affect mortality. The parameter B might be thought of as containing these other covariates, for example, evaluated at their means. That is, $B = B_o \exp\{\beta_1 x_1 + ... + \beta_n x_n\}$, where B_o is the incidence of y when all covariates in the model are zero, and $x_1, ..., x_n$ are the other covariates evaluated at their mean values. The parameter B drops out of the model, however, when changes in y are calculated, and is therefore not important.

The change in the incidence of adverse health effects can then be calculated by multiplying the change in the incidence rate, Δy, by the relevant population (e.g., if the rate is number per 100,000 population, then the relevant population is the number of 100,000s in the population).

When the PM coefficient (β) and its standard error (σ_β) are published (e.g., Ostro et al., 1989), then the coefficient estimates associated with the lower and upper bound may be calculated easily as follows:

$$\beta_{lower\ bound} = \beta - (1.96 \cdot \sigma_\beta)$$

$$\beta_{upper\ bound} = \beta + (1.96 \cdot \sigma_\beta) .$$

Epidemiological studies often report a relative risk for a given ΔPM, rather than the coefficient, β (e.g., Schwartz et al., 1995, Table 4). Recall that the relative risk (RR) is simply the ratio of two risks:

$$RR = \frac{y_0}{y_c} = e^{\beta \cdot \Delta PM} .$$

Taking the natural log of both sides, the coefficient in the C-R function underlying the relative risk can be derived as:

$$\beta = \frac{\ln(RR)}{\Delta PM} .$$

The coefficients associated with the lower and upper bounds (e.g., the 2.5 and 97.5 percentiles) can be calculated by using a published confidence interval for relative risk, and then calculating the associated coefficients.

Because of rounding of the published RR and its confidence interval, the standard error for the coefficient implied by the lower bound of the RR will not exactly equal that implied by the upper bound, so an average of the two estimates is used. The underlying standard error for the coefficient (σ_β) can be approximated by:

$$\sigma_{\beta,\,2.5\ percentile} = \frac{\beta - \beta_{2.5\ percentile}}{1.96}$$

$$\sigma_{\beta,\,97.5\ percentile} = \frac{\beta_{97.5\ percentile} - \beta}{1.96}$$

$$\sigma_\beta \cong \frac{\sigma_{\beta,\,2.5\ percentile} + \sigma_{\beta,\,97.5\ percentile}}{2} .$$

D .2.3 Logistic Model

In some epidemiological studies, a logistic model is used to estimate the probability of an occurrence of an adverse health effect. Given a vector of explanatory variables, X, the logistic model assumes the probability of an occurrence is:

$$y = prob\left(occurrence | X \cdot \beta\right) = \left(\frac{e^{X \cdot \beta}}{1 + e^{X \cdot \beta}}\right) ,$$

where β is a vector of coefficients.[7] This may be rewritten as:

$$y = \frac{e^{X \cdot \beta}}{1 + e^{X \cdot \beta}} \cdot \frac{e^{-X \cdot \beta}}{e^{-X \cdot \beta}} = \frac{1}{1 + e^{-X \cdot \beta}}$$

The odds of an occurrence is:

$$odds = \frac{y}{1 - y} = \frac{\left(\frac{1}{1 + e^{-X \cdot \beta}}\right)}{1 - \frac{1}{1 + e^{-X \cdot \beta}}}$$

$$\Rightarrow odds = \frac{\left(\frac{1}{1 + e^{-X \cdot \beta}}\right)}{1 - \frac{1}{1 + e^{-X \cdot \beta}}} = \frac{\left(\frac{1}{1 + e^{-X \cdot \beta}}\right)}{\left(\frac{e^{-X \cdot \beta}}{1 + e^{-X \cdot \beta}}\right)} = \frac{1}{e^{-X \cdot \beta}} = e^{X \cdot \beta}$$

$$\Rightarrow \ln(odds) = X \cdot \beta \ .$$

The odds ratio for the control scenario ($odds_c$) versus the baseline ($odds_0$) is then:

$$odds\,ratio = \frac{odds_c}{odds_0} = \frac{\left(\frac{y_c}{1 - y_c}\right)}{\left(\frac{y_0}{1 - y_0}\right)} = \frac{\left(\frac{1}{e^{-X_c \cdot \beta}}\right)}{\left(\frac{1}{e^{-X_0 \cdot \beta}}\right)} = \frac{e^{X_c \cdot \beta}}{e^{X_0 \cdot \beta}} .$$

[7] Greene (1997, Chapter 19) presents models with discrete dependent variables; in particular, page 874 presents the logit model. See also Judge et al. (1985, p. 763).

The *change* in the probability of an occurrence from the baseline to the control (Δy), assuming that all the other covariates remain constant, may be derived from this odds ratio:

$$odds\,ratio = \frac{\left(\frac{y_c}{1-y_c}\right)}{\left(\frac{y_0}{1-y_0}\right)} = \frac{e^{X_c\beta}}{e^{X_0\beta}} = \frac{e^{\gamma}\cdot e^{PM_c\beta}}{e^{\gamma}\cdot e^{PM_0\beta}} = e^{\Delta PM\cdot\beta}$$

$$\frac{y_c}{1-y_c} = \left(\frac{y_0}{1-y_0}\right)\cdot e^{\Delta PM\cdot\beta}$$

$$y_c = \left(1-y_c\right)\cdot\left(\frac{y_0}{1-y_0}\right)\cdot e^{\Delta PM\cdot\beta}$$

$$y_c + y_c\cdot\left(\frac{y_0}{1-y_0}\right)\cdot e^{\Delta PM\cdot\beta} = \left(\frac{y_0}{1-y_0}\right)\cdot e^{\Delta PM\cdot\beta}$$

$$y_c\cdot\left[1+\left(\frac{y_0}{1-y_0}\right)\cdot e^{\Delta PM\cdot\beta}\right] = \left(\frac{y_0}{1-y_0}\right)\cdot e^{\Delta PM\cdot\beta}$$

$$y_c = \frac{\left(\frac{y_0}{1-y_0}\right)\cdot e^{\Delta PM\cdot\beta}}{1+\left(\frac{y_0}{1-y_0}\right)\cdot e^{\Delta PM\cdot\beta}} = \frac{y_0\cdot e^{\Delta PM\cdot\beta}}{1-y_0+y_0\cdot e^{\Delta PM\cdot\beta}}.$$

Multiplying by:

$$\frac{e^{-\Delta PM\cdot\beta}}{e^{-\Delta PM\cdot\beta}},$$

gives:

$$y_c = \frac{y_0}{(1 - y_0) \cdot e^{-\Delta PM \cdot \beta} + y_0}$$

$$\Delta y = y_c - y_0 = \frac{y_0}{(1 - y_0) \cdot e^{-\Delta PM \cdot \beta} + y_0} - y_0 \ .$$

The change in the number of cases of the adverse health effect is then obtained by multiplying by the relevant population:

$$\Delta Incidence = \Delta y \cdot pop = \left[\frac{y_0}{(1 - y_0) \cdot e^{-\Delta PM \cdot \beta} + y_0} - y_0 \right] \cdot pop \ .$$

When the coefficient (β) and its standard error (σ_β) are published (e.g., Pope et al., 1991, Table 5), then the coefficient estimates associated with the lower and upper bound may be calculated easily as follows:

$$\beta_{lower\ bound} = \beta - (1.96 \cdot \sigma_\beta)$$
$$\beta_{upper\ bound} = \beta + (1.96 \cdot \sigma_\beta) \ .$$

Often the logistic regression coefficients are not published, and only the odds ratio corresponding to a specified change in PM is presented (e.g., Schwartz et al., 1994). It is easy to calculate the underlying coefficient as follows:

$$\ln(odds\ ratio) = \Delta PM \cdot \beta$$

$$\Rightarrow \ \beta = \frac{\ln(odds\,ratio)}{\Delta PM} \ .$$

The coefficients associated with the lower and upper bound estimates of the odds ratios are calculated analogously.

The underlying standard error for the coefficient (σ_β) can be approximated by:

$$\sigma_{\beta,\ 2.5\ percentile} = \frac{\beta - \beta_{2.5\ percentile}}{1.96}$$

$$\sigma_{\beta,\ 97.5\ percentile} = \frac{\beta_{97.5\ percentile} - \beta}{1.96}$$

$$\sigma_\beta \cong \frac{\sigma_{\beta,\,2.5\ percentile} + \sigma_{\beta,\,97.5\ percentile}}{2}.$$

Sometimes, however, the relative risk is presented. The relative risk does not equal the odds ratio, and a different procedure should be used to estimate the underlying coefficient.[8]

The relative risk (RR) is simply:

$$RR = \frac{y_0}{y_c},$$

where y_0 is the risk (i.e., probability of an occurrence) at the baseline PM exposure and y_c is the risk at the control PM exposure.

When the baseline incidence rate (y_0) is given, then it is easy to solve for the control incidence rate (y_c):

$$y_c = \frac{y_0}{RR}.$$

The odds ratio, may then be calculated:

$$odds\,ratio = \frac{\frac{y_0}{1-y_0}}{\frac{y_c}{1-y_c}}.$$

Given the odds ratio, the underlying coefficient (β) may be calculated as before:

$$\beta = \frac{\ln(odds\,ratio)}{\Delta PM}.$$

The odds ratio and the coefficient calculated from it are dependent on the baseline and control incidence rates. Unfortunately, it is not always clear what the baseline and control incidence rates should be. Abbey et al. (1995b, Table 2) reported that there are 117 new cases of chronic bronchitis out of a sample of 1,631, or a 7.17 percent rate. In addition, they reported the relative risk (RR = 1.81) for a new case of chronic bronchitis associated with an annual mean concentration "increment" of 45 μg/m^3 of $PM_{2.5}$ exposure.

Assuming that the baseline rate for chronic bronchitis (y_0) should be 7.17 percent, the question becomes whether the "increment" of 45 μg/m^3 should be added to or subtracted from the existing $PM_{2.5}$

[8]Note that ESEERCO (1994, p. V-21) calculated (incorrectly) the underlying regression coefficient for Abbey et al. (1993, Table 5) by taking the logarithm of the relative risk and dividing by the change in TSP.

concentration. If added the control incidence rate (y_c) would be greater than the baseline rate (y_0), while subtraction would give a control rate less than the incidence rate. In effect, one might reasonably derive two estimates of the odds ratio:

$$odds\,ratio_1 = \frac{\left(\frac{y_0}{1-y_0}\right)}{\left(\frac{y_c}{1-y_c}\right)} = \frac{\left(\frac{1.81\cdot 0.0717}{1-(1.81\cdot 0.0717)}\right)}{\left(\frac{0.0717}{1-0.0717}\right)} = 1.931$$

$$odds\,ratio_2 = \frac{\left(\frac{y_0}{1-y_0}\right)}{\left(\frac{y_c}{1-y_c}\right)} = \frac{\left(\frac{0.0717}{1-0.0717}\right)}{\left(\frac{\frac{0.0717}{1.81}}{1-\frac{0.0717}{1.81}}\right)} = 1.873$$

$$\Rightarrow \beta_1 = \frac{\ln(1.931)}{45} = 0.01462$$

$$\Rightarrow \beta_2 = \frac{\ln(1.873)}{45} = 0.01394\,.$$

An alternative is to simply assume that the relative risk (1.81) is reasonably close to the odds ratio and calculate the underlying coefficient. It is easy to show that the relative risk equals:

$$RR = \frac{y_0}{y_c} = (1-y_0)\cdot e^{-\Delta PM\cdot\beta} + y_0\ .$$

Assuming that:

$$e^{-\Delta PM\cdot\beta} \cong (1-y_0)\cdot e^{-\Delta PM\cdot\beta} + y_0$$

$$\Rightarrow RR \cong e^{-\Delta PM\cdot\beta}\ .$$

It is then possible to calculate the underlying coefficient:

$$\frac{\ln(RR)}{-\Delta PM} \cong \beta$$

$$\Rightarrow \beta_3 = \frac{\ln(1.81)}{45} = 0.01319 .$$

Since this coefficient estimate is based on the assumption that

$$e^{-\Delta PM \cdot \beta} \cong \left(1 - y_0\right) \cdot e^{-\Delta PM \cdot \beta} + y_0 \ ,$$

it should be used in a C-R function that maintains this assumption. In effect, it should be applied to a log-linear C-R function:

$$\Delta y = \left[y_0 \cdot (e^{\beta \cdot \Delta PM} - 1) \right] .$$

Using the formula for the change in the incidence rate and assuming a 10 μg/m^3 decline in $PM_{2.5}$, it is shown that this results in changes within the bounds suggested by the two estimates based on using the estimated odds ratios:

$$\Delta y_1 = \frac{.0717}{\left(1 - 0.0717\right) \cdot e^{10 \cdot 0.01462} + 0.0717} - 0.0717 = -0.00914$$

$$\Delta y_2 = \frac{.0717}{\left(1 - 0.0717\right) \cdot e^{10 \cdot 0.01394} + 0.0717} - 0.0717 = -0.00874$$

$$\Delta y_3 = 0.0717 \cdot \left(e^{-10 \cdot 0.01319} - 1\right) = -0.00886 .$$

In this instance, it seems that simply using the relative risk to estimate the underlying coefficient results in a good approximation of the change in incidence. Since it is unclear which of the two other coefficients (β_1 or β_2) should be used -- as the published work was not explicit – the coefficient based on the relative risk and the log-linear functional form seems like a reasonable approach.

D .2.4 Cox proportional Hazards Model

Use of a Cox proportional hazards model in an epidemiological study results in a C-R function that is log-linear in form. It is often used to model survival times, and as a result, this discussion focuses on mortality impacts.

The Cox proportional hazards model is based on a hazard function, defined as the probability that an individual dies at time t, conditional on having survived up to time t (Collet, 1994, p. 10). More formally, the hazard function equals the probability density function for the risk of dying divided by one minus the cumulative probability density function:

$$h(X,t) = \frac{f(X,t)}{1 - F(X,t)}.$$

The proportional hazards model takes the form:

$$h(X,t) = h_0(t)e^{X \cdot \beta}$$

where X is a vector of explanatory variables, β is a vector of coefficients, and $h_0(t)$ is the so-called "baseline hazard" rate.[9] This terminology differs from that used in most of this discussion: this "baseline hazard" is the risk when all of the covariates (X) are set to zero; this is *not* the risk in the baseline scenario.

Taking the ratio of the hazard functions for the baseline and control scenarios gives the relative risk:

$$RR = \frac{h(X_0,t)}{h(X_c,t)} = \frac{h_0(t)e^{X_0 \cdot \beta}}{h_0(t)e^{X_c \cdot \beta}} = e^{\Delta PM \cdot \beta}.$$

where it is assumed that the only difference between the baseline and control is the level of PM pollution.

The relative risk is often presented rather than the coefficient β, so it is necessary to estimate β in order to develop functional relationship between ΔPM and Δy, as described previously for log-linear C-R functions.

D .3 General Issues in Estimating Health & Welfare Benefits

Changes in air pollution result in changes in a number of health and welfare effects, or "endpoints," that society values. This chapter discusses key issues in their estimation. The first part of this section discusses the development of C-R functions, based on the results from epidemiological studies, and the second part discusses some general issues that arise with C-R functions.

D .3.1 Choosing Epidemiological Studies and Developing Concentration-Response Functions

This section reviews the steps we performed in selecting and developing C-R functions for inclusion in BenMAP. The first section of this appendix describes how we chose studies from the epidemiological literature for use in the present analysis. In any given study, there are often a large number of estimated relationships between air pollution and adverse health effects, because the estimated relationship can depend on the number and types of pollutants included in the model, among other reasons. We then describe how we chose the specific estimated relationships, or models, from among the potentially large number available in any given study. And then we briefly discuss how we convert the estimated

[9]The Cox proportional hazards model is sometimes termed a "semi-parametric" model, because the baseline hazard rate is calculated using a non-parametric method, while the impact of explanatory variables is parameterized. Collet (1994) details the estimation of Cox proportional hazards models; in particular, see Collet's discussion (pp. 95-97) of nonparametric estimation of the baseline hazard.

model into C-R functions, which then allow us to quantify the change in adverse health effects due to a change in air pollution exposure.

Study Selection

We relied on an up-to-date assessment of the published scientific literature to ascertain the relationship between particulate matter and ozone exposure and adverse human health effects. We evaluated studies using a variety of selection criteria, including: its location and design, the characteristics of the study population, and whether the study was peer-reviewed (Exhibit D-2).

In selecting studies for use in this analysis, priority was given to studies that focused on $PM_{2.5}$ and ozone, given that the emissions reductions from nonroad sources are likely to result primarily in reduced ambient $PM_{2.5}$ and ozone levels. For a given health effect, if sufficient $PM_{2.5}$ studies were available, we selected them rather than PM_{10} studies in the base analysis. In addition, results from several recent studies allowed for the inclusion of new health effects, such as myocardial infarction for $PM_{2.5}$ and school loss days for ozone.

While a broad range of serious health effects have been associated with exposure to elevated ozone and PM levels, we include only a subset of health effects in this quantified benefit analysis. Health effects are excluded from this analysis for three reasons: (i) the possibility of double counting (such as hospital admissions for specific respiratory diseases); (ii) uncertainties in applying effect relationships based on clinical studies to the affected population; or (iii) a lack of an established C-R relationship.

A more detailed description of the studies and health effects included in this analysis are presented in Appendices F and G.

Exhibit D-2. Summary of Considerations Used in Selecting C-R Functions

Consideration	Comments
Peer reviewed research	Peer reviewed research is preferred to research that has not undergone the peer review process.
Study type	Among studies that consider chronic exposure (e.g., over a year or longer) prospective cohort studies are preferred over cross-sectional studies because they control for important individual-level confounding variables that cannot be controlled for in cross-sectional studies.
Study period	Studies examining a relatively longer period of time (and therefore having more data) are preferred, because they have greater statistical power to detect effects. More recent studies are also preferred because of possible changes in pollution mixes, medical care, and life style over time. However, when there are only a few studies available, studies from all years will be included.
Population attributes	The most technically appropriate measures of benefits would be based on C-R functions that cover the entire sensitive population, but allow for heterogeneity across age or other relevant demographic factors. In the absence of C-R functions specific to age, sex, preexisting condition status, or other relevant factors, it may be appropriate to select C-R functions that cover the broadest population, to match with the desired outcome of the analysis, which is total national-level health impacts.
Study size	Studies examining a relatively large sample are preferred because they generally have more power to detect small magnitude effects. A large sample can be obtained in several ways, either through a large population, or through repeated observations on a smaller population, i.e. through a symptom diary recorded for a panel of asthmatic children.
Study location	U.S. studies are more desirable than non-U.S. studies because of potential differences in pollution characteristics, exposure patterns, medical care system, population behavior and life style.
Pollutants included in model	When modeling the effects of ozone and PM (or other pollutant combinations) jointly, it is important to use properly specified C-R functions that include both pollutants. Use of single pollutant models in cases where both pollutants are expected to affect a health outcome can lead to double-counting when pollutants are correlated.
Measure of PM	For this analysis, C-R functions based on $PM_{2.5}$ are preferred to PM_{10} because reductions in emissions from diesel engines are expected to reduce fine particles and not have much impact on coarse particles. Where $PM_{2.5}$ functions are not available, PM_{10} functions are used as surrogates, recognizing that there will be potential downward (upward) biases if the fine fraction of PM_{10} is more (less) toxic than the coarse fraction.
Economically valuable health effects	Some health effects, such as forced expiratory volume and other technical measurements of lung function, are difficult to value in monetary terms. These health effects are not quantified in this analysis.
Non-overlapping endpoints	Although the benefits associated with each individual health endpoint may be analyzed separately, care must be exercised in selecting health endpoints to include in the overall benefits analysis because of the possibility of double counting of benefits. Including emergency room visits in a benefits analysis that already considers hospital admissions, for example, will result in double counting of some benefits if the category "hospital admissions" includes emergency room visits.

Model Selection

For any given study selected for use in this analysis, there are often multiple models quantifying the relationship between air pollution exposure and adverse health. For each model, we needed to identify the specific models that we would use to develop C-R functions.

Single Pollutant versus Multipollutant Models

Many of the epidemiological studies present results both for the case where only one pollutant is entered into the health effects model, or single-pollutant models, and where two or more pollutants are entered into the health effects model, or multi-pollutant models. When attempting to quantify the impact of a single pollutant, such as $PM_{2.5}$, on adverse health effects, the use of single-pollutant models may result in biased estimates. For example, to the extent that any of the co-pollutants present in the ambient air may have contributed to the health effects attributed to $PM_{2.5}$ in single pollutant models, risks attributed to $PM_{2.5}$ might be overestimated where C-R functions are based on single-pollutant models.

In multi-pollutant models, it may be difficult to sort out which pollutants are exerting an independent effect when pollutants in a given location are highly correlated. As discussed in the 2002 draft PM CD (U.S. EPA, 2002a), inclusion of pollutants that are highly correlated with one another can lead to misleading conclusions in identifying a specific causal pollutant. When collinearity exists, multi-pollutant models would be expected to produce unstable and statistically insignificant effects estimates for both PM and the co-pollutants (U.S. EPA, 2002a, p.9-130).

Single- and multi-pollutant models each have potential advantages and disadvantages, with neither type clearly preferable over the other, however, the regulatory focus of this analysis is on PM and ozone. For regulatory analyses which consider two pollutants together, adding incidence changes for a given health endpoint, based on a single-pollutant PM model, to the incidence changes based on a single-pollutant ozone model could result in an overestimate of incidence change, if both have an effect on the health endpoint and there is some collinearity between the two pollutants.

As a result, our first choice for this analysis is to use multi-pollutant models with both PM and ozone, rather than single-pollutant models and multi-pollutant models with other pollutants. If multi-pollutant models with both PM and ozone were not available, then models with other co-pollutants were preferred to single-pollutant models. In the absence of multi-pollutant models from a given study, single pollutant models were selected for use in the analysis.

Model Selection Criteria

In many epidemiological studies of air pollution and health, researchers estimate and present numerous single pollutant and multi-pollutant models for the same pollutant and health endpoint. These models may differ from each other in a number of characteristics, including: the functional form of the model, the covariates included in the model, the pollutant exposure metric, the lag structure, and the study population.

For the purposes of estimating health benefits associated with pollutant changes, it is neither realistic nor advantageous to include every model presented in each study. However, it is important that a relatively objective process be used to select among models. Described below are the criteria that were used as guidance in the selection of a particular model from among several models presented in a study. It is not possible in all cases to select a model using a completely objective and mechanical process. In many cases, professional judgement and an understanding of the study context are necessary as well to select the most appropriate models. Exhibit D-3 summarizes the selection criteria that we used.

Exhibit D-3. Description of Selection Criteria

Selection Criteria	Description
Goodness-of-fit statistics	If an appropriate measure for model selection is reported for each of several models in a study, then this measure may be used as the basis on which to select a model.
Best captures distributed lag	Select the model that appears to best capture a distributed lag effect: If multiple single-lag models and/or moving average models are specified, select the model with the largest effect estimate, all else equal.
Best set of control variables	Select the model which includes temporal variables (i.e. season, weather patterns, day of the week) and other known non-pollutant confounders, all else equal. Select the model which uses the most sophisticated methods of capturing the relationship between these variables and the dependent variable (e.g., affords the most flexibility in fitting possible nonlinear trends).
Useful for health effects modeling	The model must be in a form that is useful for health effects modeling (e.g., the pollutant variable should be a continuous variable rather than a categorical variable).
Biologically plausible	Select only those models that are biologically plausible.
Sample size	Select the model with the larger sample size, all else equal.

Goodness-of-Fit Statistics

Model specification (or mis-specification) is one of the most important issues confronting researchers – and those who apply the results of their research. The goal is to select the "right model" – i.e., the model that has included all the variables that should be in the model (i.e., are relevant) and has not included any variables that should not be in the model (i.e., that are irrelevant). However, is not often known which model is the "right model." There are several ways of selecting one model from among several. One way is to use a goodness of fit measure, which provides a measure of how well a model fits the data. There are a variety of goodness of fit measures available, but use of such measures can at times be misleading. In order to select models based on a goodness of fit criterion, it is important to understand the meaning behind typical goodness of fit measures.

One of the most common goodness of fit measures is R^2, often called the "explained variance" or the "coefficient of multiple determination." R^2 measures the proportion of the total variability in the dependent variable (e.g., the daily incidence of a health effect) that is explained by the linear regression. The closer R^2 is to 1.0, the greater this proportion. The problem with R^2, however, is that it can be increased simply by adding more independent variables to the model, regardless of the variables' relevance to predicting the dependent variable. In the extreme, if there are N observations in the dataset, a model with N explanatory variables will result in an R^2 of 1, but it would be a meaningless model with respect to predictive value. If several models are reported in a study, all with the same number of variables, this drawback is avoided. In that case, if R^2 is reported for each estimated model, this may be a reasonable measure of the relative goodness of fit of the models and an acceptable way to select a single model from among several.

In many cases, however, R^2 is a problematic measure of goodness of fit, for the reason stated above. In view of the drawback of R^2, several alternative measures of goodness of fit have been developed which essentially penalize the model for additional variables – or, equivalently, give "points" for parsimony. Two of the more commonly used of these measures are the adjusted R^2 and Akaike's Information Criterion (AIC). The selection criterion, using the adjusted R^2, is to select the model with the largest adjusted R^2; the

selection criterion, using the AIC, is to select the model with the smallest AIC. Both of these measures offset the incremental "fit" gained by including variables in the model with a "penalty" for increasing the number of explanatory (independent) variables. It should be noted, however, that several such measures have been suggested, and there is no clear way to determine which of these measures to select the "best" model is itself the "best" measure. Nevertheless, if a measure of goodness of fit (particularly, one of the measures that consider parsimony) is presented in a paper, this provides a reasonable means by which to select one model out of several.

Often, however, no goodness of fit measure is presented in a paper. A common approach for deciding the appropriate set of independent variables to be included in a model is to include a variable if its t-value exceeds the critical value for testing whether the variable's coefficient is significantly different from zero at the 5 percent level. Several variables, or an entire model, can similarly be tested with an F-test. (If all the coefficients are being tested jointly, the null hypothesis being tested is that all the coefficients are zero, in which case the model has no more predictive value than the mean of the dependent variable.) In some cases, a comparison of F-statistics (or their corresponding p-values) can be used to select from among several models – in particular, if the F-test for one model suggests that one cannot reject the null hypothesis at the five percent level whereas the F-test for another model suggests that one should reject the null hypothesis.[10]

For example, Stieb et al. (1996) estimated the association between ozone and ER visit rates using both a linear model (in which ER visit rate was a linear function of ozone level) and a quadratic model (in which ER visit rate was a linear function of ozone level squared). No goodness of fit measure was reported in the paper. However, model p-values were reported. The linear model was not statistically significant at the 5% level, whereas the quadratic model was highly significant. This suggests that the linear model does no better in predicting ER visits than the mean of ER visits, whereas the quadratic model has predictive value. The authors of the study themselves noted that "only ozone appeared to have a nonlinear relationship with visit rates" (Stieb et al., 1996, p. 1356) and that "quadratic, linear-quadratic, and indicator models consistently fit the data better than the linear model ..." (Stieb et al., 1996, p. 1358). Based on the relative model p-values presented in the paper, corroborated by the authors' observations, the quadratic model was selected for inclusion in this analysis.

Best Captures a Distributed Lag Effect.

The question of lags and the problems of correctly specifying the lag structure in a model has been discussed extensively (U.S. EPA, 2002a, Section 8.4.4). In many time-series studies, after the basic model is fit (before considering the pollutant of interest), several different lags are typically fit in separate single-lag models and the most significant lag is chosen. The 2002 draft PM CD notes that "while this practice may bias the chance of finding a significant association, without a firm biological reason to establish a fixed pre-determined lag, it appears reasonable" (U.S. EPA, 2002a, p. 8-237).

There is recent evidence (Schwartz, 2000b) that the relationship between PM and health effects may best be described by a distributed lag (i.e., the incidence of the health effect on day n is influenced by PM concentrations on day n, day n-1, day n-2 and so on). If this is the case, a model that includes only a single lag (e.g., a 0-day lag or a 1-day lag) is likely to understate the total impact of PM. The 2002 draft

[10] If F-statistics are both (or all) greater than the critical value, it is less clear that a comparison of these F-statistics would be a good way to select a model.

PM CD makes this point, noting that "if one chooses the most significant single lag day only, and if more than one lag day shows positive (significant or otherwise) associations with mortality, then reporting a RR [relative risk] for only one lag would also underestimate the pollution effects" (U.S. EPA, 2002a, p. 8-241). The same may hold true for other pollutants that have been associated with various health effects.

Several studies report similar models with different lag structures. For example, Moolgavkar (2000c) studied the relationship between air pollution and respiratory hospital admissions in three U.S. metropolitan areas. The author reports models with PM lagged from zero to five days. Since the lagging of PM was the only difference in the models and the relationship is probably best described using a distributed lag model, any of single-lag effect estimates are likely to underestimate the full effect. Therefore, we selected the model with the largest effect estimate.

Most Sophisticated Model That Includes Temporal Variables

A correctly specified model for evaluating air pollution and health would include all variables that are relevant independent predictors of the health outcome and none that are not. If there are variables that are known from prior literature to be associated with both air pollution and the health endpoint (e.g. temperature or season), then omitting these variables is likely to result in biased effect estimates – the C-R function would attribute too much or too little of the health effect to the pollutant. Since temporal and weather patterns are known to confound the relationship between air pollution and health, we selected the models which, all else equal, adjusted for these factors over those that did not.

Useful for Health Effects Modeling

In order for a model to be selected for use, the pollutant must be a continuous variable, so that changes in incidence of the health effect can be predicted to result from any change in pollutant concentration. Those models which examine the effects of being above or below a pollutant threshold or those that look at changes in health associated with categories of pollutant levels are not useful for this purpose.

For example, in a study of the association between air pollution and emergency room (ER) visits, Stieb et al. (1996) estimated several different models. One of these models relates ER visit rates to being above or below the 95th percentile value of ozone (that is, it essentially estimates an average ER visit rate for days above the 95th percentile value of ozone and a different average ER visit rate for days below it). In another study, Peters et al. (2001) estimated a model using quintiles of PM levels. None of these models is appropriate for use in CAPMS. Instead, we selected models which associate the incidence (rate) of a health effect with the pollutant concentration.

Biologically Plausible.

If a model includes a relationship that simply doesn't make biological sense, it is probably mis-specified and should not be used for predictive purposes. It is sometimes not clear, however, what is biologically plausible and what is not. For example, Stieb et al. (1996) estimated a linear-quadratic model – i.e., a model which included both ozone and the square of ozone as independent variables – in a study of air pollution and ER visits. The coefficient of the linear term in this model was negative, while the coefficient of the quadratic term was positive. A graph of the model showed a curve which "dips" at low levels of

ozone – i.e. at low ozone levels, increases in ozone are associated with *decreases* in risk. Since ozone is not likely to be beneficial at any levels, this model is not considered to be biologically plausible.

Sample Size

Several studies report the results of an analysis using different population subsets. All else equal, the model based on the larger population, which results in more statistical power, is selected. For example, Pope et al. (1991) studied the association between PM and respiratory health in children. The authors report results for a school-based sample of 34 children and a patient-based sample of 21 residents. Since there are many more observations in the school-based sample (3,096 versus 1,912) and no other significant differences between the models, the model estimated from the school-based sample is used.

Chen et al. (2000) examined the association between air pollution and school absenteeism. The authors reported results by elementary school grade and for all grades combined. With all else equal in the models, the C-R function with the larger number of observations was selected (i.e., the model based on all grades combined).

As discussed above, these criteria are used as general guidance when it is not obvious which model should be chosen from a particular study. The purpose of this process is to provide as objective a protocol as possible for selecting C-R functions. However, model selection can never be a completely mechanical and objective process because it often depends on the specific context of the particular study. In some cases, consideration of several of the aforementioned criteria must be weighed before selecting C-R functions for use in the analysis. The C-R functions selected for use in this analysis are described below with study summaries and a description of which model was selected, when multiple models are available.

D .4 Issues in Using Concentration-Response Functions

This section briefly summarizes some of the issues that arise when using C-R functions.

D .4.1 S-Plus Issue

Recently, the Health Effects Institute (HEI) reported findings by health researchers at Johns Hopkins University and others that have raised concerns about aspects of the statistical methods used in a number of recent time-series studies of short-term exposures to air pollution and health effects (Greenbaum, 2002). The estimates derived from the long-term exposure studies, which typically account for a major share of the economic benefits, are not affected. Similarly, the time-series studies employing generalized linear models (GLMs) or other parametric methods, as well as case-crossover studies, are not affected.

As discussed in HEI materials provided to EPA and to CASAC (Greenbaum, 2002), researchers working on the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) found problems in the default "convergence criteria" used in Generalized Additive Models (GAM) and a separate issue first identified by Canadian investigators about the potential to underestimate standard errors in the same statistical package. These and other scientists have begun to reanalyze the results of several important time series studies with alternative approaches that address these issues and have found a downward revision of some results. For example, the mortality risk estimates for short-term exposure to PM_{10} from NMMAPS were overestimated (this study was not used in this benefits analysis of fine particle effects). However,

both the relative magnitude and the direction of bias introduced by the convergence issue is case-specific. In the C-R functions described in detail in Appendices F and G, we have included the available reanalyses of previous studies, such as those collected in a recent document from the Health Effects Institute (2003).

D .4.2 Thresholds

When conducting clinical (chamber) and epidemiological studies, C-R functions may be estimated with or without explicit thresholds. Air pollution levels below the threshold are assumed to have no associated adverse health effects. When a threshold is not assumed, as is often the case in epidemiological studies, any exposure level is assumed to pose a non-zero risk of response to at least one segment of the population.

The possible existence of an effect threshold is a very important scientific question and issue for policy analyses. The EPA Science Advisory Board Advisory Council for Clean Air Compliance, which provides advice and review of EPA's methods for assessing the benefits and costs of the Clean Air Act under Section 812 of the Clean Air Act, has advised EPA that there is currently no scientific basis for selecting a threshold of 15 $\mu g/m^3$ or any other specific threshold for the PM-related health effects considered in typical benefits analyses (U.S. EPA, 1999b). This is supported by the recent literature on health effects of PM exposure (Rossi et al., 1999; Daniels et al., 2000; Pope, 2000; Schwartz, 2000c) which finds in most cases no evidence of a non-linear concentration-response relationship and certainly does not find a distinct threshold for health effects. The most recent draft of the EPA Air Quality Criteria for Particulate Matter (U.S. EPA, 2002a) reports only one study, analyzing data from Phoenix, AZ, that reported even limited evidence suggestive of a possible threshold for $PM_{2.5}$ (Smith et al., 2000).

Recent cohort analyses by the Health Effects Institute (Krewski et al., 2000) and Pope et al. (Pope et al., 2002) provide additional evidence of a quasi-linear concentration-response relationship between long-term exposures to $PM_{2.5}$ and mortality. According to the latest draft PM criteria document, Krewski et al. "found a visually near-linear relationship between all-cause and cardiopulmonary mortality residuals and mean sulfate concentrations, near-linear between cardiopulmonary mortality and mean $PM_{2.5}$, but a somewhat nonlinear relationship between all-cause mortality residuals and mean $PM_{2.5}$ concentrations that flattens above about 20 $\mu g/m^3$. The confidence bands around the fitted curves are very wide, however, neither requiring a linear relationship nor precluding a nonlinear relationship if suggested by reanalyses." The Pope et al. analysis, which represented an extension to the Krewski et al. analysis, found that the concentration-response relationships relating $PM_{2.5}$ and mortality "were not significantly different from linear associations."

Daniels et al. (2000) examined the presence of threshold in PM_{10} concentration-response relationships for daily mortality using the largest 20 U.S. cities for 1987-1994. The results of their models suggest that the linear model was preferred over spline and threshold models. Thus, these results suggest that linear models without a threshold may well be appropriate for estimating the effects of PM_{10} on the types of mortality of main interest. Schwartz and Zanobetti (2000) investigated the presence of threshold by simulation and actual data analysis of 10 U.S. cities. In the analysis of real data from 10 cities, the combined concentration-response curve did not show evidence of a threshold in the PM_{10}-mortality associations. Schwartz et al. (2002) investigated thresholds by combining data on the $PM_{2.5}$-mortality relationships for six cities and found an essentially linear relationship down to 2 $\mu g/m^3$, which is at or below anthropogenic background in most areas. They also examined just traffic related particles and again found no evidence of a threshold. The Smith et al. (2000) study of associations between daily total mortality and $PM_{2.5}$ and $PM_{10-2.5}$ in Phoenix, AZ (during 1995-1997) also investigated the possibility of a threshold using

a piecewise linear model and a cubic spline model. For both the piecewise linear and cubic spline models, the analysis suggested a threshold of around 20 to 25 $\mu g/m^3$. However, the concentration-response curve for $PM_{2.5}$ presented in this publication suggests more of a U- or V-shaped relationship than the usual "hockey stick" threshold relationship.

Finally, in a recent review of methods for estimating the public health benefits of air pollution regulations, National Research Council (2002) concluded that there is no evidence for any departure from linearity in the observed range of exposure to PM_{10} or $PM_{2.5}$, nor any indication of a threshold. They cite the weight of evidence available from both short and long term exposure models and the similar effects found in cities with low and high ambient concentrations of PM.

D .4.3 Degree of Prematurity of Mortality

It is possible that the short-term studies are detecting an association between air pollution and mortality that is primarily occurring among terminally ill people. Critics of the use of short-term studies for policy analysis purposes correctly point out that an added risk factor that results in terminally ill people dying a few days or weeks earlier than they otherwise would have (referred to as "short-term harvesting") is potentially included in the measured PM mortality "signal" detected in such a study. While some of the detected excess deaths may have resulted in a substantial reduction in lifespan, others may have resulted in a relatively small decrease in lifespan. Studies by Spix et al (1993) and Pope et al. (1992) yield conflicting evidence, suggesting that harvesting may represent anywhere from zero to 50 percent of the deaths estimated in short-term studies. However, recent work by Zeger et al. (1999), Schwartz (2000a), and Zanobetti et al. (2002) that focused exclusively on this issue, reported that short-term harvesting does not play a major role in the PM-mortality relationship.[11]

Moreover, it is not likely that the excess mortality reported in a long-term prospective cohort study like Pope et al. (1995) contains any significant amount of this short-term harvesting. The Cox proportional hazard statistical model used in the Pope study examines the question of survivability throughout the study period (ten years). Deaths that are premature by only a few days or weeks within the ten-year study period (for example, the deaths of terminally ill patients, triggered by a short duration PM episode) are likely to have little impact on the calculation of the average probability of surviving the entire ten-year interval.

D .4.4 Estimating Effects for Multiple Age Groups

For analyses focusing on a year well past the year 2000, you should note that the population age distribution is expected to change over time, with a greater percentage of the population moving into older age categories. Because baseline incidence rates for older populations tend to exceed those for younger populations for several health endpoints (most importantly, for mortality), this demographic shift has important implications for the estimation of future-year incidence change. If you apply a C-R function to an entire population, using one average baseline incidence, this demographic shift would be missed, and the future-year incidence change would be significantly underestimated.

[11]Zeger et al. (1999, p. 171) reported that: "The TSP-mortality association in Philadelphia is inconsistent with the harvesting-only hypothesis, and the harvesting-resistant estimates of the TSP relative risk are actually larger – not smaller – than the ordinary estimates."

Appendix D. Types of C-R Functions & Issues in the Estimation of Adverse Health Effects

To take into account projected demographic shifts and the corresponding implications for predicted incidence change, we have included C-R functions for separate age groups within the entire population to which a C-R function is applicable, using projected populations in each age group. Projected baseline incidences (incidence rates times populations) used in the calculation of future-year pollutant-related incidence change therefore better reflect the expected demographic shifts.

The ideal approach would be have future-year incidence rates. However, these are not available. Thus to the extent that you use baseline incidence rates (which may decline slightly over time for younger age groups and increase for the oldest groups), you may be mis-estimating incidence change for particular age groups to the extent that baseline incidence rates change over time.

Appendix E: Sources of Prevalence and Incidence Data

Concentration-Response (C-R) functions developed from log-linear or logistic models estimate the percent change in an adverse health effect associated with a given pollutant change. In order to estimate the absolute change in incidence using these functions, we need the baseline incidence rate of the adverse health effect. This appendix describes the data used to estimate baseline incidence rates for the health effects considered in this analysis.

E .1 Mortality

Age, cause, and county-specific mortality rates were obtained from the U.S. Centers for Disease Control (CDC) for the years 1996 through 1998. CDC maintains an online data repository of health statistics, CDC Wonder, accessible at http://wonder.cdc.gov/. The mortality rates provided are derived from U.S. death records and U.S. Census Bureau postcensal population estimates. Mortality rates were averaged across three years (1996 through 1998) to provide more stable estimates. When estimating rates for age groups that differed from the CDC Wonder groupings, we assumed that rates were uniform across all ages in the reported age group. For example, to estimate mortality rates for individuals ages 30 and up, we scaled the 25-34 year old death count and population by one-half and then generated a population-weighted mortality rate using data for the older age groups. Population-weighted national mortality rates are presented in Exhibit E-1.

Exhibit E-1. National Mortality Rates for Selected Conditions, by Age Group

Mortality Category (ICD codes)	Mortality Rate by Age Group (deaths per 100 people per year)									
	0-17	18-24	25-29	30-34	35-44	45-54	55-64	65-74	75-84	85+
All-Cause	0.045	0.093	0.119	0.119	0.211	0.437	1.056	2.518	5.765	15.160
Non-Accidental (ICD <800)	0.025	0.022	0.057	0.057	0.150	0.383	1.006	2.453	5.637	14.859
Chronic Lung Disease (ICD 490-496)	0.000	0.001	0.001	0.001	0.002	0.009	0.046	0.166	0.367	0.561
Cardio-Pulmonary	0.004	0.005	0.013	0.013	0.044	0.143	0.420	1.163	3.179	9.846

Source: We obtained data from 1996-1998 from the CDC Wonder (http://wonder.cdc.gov/). County-specific rates are used in the C-R functions.

E .2 Hospitalizations

Regional hospitalization counts were obtained from the National Center for Health Statistics' (NCHS) National Hospital Discharge Survey (NHDS). NHDS is a sample-based survey of non-Federal, short-stay hospitals (<30 days)[12], and is the principal source of nationwide hospitalization data. The survey collects data on patient characteristics, diagnoses, and medical procedures.

[12]The following hospital types are excluded from the survey: hospitals with an average patient length of stay of greater than 30 days, federal, military, Department of Veterans Affairs hospitals, institutional hospitals (e.g. prisons), and hospitals with fewer than six beds.

Public use data files for the year 1999 survey were downloaded[13] and processed to estimate hospitalization counts by region. NCHS groups states into four regions using the following groupings defined by the U.S. Bureau of the Census:

- Northeast - Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania
- Midwest - Ohio, Indiana, Illinois, Michigan, Wisconsin, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas
- South - Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, Texas
- West - Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Washington, Oregon, California, Alaska, Hawaii

We calculated per capita hospitalization rates, by dividing these counts by the estimated regional population estimates for 1999 that we derived from the U.S. Bureau of the Census and the population projections used by NHDS to generate the counts. Note that NHDS started with hospital admission counts, based on a sample of admissions, and then they used population estimates to generate population-weighted hospital admission counts that are representative of each region. This weighting used forecasts of 1999 population data. Ideally, we would use these same forecasts to generate our admission rates. However, while NHDS presented counts of hospital admissions with a high degree of age specificity, it presented regional population data for only four age groups: 0-14, 15-44, 45-64, and 65+.[14] Using only the NHDS data, we would be limited to calculating regional admission rates for four groups. Because we are interested in a broader range of age groups, we turned to 2000 Census.

We used the 2000 Census to obtain more age specificity, and then corrected the 2000 Census figures so that the total population equaled the total for 1999 forecasted by NHDS. That is, we sued the following procedure: (1) we calculated the count of hospital admissions by region in 1999 for the age groups of interest, (2) we calculated the 2000 regional populations corresponding to these age groups, (3) calculated regional correction factors, that equal the regional total population in 1999 divided by the regional total population in 2000 by region, (4) multiplied the 2000 population estimates by these correction factors, and (5) divided the 1999 regional count of hospital admissions by the estimated 1999 population.

The endpoints in hospitalization studies are defined using different combinations of ICD codes. Rather than generating a unique baseline incidence rate for each ICD code combination, for the purposes of this analysis, we identified a core group of hospitalization rates from the studies and applied the appropriate combinations of these rates in the C-R functions:

- all respiratory (ICD-9 460-519)
- chronic lung disease (ICD-9 490-496)
- asthma (ICD-9 493)
- pneumonia (ICD-9 480-487)
- acute bronchitis (ICD-9 466)
- acute laryngitis (ICD-9 464)
- all cardiovascular (ICD-9 390-459)
- ischemic heart disease (ICD-9 410-414)

[13] Data are available at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHDS/

[14] See: 1999nhds_summary.pdf (p. 187) for published regional population estimates for 1999.

- dysrhythmia (ICD-9 427)
- congestive heart failure (ICD-9 428)

For each C-R function, we selected the baseline rate or combination of rates that most closely matches to the study endpoint definition. For studies that define chronic lung disease as ICD 490-492, 494-496, we subtracted the incidence rate for asthma (ICD 493) from the chronic lung disease rate (ICD 490-496). In some cases, the baseline rate will not match exactly to the endpoint definition in the study. For example, Burnett et al. (2001) studied the following respiratory conditions in infants <2 years of age: ICD 464.4, 466, 480-486, 493. For this C-R function we apply an aggregate of the following rates: ICD 464, 466, 480-487, 493. Although they do not match exactly, we assume that relationship observed between the pollutant and study-defined endpoint is applicable for the additional codes. Exhibit E-2 presents a summary of the national hospitalization rates for 1999 from NHDS.

Exhibit E-2. Hospitalization Rates, by Region and Age Group

Hospitalization Category		ICD-9 Codes	Hospitalization Rate by Age Group (admissions per 100 people per year)						
			0-18	18-24	25-34	35-44	45-54	55-64	65+
Respiratory	all respiratory	460-519	1.066	0.271	0.318	0.446	0.763	1.632	5.200
	acute laryngitis	464	0.055	0.002	0.001	0.002	0.008	0.000	0.005
	acute bronchitis	466	0.283	0.017	0.014	0.017	0.027	0.040	0.156
	pneumonia	480-487	0.308	0.069	0.103	0.155	0.256	0.561	2.355
	asthma	493	0.281	0.081	0.110	0.099	0.144	0.161	0.205
	chronic lung disease	490-496	0.291	0.089	0.124	0.148	0.301	0.711	1.573
Cardiovascular	all cardiovascular	390-459	0.043	0.084	0.206	0.678	1.926	4.389	11.629
	ischemic heart disease	410-414	0.004	0.008	0.031	0.231	0.902	2.021	3.708
	dysrhythmia	427	0.011	0.017	0.027	0.076	0.158	0.392	1.387
	congestive heart failure	428	0.003	0.005	0.011	0.011	0.160	0.469	2.167

Source: As described in the text, we obtained the regional count of hospital admissions from National Hospital Discharge Survey (NHDS), and we obtained the population data from the 2000 U.S. Census and NHDS.

E .3 Emergency Room Visits for Asthma

Regional asthma emergency room visit counts were obtained from the National Hospital Ambulatory Medical Care Survey (NHAMCS). NHAMCS is a sample-based survey, conducted by NCHS, designed to collect national data on ambulatory care utilization in hospital emergency and outpatient departments of non-Federal, short-stay hospitals (<30 days).[15]

[15] The target universe of the NHAMCS is in-person visits made in the United States to emergency and outpatient departments of non-Federal, short-stay hospitals (hospitals with an average stay of less than 30 days) or those whose specialty is general (medical or surgical) or children's general.

Public use data files for the year 2000 survey were downloaded[16] and processed to estimate hospitalization counts by region. We obtained population estimates from the 2000 U.S. Census. The NCHS regional groupings described above were used to estimate regional emergency room visit rates. Exhibit E-3 presents the estimated asthma emergency room rates by region.

Exhibit E-3. Emergency Room Visit Rates for Asthma, by Region and Age Group

ER Category	ICD-9 Code	Region	ER Visit Rate (visits per 100 people per year)		
			0-18	18-64	65+
asthma	493	Northeast	0.761	0.802	0.300
		Midwest	1.476	0.877	0.334
		South	1.243	0.420	0.192
		West	0.381	0.381	0.137

Source: We obtained ER visit counts for the year 2000 from the National Hospital Ambulatory Medical Care Survey (NHAMCS) and population data were obtained from the 2000 U.S. Census.

E .4 Nonfatal Heart Attacks

The relationship between short-term particulate matter exposure and heart attacks was quantified in a case-crossover analysis by Peters et al. (2001). The study population was selected from heart attack survivors in a medical clinic. Therefore, the applicable population to apply to the C-R function is all individuals surviving a heart attack in a given year. Several data sources are available to estimate the number of heart attacks per year. For example, several cohort studies have reported estimates of heart attack incidence rates in the specific populations under study. However, these rates depend on the specific characteristics of the populations under study and may not be the best data to extrapolate nationally. The American Heart Association reports approximately 540,000 new heart attacks per year using data from a multi-center study (Haase, 2002, to be published in the American Heart Association's 2003 Statistical Handbook). Exclusion of heart attack deaths reported by CDC Wonder yields approximately 330,000 nonfatal cases per year.

An alternative approach to the estimation of heart attack rates is to use data from the National Hospital Discharge Survey, assuming that all heart attacks that are not instantly fatal will result in a hospitalization. According to the National Hospital Discharge Survey, in 1999 there were approximately 829,000 hospitalizations due to heart attacks (acute myocardial infarction: ICD-9 410) (Popovic, 2001, Table 8). We used regional hospitalization rates over estimates extrapolated from cohort studies because the former is part of a nationally representative survey with a larger sample size, which is intended to provide reliable national estimates. As additional information is provided regarding the American Heart Association methodology, we will evaluate the usefulness of this estimate of heart attack incidence.

Rosamond et al. (1999) reported that approximately six percent of male and eight percent of female hospitalized heart attack patients die within 28 days (either in or outside of the hospital). We, therefore, applied a factor of 0.93 to the count of hospitalizations to estimate the number of nonfatal heart attacks per

[16] Data are available at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHAMCS/

year. To estimate the *rate* of nonfatal heart attack, we divided the count by the population estimate for 2000 from the U.S. Census. Exhibit E-4 presents the regional nonfatal heart attack incidence rates.

Exhibit E-4. Nonfatal Heart Attack Rates, by Region and Age Group

Endpoint (ICD codes)	Region	Nonfatal Heart Attack Rate (cases per 100 people per year) [a]		
		0-18	18-64	65+
nonfatal heart attacks (ICD-9 410)	Northeast	0.0000	0.2167	1.6359
	Midwest	0.0003	0.1772	1.4898
	South	0.0006	0.1620	1.1797
	West	0.0000	0.1391	1.1971

[a] Rates are based on data from the 1999 National Hospital Discharge Survey (NHDS) and an estimate from Rosamond et al. (1999) that approximately 7% of individuals hospitalized for a heart attack die within 28 days.

E .5 School Loss Days

Epidemiological studies have examined the relationship between air pollution and a variety of measures of school absence. These measures include: school loss days for all causes, illness-related, and respiratory illness-related. We have two sources of information. The first is the National Center for Education Statistics, which provided an estimate of all-cause school loss days, and the other is the National Health Interview Survey (Adams et al., 1999, Table 47), which has data on different categories of acute school loss days. Exhibit E-5 presents the illness-related rates used in this analysis.

E .5.1 All-Cause School Loss Rates

Based on data from the U.S. Department of Education (1996, Table 42-1), the National Center for Education Statistics estimates that for the 1993-1994 school year, 5.5 percent of students are absent from school on a given day. This estimate is comparable to study-specific estimates from Chen et al. (2000) and Ransom and Pope (1992), which ranged from 4.5 to 5.1 percent.

We use the total or all-cause school absence rate in C-R functions based on studies by Chen et al. (2000), Gilliland et al. (2001) and Ransom et al. (1992). We also use the all-cause school absence rate as a population adjustment in C-R functions derived from Gilliland et al. (2001), for which it is necessary to estimate the average proportion of children attending school on a given day. This is described in more detail in the specific C-R function summaries.

E .5.2 Illness-Related School Loss Rates

The National Health Interview Survey (NHIS) has regional estimates of school loss days due to a variety of acute conditions (Adams et al., 1999). NHIS is a nationwide sample-based survey of the health of the noninstitutionalized, civilian population, conducted by NCHS. The survey collects data on acute conditions, prevalence of chronic conditions, episodes of injury, activity limitations, and self-reported health status. However, it does not provide an estimate of all-cause school loss days.

In estimating illness-related school loss days, we started with school loss days due to acute problems (Adams et al., 1999, Table 47) and subtracted lost days due to injuries, in order to match the definition of the study used in the C-R function to estimate illness-related school absences (Gilliland et al., 2001). We then divided by 180 school days per to estimate *illness*-related school absence rates per school day. Similarly, when estimating respiratory illness-related school loss days, we use data from Adams et al. (1999, Table 47). Note that we estimated 180 school days in a year to calculate respiratory illness-related school absence rates per year.

Exhibit E-5. School Loss Day Rates

Type of School Loss Day [a]	Absence Rate by Region (cases per 100 students per year)			
	Northeast	Midwest	South	West
Respiratory illness-related absences	131.4	165.6	109.8	223.2
Illness-related absences	244.8	262.8	255.6	370.8
All-cause	990.0	990.0	990.0	990.0

[a] We based illness-related school loss day rates on data from the 1996 NHIS (Adams et al., 1999, Table 47) and an estimate of 180 school days per year. This excludes school loss days due to injuries. We based the all-cause school loss day rate on data from the National Center for Education Statistics (U.S. Department of Education, 1996, Table 42-1).

E .6 Other Acute and Chronic Effects

For many of the minor effect studies, baseline rates from a single study are often the only source of information, and we assume that these rates hold for locations in the U.S. The use of study-specific estimates are likely to increase the uncertainty around the estimate because they are often estimated from a single location using a relatively small sample. These endpoints include: acute bronchitis, chronic bronchitis, upper respiratory symptoms, lower respiratory symptoms. Exhibit E-6 presents a summary of these baseline rates.

Exhibit E-6. Selected Acute and Chronic Effects Rates

Endpoint	Age	Parameter [a]	Rate	Source
Acute Bronchitis	8-12	Incidence	4.300	(American Lung Association, 2002a, Table 11)
Chronic Bronchitis	27+	Incidence	0.378	(Abbey et al., 1993, Table 3)
Chronic Bronchitis	18+	Prevalence	4.43%	(American Lung Association, 2002b, Table 4)
	18-44		3.67%	
	45-64		5.05%	
	65+		5.87%	
Lower Respiratory Symptoms (LRS)	7-14	Incidence	43.8	(Schwartz et al., 1994, Table 2)
Minor Restricted Activity Days (MRAD)	18-64	Incidence	780.0	(Ostro and Rothschild, 1989, p. 243)
Work Loss Day (WLD)	18-64	Incidence	217.2	(Adams et al., 1999, Table 41); (U.S. Bureau of the Census, 1997, No. 22)
	18-24		197.1	
	25-44		247.5	
	45-64		179.6	

[a] The incidence rate is the number of cases per 100 people per year. Prevalence refers to the fraction of people that have a particular illness during a particular time period.

E .6.1 Acute Bronchitis

The annual rate of acute bronchitis for children ages 5 to 17 was obtained from the American Lung Association (2002a, Table 11). The authors reported an annual incidence rate per person of 0.043, derived from the 1996 National Health Interview Survey.

E .6.2 Chronic Bronchitis Incidence Rate

The annual incidence rate for chronic bronchitis is estimated from data reported by Abbey et al.(1993, Table 3). The rate is calculated by taking the number of new cases (234), dividing by the number of individuals in the sample (3,310), dividing by the ten years covered in the sample, and then multiplying by one minus the reversal rate (estimated to be 46.6% based on Abbey et al. (1995a, Table 1)). We then multiplied this result by 100 to calculate an annual incidence rate per 100 people of 0.378.

Age-specific incidence rates are not available. Abbey et al. (1995a, Table 1) did report the incidences by three age groups (25-54, 55-74, and 75+) for "cough type" and "sputum type" bronchitis. However, they did not report an overall incidence rate for bronchitis by age-group. Since, the cough and sputum types of bronchitis overlap to an unknown extent, we did not attempt to generate age-specific incidence rates for the over-all rate of bronchitis.

E .6.3 Chronic Bronchitis Prevalence Rate

We obtained the annual prevalence rate for chronic bronchitis from the American Lung Association (2002b, Table 4). Based on an analysis of 1999 National Health Interview Survey data, they estimated a rate of 0.0443 for persons 18 and older, they also reported the following prevalence rates for people in the age groups 18-44, 45-64, and 65+: 0.0367, 0.0505, and 0.0587, respectively.

E .6.4 Lower Respiratory Symptoms

Lower respiratory symptoms (LRS) are defined as two or more of the following: cough, chest pain, phlegm, wheeze. The proposed yearly incidence rate for 100 people, 43.8, is based on the percentiles in Schwartz et al. (Schwartz et al., 1994, Table 2). The authors did not report the mean incidence rate, but rather reported various percentiles from the incidence rate distribution. The percentiles and associated per person per day values are 10^{th} = 0 percent, 25^{th} = 0 percent, 50^{th} = 0 percent, 75^{th} = 0.29 percent, and 90^{th} = 0.34 percent. The most conservative estimate consistent with the data are to assume the incidence per person per day is zero up to the 75^{th} percentile, a constant 0.29 percent between the 75^{th} and 90^{th} percentiles, and a constant 0.34 percent between the 90^{th} and 100^{th} percentiles. Alternatively, assuming a linear slope between the 50^{th} and 75^{th}, 75^{th} and 90^{th}, and 90^{th} to 100^{th} percentiles, the estimated mean incidence rate per person per day is 0.12 percent.[17] We used the latter approach in this analysis, and then multiplied by 100 and by 365 to calculate the incidence rate per 100 people per year.

E .6.5 Minor Restricted Activity Days (MRAD)

Ostro and Rothschild (1989, p. 243) provide an estimate of the annual incidence rate of MRADs (7.8). We multiplied this estimate by 100 to get an annual rate per 100 people.

E .6.6 Work Loss Days

The yearly work-loss-day incidence rate per 100 people is based on estimates from the 1996 National Health Interview Survey (Adams et al., 1999, Table 41). They reported a total annual work loss days of 352 million for individuals ages 18 to 65. The total population of individuals of this age group in 1996 (162 million) was obtained from (U.S. Bureau of the Census, 1997, No. 22). The average annual rate of work loss days per individual (2.17) was multiplied by 100 to obtain the average yearly work-loss-day rate of 217 per 100 people. Using a similar approach, we calculated work-loss-day rates for ages 18-24, 25-44, and 45-64, respectively.

E .7 Asthma-Related Health Effects

Several studies have examined the impact of air pollution on asthma development or exacerbation. Many of the baseline incidence rates used in the C-R functions are based on study-specific estimates. The baseline rates for the various endpoints are described below and summarized in Exhibit E-7.

[17] For example, the 62.5^{th} percentile would have an estimated incidence rate per person per day of 0.145 percent.

Exhibit E-7. Asthma-Related Health Effects Rates

Endpoint	Age	Parameter [a]	Rate	Source
Acute Bronchitis	9-15	Incidence	32.6	(McConnell et al., 1999, Table 2)
Asthma Attacks	18+	Incidence	2008	1999 National Health Interview Survey
Asthma Exacerbation, Shortness of Breath, African American	8-13	Incidence	1351	(Ostro et al., 2001, p.202)
	8-13	Prevalence	7.40%	
Asthma Exacerbation, Wheeze, African American	8-13	Incidence	2774	(Ostro et al., 2001, p.202)
	8-13	Prevalence	17.30%	
Asthma Exacerbation, Cough, African American	8-13	Incidence	2446	(Ostro et al., 2001, p.202)
	8-13	Prevalence	14.50%	
Asthma Exacerbation, Cough	6-13	Incidence	3139	(Vedal et al., 1998, Table 1 p. 1038)
Asthma Exacerbation, One or more symptoms	5-13	Incidence	21900	(Yu et al., 2000, Table 2 p. 1212)
Chronic Asthma, Male	27+	Incidence	0.219	(McDonnell et al., 1999, Table 4)
Phlegm	9-15	Incidence	25.7	(McConnell et al., 1999, Table 2)
Upper Respiratory Symptoms (URS)[2]	9-11	Incidence	12479	(Pope et al., 1991, Table 2)

[a] The incidence rate is the number of cases per 100 people per year. Prevalence refers to the fraction of people that have a particular illness during a particular time period.

E .7.1 Asthma Attacks

The annual rate of asthma attacks among asthmatics is estimated from the 1999 National Health Interview Survey. Individuals with asthma were asked about the number of wheezing attacks per year. The average number of wheezing attacks per year was multiplied by 100 to obtain a wheezing attack rate per year per 100 people for individuals 18 and older. We assume that this rate of wheezing attacks can be used as a surrogate for asthma attacks.

Note that the same survey examined wheezing attacks for children. However, the number of wheezing attacks per year were censored at 12 (compared to censoring at 95 for adults). Due to the potential for underestimation of the number of children's wheezing attacks, we used the adult rate for all individuals.

E .7.2 Asthma Exacerbation

There are a variety of types of symptoms for asthma exacerbation. We calculated rates for shortness of breath, wheeze, cough, and other asthma related effects.

E .7.3 Shortness of Breath

To estimate the annual rate of new shortness of breath episodes among African-American asthmatics, ages 8-13, we used the rate reported by Ostro et al. (2001, p.202). We estimated the daily prevalence of shortness of breath episodes among African-American asthmatics, ages 8-13, by taking a weighted average of the reported rates in Ostro et al. (2001, p.202).

E .7.4 Wheeze

The daily rate of new wheeze episodes among African-American asthmatics, ages 8-13, is reported by Ostro et al. (2001, p.202) as 0.076. We multiplied this value by 100 and by 365 to get the annual incidence rate per 100 people. The daily rate of prevalent wheeze episodes (0.173) among African-American asthmatics, ages 8-13, is estimated by taking a weighted average of the reported rates in Ostro et al. (2001, p.202).

E .7.5 Cough

The daily rate of new cough episodes among African-American asthmatics, ages 8-13, is reported by Ostro et al. (2001, p.202) as 0.067. We multiplied this value by 100 and by 365 to get the annual incidence rate per 100 people. The daily rate of prevalent cough episodes (0.145) among African-American asthmatics, ages 8-13, is estimated by taking a weighted average of the reported rates in Ostro et al. (2001, p.202).

E .7.6 One or More Symptoms

Yu et al. (2000, Table 2, p. 1212) reported a daily rate of at least one asthma episode per asthmatic child ages 5-13. An asthma episode is defined as at least one of the following asthma symptoms: wheezing, coughing, chest tightness, or shortness of breath.

E .7.7 Chronic Asthma

We derived the annual incidence rate per 100 people by taking the number of new cases (32), dividing by the number of individuals in the sample (972), as reported by (McDonnell et al., 1999, Table 4), and then dividing by the 15 years in the sample. We then multiplied by 100 to get the annual incidence rate per 100 people.

E .7.8 Upper Respiratory Symptoms

Upper Respiratory Symptoms are defined as one or more of the following: runny or stuffy nose; wet cough; burning, aching, or red eyes. Using the incidence rates for upper respiratory symptoms among asthmatics, published in Pope et al. (1991, Table 2), we calculated a sample size-weighted average incidence rate.

E .7.9 Asthma Population Estimates

In studies examining the association between air pollution and the development or exacerbation of asthma, often times an estimate of the percent of the population with asthma is required. Asthma percentages were obtained either directly from the National Health Interview Survey (NHIS) or an American Lung Association (2002c) report summarizing data from NHIS. Exhibit E-8 presents asthma prevalence rates used define asthmatic populations in the C-R functions.

Exhibit E-8. Asthma Prevalence Rates Used to Estimate Asthmatic Populations

Population Group	Prevalence	Source
All Ages	3.86%	American Lung Association (2002c, Table 7) [a]
<18	5.27%	
5-17	5.67%	
18-44	3.71%	
45-64	3.33%	
65+	2.21%	
African-American, 5 to 17	7.26%	American Lung Association (2002c, Table 9) [a]
African-American, <18	7.35%	
Male, 27+	2.10%	2000 NHIS public use data files [b]

[a] The work by the American Lung Association is based on the 1999 National Health Interview Survey.

[b] See ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/

Appendix F: Particulate Matter Concentration-Response Functions

In this Appendix, we present the concentration-response (C-R) functions used to estimate PM-related adverse health effects. Each sub-section has an Exhibit with a brief description of the C-R function and the underlying parameters. Following each Exhibit, we present a brief summary of each of the studies and any items that are unique to the study.

Note that the main text describes the methods that we used to choose these C-R functions from the wide range available in the literature. In addition, Appendix D mathematically derives the standard types of C-R functions that we encountered in the epidemiological literature, such as, log-linear, logistic and linear, so we simply note here the type of functional form. Finally, Appendix E presents a description of the sources for the incidence and prevalence data used in these C-R functions.

Appendix F. Particulate Matter C-R Functions

Exhibit F-1. Concentration-Response (C-R) Functions for Particulate Matter and Long-Term Mortality

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time	Beta	Std Error	Functional Form	Notes
All Cause	$PM_{2.5}$	Krewski et al.	2000	63 cities	30+	All	All	None	Annual Avg	0.004626	0.001205	Log-linear	ACS reanalysis
All Cause	$PM_{2.5}$	Krewski et al.	2000	50 cities	30+	All	All	None	Annual Median	0.005348	0.001464	Log-linear	ACS reanalysis
All Cause	$PM_{2.5}$	Krewski et al.	2000	nationwide	30+	All	All	None	Annual Median	0.010394	0.002902	Log-linear	ACS reanalysis; RE Ind Cities
All Cause	$PM_{2.5}$	Krewski et al.	2000	nationwide	30+	All	All	None	Annual Median	0.006058	0.003383	Log-linear	ACS reanalysis; RE Reg Adj
All Cause	$PM_{2.5}$	Krewski et al.	2000	6 cities	25+	All	All	None	Annual Avg	0.013272	0.004070	Log-linear	Six Cities reanalysis
All Cause	$PM_{2.5}$	Pope et al.	1995	50 cities	30+	All	All	None	Annual Median	0.006408	0.001509	Log-linear	
All Cause	$PM_{2.5}$	Dockery et al.	1993	6 cities	25+	All	All	None	Annual Avg	0.012425	0.004228	Log-linear	
All Cause	$PM_{2.5}$	Pope et al.	2002	61 cities	30+	All	All	None	Annual Avg	0.004018	0.001642	Log-linear	'79-'83 air data
All Cause	$PM_{2.5}$	Pope et al.	2002	51 cities	30+	All	All	None	Annual Avg	0.006015	0.002257	Log-linear	Average of ’79-’83 and ’99-’00 air data
All Cause	Sulfate	Pope et al.	2002	53 cities	30+	All	All	None	Annual Avg	0.008964	0.001778	Log-linear	
Cardiopulmonary	$PM_{2.5}$	Pope et al.	2002	61 cities	30+	All	All	None	Annual Avg	0.005733	0.002167	Log-linear	'79-'83 air data
Cardiopulmonary	$PM_{2.5}$	Pope et al.	2002	51 cities	30+	All	All	None	Annual Avg	0.008893	0.002914	Log-linear	Average of ’79-’83 and ’99-’00 air data
Cardiopulmonary	Sulfate	Pope et al.	2002	53 cities	30+	All	All	None	Annual Avg	0.007506	0.002690	Log-linear	
Lung Cancer	$PM_{2.5}$	Pope et al.	2002	61 cities	30+	All	All	None	Annual Avg	0.007881	0.003463	Log-linear	'79-'83 air data
Lung Cancer	$PM_{2.5}$	Pope et al.	2002	51 cities	30+	All	All	None	Annual Avg	0.012663	0.004265	Log-linear	Average of ’79-’83 and ’99-’00 air data
Lung Cancer	Sulfate	Pope et al.	2002	53 cities	30+	All	All	None	Annual Avg	0.013962	0.004048	Log-linear	
Infant	PM_{10}	Woodruff et al.	1997	86 cities	<1	All	All	None	Annual Avg	0.003922	0.001221	Logistic	

F .1 Long-term Mortality

There are two types of exposure to PM that may result in premature mortality. Short-term exposure may result in excess mortality on the same day or within a few days of exposure. Long-term exposure over, say, a year or more, may result in mortality in excess of what it would be if PM levels were generally lower, although the excess mortality that occurs will not necessarily be associated with any particular episode of elevated air pollution levels. In other words, long-term exposure may capture a facet of the association between PM and mortality that is not captured by short-term exposure.

F .1.1 Mortality - Mean, All Cause (Krewski et al., 2000) - Reanalysis of Pope et al. (1995)

The Krewski et al. (2000) reanalysis of Pope et al. (1995) used a Cox proportional hazard model to estimate the impact of long-term PM exposure. The original investigation followed 295,223 individuals[18] ages 30 and over in 50 cities from September 1, 1982 to December 31, 1989, and related their survival to median $PM_{2.5}$ concentrations for 1979 to 1983. Krewski et al. (2000) independently estimated city-specific annual *mean* values from EPA's Inhalable Particle Monitoring Network (IPMN) for the same years (1979-1983). Krewski et al. (2000) followed Pope et al. (1995, Table 2) and reported results for all-cause deaths, lung cancer (ICD-9 code: 162), cardiopulmonary deaths (ICD-9 codes: 401-440 and 460-519), and "all other" deaths,[19] and found that mean $PM_{2.5}$ is significantly related to all-cause and cardiopulmonary mortality. Krewski et al. included only PM, so it is unclear to what extent it may be including the impacts of ozone or other gaseous pollutants.

Pope et al. (1995) is the better of the two published prospective cohort studies: it has a larger population and includes more cities than the prospective cohort study by Dockery et al. (1993). Pope et al.'s study has several further advantages. The population followed in this study was largely Caucasian and middle class, decreasing the likelihood that interlocational differences in premature mortality were due in part to differences in race, socioeconomic status, or related factors. In addition, the PM coefficient in Pope et al. is likely to be biased downward, counteracting a possible upward bias associated with historical air quality trends discussed earlier. One source of this downward bias is the generally healthier and study population, in comparison to poorer minority populations. Krewski et al. (2000, Part II - Table 52) found that educational status was a strong effect modifier of the PM - mortality relationship in both studies, with the strongest effect seen among the less educated. In fact, much of the differences in magnitude of effect between the studies was made up when assessing risk across comparable levels of educational attainment.

Another source of downward bias is that intercity movement of cohort members was not considered in the original study and therefore could not be evaluated in the reanalysis. Migration across study cities would result in exposures of cohort members being more similar than would be indicated by assigning city-specific annual average pollution levels to each member of the cohort. The more intercity migration there is, the more exposure will tend toward an intercity mean. If this is ignored, differences in exposure levels, that are proxied by differences in city-specific annual average PM levels, will be exaggerated, and will result in a downward bias of the PM coefficient (because a given difference in mortality rates is being associated with a larger difference in PM levels than is actually the case).

[18] The total study population was 552,138 in 151 cities, however, only 295,223 individuals resided in 50 cities with fine particle data.

[19] All-cause mortality includes accidents, suicides, homicides and legal interventions. The category "all other" deaths is all-cause mortality less lung cancer and cardiopulmonary deaths.

Single Pollutant Model

The coefficient and standard error are estimated from the relative risk (1.12) and 95% confidence interval (1.06-1.19) associated with a change in *annual mean* $PM_{2.5}$ exposure of 24.5 $\mu g/m^3$ (based on the range from the original ACS study) (Krewski et al., 2000, Part II - Table 31).

Functional Form: Log-linear
Coefficient: 0.004626
Standard Error: 0.001205
Incidence Rate: county-specific annual all cause mortality rate per person ages 30 and older
Population: population of ages 30 and older

F .1.2 Mortality - Median, All Cause (Krewski et al., 2000) - Reanalysis of Pope et al. (1995)

Krewski et al. (2000) performed an analysis of Pope et al. (2000) using independently estimated city-specific annual *median* values as well. Fine particle estimates were obtained from EPA's Inhalable Particle Monitoring Network (IPMN) for the years 1979-1983 for the same 50 cities. Overall, the estimates showed good agreement with the median values used in the original investigation with one exception. The median fine particle concentration for Denver dropped from 16.1 to 7.8 $\mu g/m^3$, resulting in a larger range between the least and most polluted cities and a reduced relative risk. Since the original estimate could not be audited, Denver is included in the subsequent C-R function as there is no reason to believe that the monitoring data is invalid.

Single Pollutant Model

The coefficient and standard error are estimated from the relative risk (1.14) and 95% confidence interval (1.06-1.22) associated with a change in *annual median* $PM_{2.5}$ exposure of 24.5 $\mu g/m^3$ (based on the range from the original ACS study) (Krewski et al., 2000, Part II - Table 31).

Functional Form: Log-linear
Coefficient: 0.005348
Standard Error: 0.001464
Incidence Rate: county-specific annual all cause mortality rate per person ages 30 and older
Population: population of ages 30 and older

F .1.3 Mortality - Median, Random Effects with Regional Adjustment (Krewski et al., 2000) - Reanalysis of Pope et al. (1995)

Krewski et al. (2000) also performed an analysis of Pope et al. (2000) using a random effects model to estimate a regionally-adjusted relative risk. The authors used an indicator variable representing seven regions of the U.S. The regionally-adjusted estimate was comparable with the results from the standard Cox Proportional Hazards Model, which assumes that all observations are statistically independent.

Single Pollutant Model

The coefficient and standard error are estimated from the relative risk (1.16) and 95% confidence interval (0.99-1.37) associated with a change in *annual median* $PM_{2.5}$ exposure of 24.5 μg/m^3 (based on the range from the original ACS study) (Krewski et al., 2000, Part II - Table 46).

Functional Form: Log-linear
Coefficient: 0.006058
Standard Error: 0.003383
Incidence Rate: county-specific annual all cause mortality rate per person ages 30 and older
Population: population of ages 30 and older

F .1.4 Mortality - Median, Random Effects with Independent Cities (Krewski et al., 2000) - Reanalysis of Pope et al. (1995)

Krewski et al. (2000) also performed an analysis of Pope et al. (2000) using a random effects approach to estimate an independent cities model. This approach incorporates between-city variation into second-stage modeling weights, thereby avoiding the assumption of independent observations. However, potential regional patterns in mortality may be overlooked, because the approach assumes that city-specific mortality rates are statistically independent. The independent cities estimate is considerably larger than the standard Cox Proportional Hazards Model, which assumes that all observations are statistically independent.

Single Pollutant Model

The coefficient and standard error are estimated from the relative risk (1.29) and 95% confidence interval (1.12-1.48) associated with a change in *annual median* $PM_{2.5}$ exposure of 24.5 μg/m^3 (based on the range from the original ACS study) (Krewski et al., 2000, Part II - Table 46).

Functional Form: Log-linear
Coefficient: 0.010394
Standard Error: 0.002902
Incidence Rate: county-specific annual all cause mortality rate per person ages 30 and older
Population: population of ages 30 and older

F .1.5 Mortality (Krewski et al., 2000) - Reanalysis of Dockery et al. (1993)

Krewski et al. (2000) performed a validation and replication analysis of Dockery et al. (1993). The originial investigators examined the relationship between PM exposure and mortality in a cohort of 8,111 individuals aged 25 and older, living in six U.S. cities. They surveyed these individuals in 1974-1977 and followed their health status until 1991. While they used a smaller sample of individuals from fewer cities than the study by Pope et al., they used improved exposure estimates, a slightly broader study population (adults aged 25 and older; a higher proportion without a high school education), and a follow-up period nearly twice as long as that of Pope et al. (1995). Krewski et al. (2000, Part II - Table 52) found that educational status was a strong effect modifier of the PM - mortality relationship in both studies, with the strongest effect seen among the less educated. Perhaps because of these differences, Dockery et al. study found a larger effect of PM on premature mortality than that found by Pope et al.

After an audit of the air pollution data, demographic variables, and cohort selection process, Krewski et al. (2000) noted that a small portion of study participants were mistakenly censored early. The following C-R function is based on the risk estimate from the audited data, with the inclusion of those person-years mistakenly censored early.

Single Pollutant Model

The coefficient and standard error are estimated from the relative risk (1.28) and 95% confidence interval (1.10-1.48) associated with a change in *annual mean* $PM_{2.5}$ exposure of 18.6 $\mu g/m^3$ to 29.6 $\mu g/m^3$ (Krewski et al., 2000, Part I - Table 19c).

Functional Form: Log-linear
Coefficient: 0.013272
Standard Error: 0.004070
Incidence Rate: county-specific annual all cause mortality rate per person ages 25 and older
Population: population of ages 25 and older

F .1.6 Mortality, All Cause (Pope et al., 1995)

Pope et al. (1995) used a Cox proportional hazard model to estimate the impact of long-term PM exposure. They followed 295,223 individuals[20] ages 30 and over in 50 cities from September 1, 1982 to December 31, 1989, and related their survival to median $PM_{2.5}$ concentrations for 1979 to 1983. Pope et al. (1995, Table 2) reported results for all-cause deaths, lung cancer (ICD-9 code: 162), cardiopulmonary deaths (ICD-9 codes: 401-440 and 460-519), and "all other" deaths,[21] and found that median $PM_{2.5}$ is significantly related to all-cause and cardiopulmonary mortality. Pope et al. included only PM, so it is unclear to what extent it may be including the impacts of ozone or other gaseous pollutants.

Pope et al. (1995) is the better of the two published prospective cohort studies: it has a larger population and includes more cities than the prospective cohort study by Dockery et al. (1993). Pope et al.'s study has several further advantages. The population followed in this study was largely Caucasian and middle class, decreasing the likelihood that interlocational differences in premature mortality were due in part to differences in race, socioeconomic status, or related factors. In addition, the PM coefficient in Pope et al. is likely to be biased downward, counteracting a possible upward bias associated with historical air quality trends discussed earlier. One source of this downward bias is the generally healthier study population, in comparison to poorer minority populations. Another source of downward bias is that intercity movement of cohort members was not considered in this study. Migration across study cities would result in exposures of cohort members being more similar than would be indicated by assigning city-specific annual average pollution levels to each member of the cohort. The more intercity migration there is, the more exposure will tend toward an intercity mean. If this is ignored, differences in exposure levels, that are proxied by differences in city-specific annual average PM levels, will be exaggerated, and will result in a downward bias of the PM coefficient (because a given difference in mortality rates is being associated with a larger difference in PM levels than is actually the case).

[20] The total study population was 552,138 in 151 cities, however, only 295,223 individuals resided in 50 cities with fine particle data.

[21] All-cause mortality includes accidents, suicides, homicides and legal interventions. The category "all other" deaths is all-cause mortality less lung cancer and cardiopulmonary deaths.

Single Pollutant Model

The coefficient and standard error are estimated from the relative risk (1.17) and 95% confidence interval (1.09-1.26) associated with a change in *annual median* $PM_{2.5}$ exposure of 24.5 μg/m^3 (Pope et al., 1995, Table 2).

Functional Form: Log-linear
Coefficient: 0.006408
Standard Error: 0.001509
Incidence Rate: county-specific annual all cause mortality rate per person ages 30 and older
Population: population of ages 30 and older

F .1.7 Mortality, All Cause (Dockery et al., 1993)

Dockery et al. (1993) examined the relationship between PM exposure and mortality in a cohort of 8,111 individuals aged 25 and older, living in six U.S. cities. They surveyed these individuals in 1974-1977 and followed their health status until 1991. While they used a smaller sample of individuals from fewer cities than the study by Pope et al., they used improved exposure estimates, a slightly broader study population (adults aged 25 and older), and a follow-up period nearly twice as long as that of Pope et al. (1995). Perhaps because of these differences, Dockery et al. study found a larger effect of PM on premature mortality than that found by Pope et al.

Single Pollutant Model

The coefficient and standard error are estimated from the relative risk (1.26) and 95% confidence interval associated (1.08-1.47) with a change in *annual mean* $PM_{2.5}$ exposure of 18.6 μg/m^3 (Dockery et al., 1993, Tables 1 and 5).

Functional Form: Log-linear
Coefficient: 0.012425
Standard Error: 0.004228
Incidence Rate: county-specific annual all cause mortality rate per person ages 25 and older
Population: population of ages 25 and older

F .1.8 Mortality, All Cause (Pope et al., 2002) - Based on ACS Cohort

The Pope et al. (2002) analysis is a longitudinal cohort tracking study that uses the same American Cancer Society (ACS) cohort as the original Pope et al. (1995) study, and the Krewski et al. (2000) reanalysis. Pope et al. (2002) analyzed survival data for the cohort from 1982 through 1998, 9 years longer than the original Pope study. Pope et al. (2002) also obtained $PM_{2.5}$ data in 116 metropolitan areas collected in 1999, and the first three quarters of 2000. This is more metropolitan areas with $PM_{2.5}$ data than was available in the Krewski reanalysis (61 areas), or the original Pope study (50 areas), providing a larger size cohort.

They used a Cox proportional hazard model to estimate the impact of long-term PM exposure using three alternative measures of $PM_{2.5}$ exposure; metropolitan area-wide annual mean PM levels from the beginning of tracking period (’79-’83 PM data, conducted for 61 metropolitan areas with 359,000 individuals), annual mean PM from the end of the tracking period (’99-’00, for 116 areas with 500,000

individuals), and the average annual mean PM levels of the two periods (for 51 metropolitan areas, with 319,000 individuals). PM levels were lower in '99-00 than in '79 - '83 in most cities, with the largest improvements occurring in cities with the highest original levels.

Pope et al. (2002) followed Krewski et al. (2000) and Pope et al. (1995, Table 2) and reported results for all-cause deaths, lung cancer (ICD-9 code: 162), cardiopulmonary deaths (ICD-9 codes: 401-440 and 460-519), and "all other" deaths.[22] Like the earlier studies, Pope et al. (2002) found that mean $PM_{2.5}$ is significantly related to all-cause and cardiopulmonary mortality. In addition, Pope et al. (2002) found a significant relationship with lung cancer mortality, which was not found in the earlier studies. None of the three studies found a significant relationship with "all other" deaths.

Pope et al. (2002) obtained ambient data on gaseous pollutants routinely monitored by EPA during the 1982-1998 observation period, including SO_2, NO_2, CO, and ozone. They did not find significant relationships between NO_2, CO, and ozone and premature mortality, but there were significant relationships between SO_4 (as well as SO_2), and all-cause, cardiopulmonary, lung cancer and "all other" mortality.

$PM_{2.5}$ Function(s)

'79-'83 Exposure

The coefficient and standard error for $PM_{2.5}$ using the '79-'83 PM data are estimated from the relative risk (1.041) and 95% confidence interval (1.008-1.075) associated with a change in *annual mean* exposure of 10.0 $\mu g/m^3$. Pope et al. (2002, Table 2).[23]

Functional Form: Log-linear
Coefficient: 0.004018
Standard Error: 0.001642
Incidence Rate: county-specific annual all cause mortality rate per person ages 30 and older
Population: population of ages 30 and older

Average of '79-'83 and '99-'00 Exposure

The coefficient and standard error for $PM_{2.5}$ using the average of '79-'83 and '99-'00 PM data are estimated from the relative risk (1.062) and 95% confidence interval (1.016-1.110) associated with a change in *annual mean* exposure of 10.0 $\mu g/m^3$. Pope et al. (2002, Table 2).[24]

Functional Form: Log-linear
Coefficient: 0.006015
Standard Error: 0.002257
Incidence Rate: county-specific annual all cause mortality rate per person ages 30 and older
Population: population of ages 30 and older

---

[22] All-cause mortality includes accidents, suicides, homicides and legal interventions. The category "all other" deaths is all-cause mortality less lung cancer and cardiopulmonary deaths.

[23] Note that we used an unpublished, final version of the paper that presents the relative risks with one more significant digit than that found in the published version. We chose to use this extra information to increase the precision of our estimates.

[24] Note that we used an unpublished, final version of the paper that presents the relative risks with one more significant digit than that found in the published version. We chose to use this extra information to increase the precision of our estimates.

SO_4 Function(s)

The coefficient and standard error for SO_4 using '80-'81 data are estimated from the relative risk (1.060) and 95% confidence interval (1.036-1.084) associated with a change in *annual mean* exposure of 6.5 μg/m^3. The relative risk and confidence interval were provided by C.A. Pope III over the phone.

Functional Form: Log-linear
Coefficient: 0.008964
Standard Error: 0.001778
Incidence Rate: county-specific annual all cause mortality rate per person ages 30 and older
Population: population of ages 30 and older

F .1.9 Mortality, Cardiopulmonary (Pope et al., 2002) - Based on ACS Cohort

Pope et al. (2002) followed Krewski et al. (2000) and Pope et al. (1995, Table 2) and reported results for all-cause deaths, lung cancer (ICD-9 code: 162), cardiopulmonary deaths (ICD-9 codes: 401-440 and 460-519), and "all other" deaths.[25] Like the earlier studies, Pope et al. (2002) found that mean $PM_{2.5}$ and SO_4 (as well as SO_2) is significantly related to all-cause and cardiopulmonary mortality. In addition, Pope et al. (2002) found a significant relationship with lung cancer mortality, which was not found in the earlier studies. None of the three studies found a significant relationship with "all other" deaths.

$PM_{2.5}$ Function(s)

'79-'83 Exposure

The coefficient and standard error for $PM_{2.5}$ using the '79-'83 PM data are estimated from the relative risk (1.059) and 95% confidence interval (1.015-1.105) associated with a change in *annual mean* exposure of 10.0 μg/m^3. Pope et al. (2002, Table 2).[26]

Functional Form: Log-linear
Coefficient: 0.005733
Standard Error: 0.002167
Incidence Rate: county-specific annual cardiopulmonary mortality rate (ICD codes 401-440, 460-519) per person ages 30 and older
Population: population of ages 30 and older

[25] All-cause mortality includes accidents, suicides, homicides and legal interventions. The category "all other" deaths is all-cause mortality less lung cancer and cardiopulmonary deaths.

[26] Note that we used an unpublished, final version of the paper that presents the relative risks with one more significant digit than that found in the published version. We chose to use this extra information to increase the precision of our estimates.

Average of '79-'83 and '99-'00 Exposure

The coefficient and standard error for $PM_{2.5}$ using the average of '79-'83 and '99-'00 PM data are estimated from the relative risk (1.093) and 95% confidence interval (1.033-1.158) associated with a change in *annual mean* exposure of 10.0 μg/m^3. Pope et al. (2002, Table 2).[27]

Functional Form: Log-linear
Coefficient: 0.008893
Standard Error: 0.002914
Incidence Rate: county-specific annual cardiopulmonary mortality rate (ICD codes 401-440, 460-519) per person ages 30 and older
Population: population of ages 30 and older

SO_4 Function(s)

The coefficient and standard error for SO_4 using '80-'81 data are estimated from the relative risk (1.050) and 95% confidence interval (1.015-1.087) associated with a change in *annual mean* exposure of 6.5 μg/m^3. The relative risk and confidence interval were provided by C.A. Pope III over the phone.

Functional Form: Log-linear
Coefficient: 0.007506
Standard Error: 0.002690
Incidence Rate: county-specific annual cardiopulmonary mortality rate (ICD codes 401-440, 460-519) per person ages 30 and older
Population: population of ages 30 and older

F .1.10 Mortality, Lung Cancer (Pope et al., 2002) - Based on ACS Cohort

Pope et al. (2002) followed Krewski et al. (2000) and Pope et al. (1995, Table 2) and reported results for all-cause deaths, lung cancer (ICD-9 code: 162), cardiopulmonary deaths (ICD-9 codes: 401-440 and 460-519), and "all other" deaths.[28] Like the earlier studies, Pope et al. (2002) found that mean $PM_{2.5}$ SO_4 (as well as SO_2) is significantly related to all-cause and cardiopulmonary mortality. In addition, Pope et al. (2002) found a significant relationship with lung cancer mortality, which was not found in the earlier studies. None of the three studies found a significant relationship with "all other" deaths.

[27] Note that we used an unpublished, final version of the paper that presents the relative risks with one more significant digit than that found in the published version. We chose to use this extra information to increase the precision of our estimates.

[28] All-cause mortality includes accidents, suicides, homicides and legal interventions. The category "all other" deaths is all-cause mortality less lung cancer and cardiopulmonary deaths.

$PM_{2.5}$ Function(s)

'79-'83 Exposure

The coefficient and standard error for $PM_{2.5}$ using the '79-'83 PM data are estimated from the relative risk (1.082) and 95% confidence interval (1.011-1.158) associated with a change in *annual mean* exposure of 10.0 μg/m^3. Pope et al. (2002, Table 2).[29]

Functional Form: Log-linear
Coefficient: 0.007881
Standard Error: 0.003463
Incidence Rate: county-specific annual lung cancer mortality rate (ICD code 162) per person ages 30 and older
Population: population of ages 30 and older

Average of '79-'83 and '99-'00 Exposure

The coefficient and standard error for $PM_{2.5}$ using the average of '79-'83 and '99-'00 PM data are estimated from the relative risk (1.135) and 95% confidence interval (1.044-1.234) associated with a change in *annual mean* exposure of 10.0 μg/m^3. Pope et al. (2002, Table 2).[30]

Functional Form: Log-linear
Coefficient: 0.012663
Standard Error: 0.004265
Incidence Rate: county-specific annual lung cancer mortality rate (ICD code 162) per person ages 30 and older
Population: population of ages 30 and older

SO_4 Function(s)

The coefficient and standard error for SO_4 using '80-'81 data are estimated from the relative risk (1.095) and 95% confidence interval (1.040-1.153) associated with a change in *annual mean* exposure of 6.5 μg/m^3. The relative risk and confidence interval were provided by C.A. Pope III over the phone.

Functional Form: Log-linear
Coefficient: 0.013962
Standard Error: 0.004048
Incidence Rate: county-specific annual lung cancer mortality rate (ICD code 162) per person ages 30 and older
Population: population of ages 30 and older

[29] Note that we used an unpublished, final version of the paper that presents the relative risks with one more significant digit than that found in the published version. We chose to use this extra information to increase the precision of our estimates.

[30] Note that we used an unpublished, final version of the paper that presents the relative risks with one more significant digit than that found in the published version. We chose to use this extra information to increase the precision of our estimates.

F .1.11 Infant Mortality (Woodruff et al., 1997)

In a study of four million infants in 86 U.S. metropolitan areas conducted from 1989 to 1991, Woodruff et al. (1997) found a significant link between PM_{10} exposure in the first two months of an infant's life with the probability of dying between the ages of 28 days and 364 days. PM_{10} exposure was significant for all-cause mortality. PM_{10} was also significant for respiratory mortality in average birth-weight infants, but not low birth-weight infants.

In addition to the work by Woodruff et al., work in Mexico City (Loomis et al., 1999), the Czech Republic (Bobak and Leon, 1992), Sao Paulo (Saldiva et al., 1994; Pereira et al., 1998), and Beijing (Wang et al., 1997) provides additional evidence that particulate levels are significantly related to infant or child mortality, low birth weight or intrauterine mortality.

Conceptually, neonatal or child mortality could be added to the premature mortality predicted by Pope et al. (1995), because the Pope function covers only the population over 30 years old.[31] However, the EPA Science Advisory Board recently advised the Agency not to include post-neonatal mortality in this analysis because the study is of a new endpoint and the results have not been replicated in other studies (U.S. EPA, 1999a, p. 12). The estimated avoided incidences of neonatal mortality are estimated and presented as a sensitivity analysis, and are not included in the primary analysis.

Single Pollutant Model

The coefficient and standard error are based on the odds ratio (1.04) and 95% confidence interval (1.02-1.07) associated with a 10 $\mu g/m^3$ change in PM_{10} (Woodruff et al., 1997, Table 3).

Functional Form: Logistic
Coefficient: 0.003922
Standard Error: 0.001221
Incidence Rate: county-specific annual postneonatal[32] infant deaths per infant under the age of one
Population: population of infants under one year old

[31] Predicted neonatal mortality could not be added to the premature mortality predicted by the daily (short-term exposure) mortality studies, however, because these studies cover all ages.

[32] Post-neonatal refers to infants that are 28 days to 364 days old.

Appendix F. Particulate Matter C-R Functions

Exhibit F-2. Concentration-Response (C-R) Functions for Particulate Matter and Short-Term Mortality

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Beta	Std Error	Functional Form	Notes
Non-Accidental	$PM_{2.5}$	Fairley	2003	Santa Clara County	All	All	All	None	24-hr avg	0.003143	0.001283	Log-linear	Reanalysis of Fairley, 1999
Non-Accidental	$PM_{2.5}$	Fairley	2003	Santa Clara County	All	All	All	O_3	24-hr avg	0.003404	0.001300	Log-linear	Reanalysis of Fairley, 1999
Non-Accidental	$PM_{2.5}$	Ito	2003	Detroit, MI	All	All	All	None	24-hr avg	0.000740	0.000752	Log-linear	Reanalysis of Lippmann et al., 2000
Non-Accidental	$PM_{2.5}$	Klemm and Mason	2003	6 Cities	All	All	All	None	24-hr avg	0.001193	0.000202	Log-linear	Reanalysis of Klemm and Mason et al., 2000
Non-Accidental	$PM_{2.5}$	Moolgavkar	2003	Los Angeles, CA	All	All	All	None	24-hr avg	0.000588	0.000300	Log-linear	Reanalysis of Moolgavkar, 2000a,b,c
Non-Accidental	$PM_{2.5}$	Schwartz et al.	1996	6 cities	All	All	All	None	24-hr avg	0.001433	0.000129	Log-linear	Standard and Lag adjusted[2]
Non-Accidental	$PM_{2.5}$	Schwartz	2003	6 cities	All	All	All	None	24-hr avg	0.00137	0.00020	Log-linear	Standard and Lag adjusted[2]
Chronic Lung	$PM_{2.5}$	Schwartz et al.	1996	6 cities	All	All	All	None	24-hr avg	0.003247	--	Log-linear	Lag adjusted[2]

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.
2. Refer to the study summaries below for a discussion of the lag adjustment used for these functions.

F .2 Short-term Mortality

Short-term mortality studies are those that typically link daily air pollution levels with daily changes in mortality rates.

F .2.1 Short-Term Mortality, Non-Accidental (Fairley, 2003)

Using data from 1989-1996 in Santa Clara County, California, Fairley et al. (1999) examined the relationship between daily non-accidental mortality and fluctuations in a variety of pollutants, including $PM_{2.5}$, coarse PM_{10} (i.e., $PM_{2.5-10}$), nitrate (NO_3), SO_4, coefficient of haze (COH), ozone, CO, and NO_2. They reported that $PM_{2.5}$ and NO_3 were significant in single-pollutant models, as well as two-pollutant models. $PM_{2.5}$ was only insignificant when paired with PM_{10} and NO_3, and NO_3 was only insignificant when paired with $PM_{2.5}$. The other pollutants were insignificant when paired with either $PM_{2.5}$ or NO_3.

The analysis by Fairly et al. (1999) relied on a generalized additive model based on the Splus software. Because of potential bias from using Splus, Fairley (2003) conducted a reanalysis, and reported that the conclusions of the original study were unchanged. Both $PM_{2.5}$ and NO_3 appear significantly related to non-accidental mortality.

Single Pollutant Model

The coefficient and standard error for $PM_{2.5}$ are estimated from the relative risk (1.092) and 95% confidence interval (1.018-1.172) for a 28 $\mu g/m^3$ increase in $PM_{2.5}$ in the 0-day lag GAM stringent ('New GAM') model (Fairley, 2003, Table 1a).

Functional Form: Log-linear
Coefficient: 0.003143
Standard Error: 0.001283
Incidence Rate: county-specific annual daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

Multipollutant Model ($PM_{2.5}$ and ozone)

In a model with 8-hour averaged ozone, the coefficient and standard error for $PM_{2.5}$ are estimated from the relative risk (1.100) and 95% confidence interval (1.024-1.181) for a 28 $\mu g/m^3$ increase in $PM_{2.5}$ in the 0-day lag GAM stringent ('New GAM') model (Fairley, 2003, Table 1b).

Functional Form: Log-linear
Coefficient: 0.003404
Standard Error: 0.001300
Incidence Rate: county-specific annual daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

F .2.2 Short-Term Mortality, Non-Accidental (Ito, 2003)

Ito (2003) reanalyzed a study by Lippmann et al. (2000) who examined the associations between PM components and daily mortality and elderly hospital admissions in Detroit, Michigan. The reanalysis

by Ito reported that more generalized additive models with stringent convergence criteria and generalized linear models resulted in smaller relative risk estimates.

Single Pollutant Model

The coefficient and standard error for $PM_{2.5}$ are estimated from the relative risk (1.027) and 95% confidence interval (0.974-1.083) for a 36 $\mu g/m^3$ increase in $PM_{2.5}$ in the 3-day lag GAM stringent model (Ito, 2003, Table 4).

Functional Form: Log-linear
Coefficient: 0.000740
Standard Error: 0.000752
Incidence Rate: region-specific annual daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

F .2.3 Short-Term Mortality, Non-Accidental (Klemm and Mason, 2003)

Klemm and Mason (2003) reanalyzed a prior study by Klemm and Mason (2000), who conducted a replication of work by Schwartz et al. (1996). In the updated work using more stringent convergence criteria and generalized linear models, Klemm and Mason (2003) reported a generally smaller relationship between daily $PM_{2.5}$ levels and premature mortality.

Single Pollutant Model

The coefficient and standard error for $PM_{2.5}$ are estimated from the relative risk (1.012) and 95% confidence interval (0.008-1.016) for a 10 $\mu g/m^3$ increase in $PM_{2.5}$ in the 0-day lag GAM stringent ('GAM 2002') model (Klemm and Mason, 2003, Table 1).

Functional Form: Log-linear
Coefficient: 0.001193
Standard Error: 0.000202
Incidence Rate: county-specific annual daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

F .2.4 Short-Term Mortality, Non-Accidental (Moolgavkar, 2003)

Moolgavkar (2003) reanalyzed a study by Moolgavkar (2000b) who examined the relationships between daily mortality and hospital admissions in Los Angeles and Cook Counties. The reanalysis by Moolgavkar reported that more generalized additive models with stringent convergence criteria and generalized linear models generally resulted in smaller relative risk estimates, and that gases such as CO were often more closely associated with health endpoints than particulate matter.

Single Pollutant Model

The coefficient and standard error for $PM_{2.5}$ are estimated from the relative risk (1.0059) and the t-statistic (1.96) for a 10 $\mu g/m^3$ increase in $PM_{2.5}$ in the 1-day lag GAM-30df stringent (10^{-8}) model (Moolgavkar, 2003, Table 1)

Functional Form: Log-linear
Coefficient: 0.000588
Standard Error: 0.000300
Incidence Rate: county-specific annual daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

F .2.5 Short-Term Mortality, Non-Accidental (Schwartz et al., 1996)

Schwartz et al. (1996) pooled the results from six cities in the U.S. and found a significant relationship between daily $PM_{2.5}$ concentration and non-accidental mortality.[33] Abt Associates Inc. (1996b, p. 52) used the six $PM_{2.5}$ relative risks reported by Schwartz et al. in a three-step procedure to estimate a pooled $PM_{2.5}$ coefficient and its standard error. The first step estimates a random-effects pooled estimate of β; the second step uses an "empirical Bayes" procedure to reestimate the β for each study as a weighted average of the β reported for that location and the random effects pooled estimate; the third step estimates the underlying distribution of β, and uses a Monte Carlo procedure to estimate the standard error (Abt Associates Inc., 1996a, p. 65).

Single Pollutant Model

The C-R function to estimate the change in mortality associated with daily changes in $PM_{2.5}$ is:

Functional Form: Log-linear
Coefficient: 0.001433 (Abt Associates Inc., 1996a, Exhibit 7.2)
Standard Error: 0.000129 (Abt Associates Inc., 1996a, Exhibit 7.2)
Incidence Rate: county-specific annual daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

Single Pollutant Model, Lag Adjusted

Recent studies have found that an increase in PM levels on a given day can elevate mortality for several days following the exposure (Samet et al., 2000; Schwartz, 2000b). These studies have reported the results of distributed lag models for the relationship between PM_{10} and daily mortality. Schwartz (2000b) examined the relationship between PM_{10} and daily mortality and reported results both for a single day lag model and an unconstrained distributed lag model. The unconstrained distributed lag model coefficient estimate is 0.0012818 and the single-lag model coefficient estimate is 0.0006479. A distributed lag adjustment factor can be constructed as the ratio of the estimated coefficient from the unconstrained distributed lag model to the estimated coefficient from the single-lag model reported in Schwartz (2000). The ratio of these estimates is 1.9784. In order to estimate the full impact of daily PM levels on daily mortality, we applied this ratio to the coefficient obtained from Schwartz et al. (1996) for the association between $PM_{2.5}$ and daily mortality.

In applying the ratio derived from a PM_{10} study to $PM_{2.5}$, we assume that the same relationship between the distributed lag and single day estimates would hold for $PM_{2.5}$. Effect estimates for the PM_{10}-daily mortality relationship tend to be lower in magnitude than for $PM_{2.5}$, because fine particles are believed to be more closely associated with mortality than the coarse fraction of PM. If most of the increase in

[33] Schwartz et al. (1996, p. 929) defined non-accidental mortality as all-cause mortality less deaths due to accidents and other external causes (ICD-9 codes: 800-999). Other external causes includes suicide, homicide, and legal intervention (National Center for Health Statistics, 1994).

mortality is expected to be associated with the fine fraction of PM_{10}, then it is reasonable to assume that the same proportional increase in risk would be observed if a distributed lag model were applied to the $PM_{2.5}$ data.

The distributed lag model coefficient is estimated by multiplying the distributed lag adjustment factor of 1.9784 with the pooled $PM_{2.5}$ coefficient. Note that the distributed lag adjustment C-R function is only run for the point estimate, as the standard error of this modified coefficient has not been estimated.

Functional Form: Log-linear
Coefficient: 0.001433
Lag Adjustment: 1.9784
Incidence Rate: county-specific annual daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

F .2.6 Short-Term Mortality, Non-Accidental (Schwartz, 2003)

Schwartz et al. (1996) pooled the results from six cities in the U.S. and found a significant relationship between daily $PM_{2.5}$ concentration and non-accidental mortality.[34] In a reanalysis of this work, Schwartz (2003) reported that the coefficients are somewhat smaller and less stable, but that the overall relationship between $PM_{2.5}$ and mortality remained unchanged.

Single Pollutant Model

The coefficient and standard error are provided Schwartz (2003, Table 1) (see: combined estimate, mean of lag 0 and 1, New Convergence – GAM stringent).

Functional Form: Log-linear
Coefficient: 0.00137
Standard Error: 0.0002
Incidence Rate: county-specific annual daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

Single Pollutant Model, Lag Adjusted

As noted in the section on the Schwartz et al. (1996) C-R function, we have added a distributed lag adjustment factor. The distributed lag model coefficient is estimated by multiplying the distributed lag adjustment factor of 1.9784 with the $PM_{2.5}$ coefficient. Note that the distributed lag adjustment C-R function is only run for the point estimate, as the standard error of this modified coefficient has not been estimated.

Functional Form: Log-linear
Coefficient: 0.00137 (Schwartz, 2003, Table 1)
Lag Adjustment: 1.9784
Incidence Rate: county-specific annual daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

[34] Schwartz et al. (1996, p. 929) defined non-accidental mortality as all-cause mortality less deaths due to accidents and other external causes (ICD-9 codes: 800-999). Other external causes includes suicide, homicide, and legal intervention (National Center for Health Statistics, 1994).

F .2.7 Short-Term Mortality, Chronic Lung Disease - Lag Adjusted (Schwartz et al., 1996)

Schwartz et al. (1996) evaluated the relationship between daily $PM_{2.5}$ levels and short-term mortality in six U.S. cities. Schwartz pooled results across the six cities and found statistically significant associations between daily $PM_{2.5}$ levels and non-accidental mortality (ICD codes <800), along with mortality for ischemic heart disease (ICD codes 410-414), COPD (ICD codes 490-496), and pneumonia (ICD codes 480-486). A smaller association was found for PM_{10} and no significant associations were reported for $PM_{10-2.5}$. The C-R function for chronic lung disease mortality is based on the results of a single pollutant model using a two-day average of $PM_{2.5}$ (Schwartz et al., 1996, Table 7). In order to estimate the impact of daily $PM_{2.5}$ levels on daily mortality if a distributed lag model had been fit, the $PM_{2.5}$ coefficient is adjusted as described below.

Single Pollutant Model

The $PM_{2.5}$ coefficient is based on a reported 3.3% increase in COPD mortality associated with a 10 $\mu g/m^3$ change in two-day average $PM_{2.5}$ levels (Schwartz et al., 1996, Table 7). This coefficient (0.003247) is then multiplied by the distributed lag adjustment factor of 1.9784 to estimate a distributed lag model coefficient.

Functional Form: Log-linear
Coefficient: 0.003247
Lag Adjustment: 1.9784
Incidence Rate: county-specific annual daily chronic lung disease mortality rate (ICD codes 490-496)
Population: population of all ages

Appendix F. Particulate Matter C-R Functions

Exhibit F-3. Concentration-Response (C-R) Functions for Particulate Matter and Chronic Illness

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time	Beta	Std Error	Functional Form
Chronic Bronchitis	$PM_{2.5}$	Abbey et al.	1995	SF, SD, South Coast Air Basin	27+	All	All	None	Annual Avg	0.0137	0.00680	Logistic
Chronic Bronchitis, Reversals	$PM_{2.5}$	Abbey et al.	1995	SF, SD, South Coast Air Basin	27+	All	All	None	Annual Avg	0.0137	0.00680	Logistic
Chronic Bronchitis	PM_{10}	Schwartz	1993	53 cities	30+	All	All	None	Annual Avg	0.0123	0.00434	Logistic

F .3 Chronic Illness

Schwartz (1993) and Abbey et al. (1993; 1995c) provide evidence that PM exposure over a number of years gives rise to the development of chronic bronchitis in the U.S., and a recent study by McDonnell et al. (1999) provides evidence that ozone exposure is linked to the development of asthma in adults. These results are consistent with research that has found chronic exposure to pollutants leads to declining pulmonary functioning (Detels et al., 1991; Ackermann-Liebrich et al., 1997; Abbey et al., 1998).[35]

F .3.1 Chronic Bronchitis (Abbey et al., 1995c, California)

Abbey et al. (1995c) examined the relationship between estimated $PM_{2.5}$ (annual mean from 1966 to 1977), PM_{10} (annual mean from 1973 to 1977) and TSP (annual mean from 1973 to 1977) and the same chronic respiratory symptoms in a sample population of 1,868 Californian Seventh Day Adventists. The initial survey was conducted in 1977 and the final survey in 1987. To ensure a better estimate of exposure, the study participants had to have been living in the same area for an extended period of time. In single-pollutant models, there was a statistically significant $PM_{2.5}$ relationship with development of chronic bronchitis, but not for AOD or asthma; PM_{10} was significantly associated with chronic bronchitis and AOD; and TSP was significantly associated with all cases of all three chronic symptoms. Other pollutants were not examined. The C-R function is based on the results of the single pollutant model presented in Table 2.

Single Pollutant Model (Chronic Bronchitis)

The estimated coefficient (0.0137) is presented for a one $\mu g/m^3$ change in $PM_{2.5}$ (Abbey et al., 1995c, Table 2). The standard error is calculated from the reported relative risk (1.81) and 95% confidence interval (0.98-3.25) for a 45 $\mu g/m^3$ change in $PM_{2.5}$ (Abbey et al., 1995c, Table 2).

Functional Form: Logistic
Coefficient: 0.0137
Standard Error: 0.00680
Incidence Rate: annual bronchitis incidence rate per person (Abbey et al., 1993, Table 3) = 0.00378
Population: population of ages 27 and older[36] without chronic bronchitis = 95.57%[37] of population 27+

Single Pollutant Model (Chronic Bronchitis, Reversals)

In developing the C-R function for chronic bronchitis, it is necessary to estimate its annual incidence rate. The annual incidence rate is derived by taking the number of new cases (234), dividing by the number of individuals in the sample (3,310), as reported by Abbey et al.(1993, Table 3), dividing by the ten years covered in the sample, and then multiplying by one minus the reversal rate.[38] Reversals refer to

[35] There are a limited number of studies that have estimated the impact of air pollution on chronic bronchitis. An important hindrance is the lack of health data and the associated air pollution levels over a number of years.

[36] Using the same data set, Abbey et al. (1995a, p. 140) reported that the respondents in 1977 ranged in age from 27 to 95.

[37] The American Lung Association (2002b, Table 4) reports a chronic bronchitis prevalence rate for ages 18 and over of 4.43% (American Lung Association, 2002b, Table 4).

[38] The percentage of reversals is estimated to be 46.6% based on Abbey et al. (1995a, Table 1).

those cases of chronic bronchitis that were reported at the start of the Abbey et al. survey, but were subsequently not reported at the end of the survey. Since we assume that chronic bronchitis is a permanent condition, we subtract these reversals from the C-R function for chronic bronchitis. Nevertheless, reversals may likely represent a real effect that should be included in an analysis.

The estimated coefficient (0.0137) is presented for a one $\mu g/m^3$ change in $PM_{2.5}$ (Abbey et al., 1995c, Table 2). The standard error is calculated from the reported relative risk (1.81) and 95% confidence interval (0.98-3.25) for a 45 $\mu g/m^3$ change in $PM_{2.5}$ (Abbey et al., 1995c, Table 2).

Functional Form: Logistic
Coefficient: 0.0137
Standard Error: 0.00680
Incidence Rate: annual bronchitis incidence rate per person for chronich bronchitis that eventually resolves itself. Based on the percentage of reversals (46.6%) from Abbey et al. (1995a, Table 1) and chronic bronchitis cases from (Abbey et al., 1993, Table 3) = 0.00325
Population: population of ages 27 and older[39] without chronic bronchitis = 95.57%[40] of population 27+

F .3.2 Chronic Bronchitis (Schwartz, 1993)

Schwartz (1993) examined survey data collected from 3,874 adults ranging in age from 30 to 74, and living in 53 urban areas in the U.S. The survey was conducted between 1974 and 1975, as part of the National Health and Nutrition Examination Survey, and is representative of the non-institutionalized U.S. population. Schwartz (1993, Table 3) reported chronic bronchitis prevalence rates in the study population by age, race, and gender. Non-white males under 52 years old had the lowest rate (1.7%) and white males 52 years and older had the highest rate (9.3%). The study examined the relationship between the prevalence of reported chronic bronchitis, asthma, shortness of breath (dyspnea) and respiratory illness[41], and the annual levels of TSP, collected in the year prior to the survey (TSP was the only pollutant examined in this study). TSP was significantly related to the prevalence of chronic bronchitis, and marginally significant for respiratory illness. No effect was found for asthma or dyspnea. The C-R function for PM_{10} is estimated from the results of the single pollutant model reported for TSP.

Single Pollutant Model

The estimated coefficient is based on the odds ratio (1.07) associated with 10 $\mu g/m^3$ change in TSP (Schwartz, 1993, p. 9). Assuming that PM_{10} is 55 percent of TSP[42] and that particulates greater than ten micrometers are harmless, the coefficient is calculated by dividing the TSP coefficient by 0.55. The standard error for the coefficient is calculated from the 95% confidence interval for the odds ratio (1.02 to 1.12) (Schwartz, 1993, p. 9).

---

[39] Using the same data set, Abbey et al. (1995a, p. 140) reported that the respondents in 1977 ranged in age from 27 to 95.

[40] The American Lung Association (2002b, Table 4) reports a chronic bronchitis prevalence rate for ages 18 and over of 4.43% (American Lung Association, 2002b, Table 4).

[41] Respiratory illness defined as a significant condition, coded by an examining physician as ICD-8 code 460-519.

[42] The conversion of TSP to PM_{10} is from ESEERCO (1994, p. V-5), who cited studies by EPA (1986) and the California Air Resources Board (1982).

Schwartz (1993) examined the *prevalence* of chronic bronchitis, not its *incidence*. To use Schwartz's study and still estimate the change in incidence, there are at least two possible approaches. The first is to simply assume that it is appropriate to use the baseline *incidence* of chronic bronchitis in a C-R function with the estimated coefficient from Schwartz's study, to directly estimate the change in incidence. The second is to estimate the percentage change in the prevalence rate for chronic bronchitis using the estimated coefficient from Schwartz's study in a C-R function, and then to assume that this percentage change applies to a baseline incidence rate obtained from another source. (That is, if the prevalence declines by 25 percent with a drop in PM, then baseline incidence drops by 25 percent with the same drop in PM.) This analysis is using the latter approach, and estimates a percentage change in prevalence which is then applied to a baseline incidence rate. The scaling factor used in the C-R function is the ratio of chronic bronchitis incidence rate (estimated from Abbey et al. (1993)) to chronic bronchitis prevalence rate (estimated from American Lung Association (2002b, Table 4)).

Functional Form: Logistic
Coefficient: 0.0123
Standard Error: 0.00434
Prevalence Rate: annual chronic bronchitis prevalence rate per person (American Lung Association, 2002b, Table 4) = 0.0443
Population: population of ages 30 and older without chronic bronchitis = 95.57%[43] of population 30+
Adjustment Factor: ratio of chronic bronchitis incidence to chronic bronchitis prevalence = 0.00378/0.0443 = 0.085 (Abbey et al., 1993, Table 3; American Lung Association, 2002b, Table 4)

[43] The American Lung Association (2002b, Table 4) reports a chronic bronchitis prevalence rate for ages 18 and over of 4.43% (American Lung Association, 2002b, Table 4).

Appendix F. Particulate Matter C-R Functions

Exhibit F-4. Concentration-Response (C-R) Functions for Particulate Matter and Hospital Admissions

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time	Beta	Std Error	Functional Form	Notes
All Respiratory	$PM_{2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	None	24-hr avg	0.003303	0.001004	Log-linear	
All Respiratory	$PM_{2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	O_3	24-hr avg	0.002422	0.001039	Log-linear	
All Respiratory	$PM_{2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	NO_2, O_3, SO_2	24-hr avg	-0.000091	0.000910	Log-linear	
All Respiratory	$PM_{10-2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	None	24-hr avg	0.004787	0.001404	Log-linear	
All Respiratory	$PM_{10-2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	O_3	24-hr avg	0.004169	0.001371	Log-linear	
All Respiratory	$PM_{10-2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	NO_2, O_3, SO_2	24-hr avg	0.001469	0.001791	Log-linear	
All Respiratory	PM_{10}	Burnett et al.	1997	Toronto, CAN	All	All	All	None	24-hr avg	0.002074	0.000607	Log-linear	
All Respiratory	PM_{10}	Burnett et al.	1997	Toronto, CAN	All	All	All	O_3	24-hr avg	0.001870	0.000592	Log-linear	
All Respiratory	PM_{10}	Burnett et al.	1997	Toronto, CAN	All	All	All	NO_2, O_3, SO_2	24-hr avg	0.000280	0.000778	Log-linear	
All Respiratory	$PM_{2.5}$	Burnett et al.	2001	Toronto, CAN	<2	All	All	None	24-hr avg	0.008150	0.002477	Log-linear	
All Respiratory	$PM_{2.5}$	Burnett et al.	2001	Toronto, CAN	<2	All	All	O_3	24-hr avg	0.000772	0.003218	Log-linear	
All Respiratory	$PM_{10-2.5}$	Burnett et al.	2001	Toronto, CAN	<2	All	All	None	24-hr avg	0.010374	0.002752	Log-linear	
All Respiratory	$PM_{10-2.5}$	Burnett et al.	2001	Toronto, CAN	<2	All	All	O_3	24-hr avg	0.002717	0.003774	Log-linear	
All Respiratory	PM_{10}	Schwartz	1995	New Haven, CT	65+	All	All	None	24-hr avg	0.001165	0.000624	Log-linear	
All Respiratory	PM_{10}	Schwartz	1995	New Haven, CT	65+	All	All	O_3	24-hr avg	0.001724	0.000930	Log-linear	
All Respiratory	PM_{10}	Schwartz	1995	Tacoma, WA	65+	All	All	None	24-hr avg	0.001906	0.000650	Log-linear	
All Respiratory	PM_{10}	Schwartz	1995	Tacoma, WA	65+	All	All	O_3	24-hr avg	0.002267	0.001455	Log-linear	
All Respiratory	$PM_{2.5}$	Thurston et al.	1994	Toronto, CAN	All	All	All	None	24-hr avg	0.0828	0.0367	Linear	
All Respiratory	$PM_{2.5}$	Thurston et al.	1994	Toronto, CAN	All	All	All	O_3	24-hr avg	0.0434	0.0429	Linear	
All Respiratory	$PM_{10-2.5}$	Thurston et al.	1994	Toronto, CAN	All	All	All	None	24-hr avg	0.1228	0.0895	Linear	
All Respiratory	PM_{10}	Thurston et al.	1994	Toronto, CAN	All	All	All	None	24-hr avg	0.0642	0.0290	Linear	
All Respiratory	PM_{10}	Thurston et al.	1994	Toronto, CAN	All	All	All	O_3	24-hr avg	0.0339	0.0344	Linear	
Asthma	$PM_{2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.002499	0.000776	Log-linear	
Asthma	$PM_{10-2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.004194	0.000999	Log-linear	

Appendix F. Particulate Matter C-R Functions

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time	Beta	Std Error	Functional Form	Notes
Asthma	$PM_{10-2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	CO, O_3	24-hr avg	0.003215	0.001058	Log-linear	
Asthma	PM_{10}	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.001701	0.000502	Log-linear	
Asthma	$PM_{2.5}$	Lin et al.	2002	Toronto, CAN	6-12	All	Male	None	24-hr avg	-0.004389	0.003130	Log-linear	
Asthma	$PM_{2.5}$	Lin et al.	2002	Toronto, CAN	6-12	All	Male	CO, NO_2, O_3, SO_2	24-hr avg	-0.006653	0.003779	Log-linear	
Asthma	$PM_{2.5}$	Lin et al.	2002	Toronto, CAN	6-12	All	Female	None	24-hr avg	0.006265	0.008377	Log-linear	
Asthma	$PM_{2.5}$	Lin et al.	2002	Toronto, CAN	6-12	All	Female	CO, NO_2, O_3, SO_2	24-hr avg	-0.002172	0.005001	Log-linear	
Asthma	$PM_{10-2.5}$	Lin et al.	2002	Toronto, CAN	6-12	All	Male	None	24-hr avg	0.013492	0.004346	Log-linear	
Asthma	$PM_{10-2.5}$	Lin et al.	2002	Toronto, CAN	6-12	All	Male	CO, NO_2, O_3, SO_2	24-hr avg	0.016638	0.005007	Log-linear	
Asthma	$PM_{10-2.5}$	Lin et al.	2002	Toronto, CAN	6-12	All	Female	None	24-hr avg	0.021705	0.005583	Log-linear	
Asthma	$PM_{10-2.5}$	Lin et al.	2002	Toronto, CAN	6-12	All	Female	CO, NO_2, O_3, SO_2	24-hr avg	0.016638	0.006836	Log-linear	
Asthma	PM_{10}	Lin et al.	2002	Toronto, CAN	6-12	All	Male	None	24-hr avg	0.000672	0.002211	Log-linear	
Asthma	PM_{10}	Lin et al.	2002	Toronto, CAN	6-12	All	Male	CO, NO_2, O_3, SO_2	24-hr avg	0.001338	0.002865	Log-linear	
Asthma	PM_{10}	Lin et al.	2002	Toronto, CAN	6-12	All	Female	None	24-hr avg	0.004572	0.002906	Log-linear	
Asthma	PM_{10}	Lin et al.	2002	Toronto, CAN	6-12	All	Female	CO, NO_2, O_3, SO_2	24-hr avg	0.001997	0.003660	Log-linear	
Asthma	$PM_{2.5}$	Sheppard et al.	2003	Seattle, WA	<65	All	All	None	24-hr avg	0.003324	0.001045	Log-linear	Reanalysis of Sheppard et al., 1999
Asthma	$PM_{2.5}$	Sheppard et al.	1999	Seattle, WA	<65	All	All	CO	24-hr avg	0.002505	0.001045	Log-linear	
Asthma	$PM_{10-2.5}$	Sheppard et al.	1999	Seattle, WA	<65	All	All	None	24-hr avg	0.004217	0.001583	Log-linear	
Asthma	PM_{10}	Sheppard et al.	1999	Seattle, WA	<65	All	All	None	24-hr avg	0.002568	0.000767	Log-linear	
Asthma	$PM_{2.5}$	Thurston et al.	1994	Toronto, CAN	All	All	All	None	24-hr avg	0.0334	0.0241	Linear	
Asthma	$PM_{2.5}$	Thurston et al.	1994	Toronto, CAN	All	All	All	O_3	24-hr avg	0.0132	0.0273	Linear	
Asthma	$PM_{10-2.5}$	Thurston et al.	1994	Toronto, CAN	All	All	All	None	24-hr avg	0.0670	0.0571	Linear	
Asthma	PM_{10}	Thurston et al.	1994	Toronto, CAN	All	All	All	None	24-hr avg	0.0248	0.0180	Linear	

Appendix F. Particulate Matter C-R Functions

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time	Beta	Std Error	Functional Form	Notes
Asthma	PM_{10}	Thurston et al.	1994	Toronto, CAN	All	All	All	O_3	24-hr avg	0.0039	0.0208	Linear	
Chronic Lung Disease	$PM_{2.5}$	Ito	2003	Detroit, MI	65+	All	All	None	24-hr avg	0.001169	0.002064	Log-linear	Reanalysis of Lippmann et al., 2000
Chronic Lung Disease	$PM_{2.5}$	Lippmann et al.	2000	Detroit, MI	65+	All	All	O_3	24-hr avg	0.001089	0.002420	Log-linear	
Chronic Lung Disease	$PM_{2.5}$	Moolgavkar	2003	Los Angeles, CA	65+	All	All	None	24-hr avg	0.001833	0.000519	Log-linear	Reanalysis of Moolgavkar, 2000b
Chronic Lung Disease	$PM_{2.5}$	Moolgavkar	2003	Los Angeles, CA	65+	All	All	NO_2	24-hr avg	0.000419	0.000676	Log-linear	Reanalysis of Moolgavkar, 2000b
Chronic Lung Disease	$PM_{2.5}$	Moolgavkar	2000	Los Angeles, CA	65+	All	All	CO	24-hr avg	0.0008	0.001000	Log-linear	
Chronic Lung Disease	$PM_{2.5}$	Moolgavkar	2000	Los Angeles, CA	18-64	All	All	None	24-hr avg	0.0022	0.000733	Log-linear	
Chronic Lung Disease	$PM_{2.5}$	Moolgavkar	2000	Los Angeles, CA	18-64	All	All	CO	24-hr avg	0.0020	0.000909	Log-linear	
Chronic Lung Disease	PM_{10}	Moolgavkar et al.	1997	Minneapolis, MN	65+	All	All	CO, O_3	24-hr avg	0.000877	0.000777	Log-linear	
Chronic Lung Disease	PM_{10}	Schwartz	1994	Minneapolis, MN	65+	All	All	None	24-hr avg	0.003853	0.001461	Log-linear	
Chronic Lung Disease (less Asthma)	$PM_{2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.001868	0.000988	Log-linear	
Chronic Lung Disease (less Asthma)	$PM_{2.5}$	Moolgavkar	2000	Los Angeles, CA	18-64	All	All	CO	24-hr avg	0.0020	0.000909	Log-linear	
Chronic Lung Disease (less Asthma)	$PM_{10-2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.004830	0.001482	Log-linear	
Chronic Lung Disease (less Asthma)	$PM_{10-2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	CO, O_3	24-hr avg	0.003104	0.001634	Log-linear	

Appendix F. Particulate Matter C-R Functions

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time	Beta	Std Error	Functional Form	Notes
Chronic Lung Disease (less Asthma)	PM_{10}	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.001334	0.000547	Log-linear	
Chronic Lung Disease (less Asthma)	PM_{10}	Samet et al.	2000	14 cities	65+	All	All	None	24-hr avg	0.002839	0.001351	Log-linear	
Chronic Lung Disease (less Asthma)	PM_{10}	Schwartz	1994	Detroit, MI	65+	All	All	O_3	24-hr avg	0.00202	0.00059	Log-linear	
Pneumonia	$PM_{2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.004090	0.000672	Log-linear	
Pneumonia	$PM_{2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	NO_2, O_3	24-hr avg	0.003279	0.000735	Log-linear	
Pneumonia	$PM_{10-2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.003561	0.000890	Log-linear	
Pneumonia	PM_{10}	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.002656	0.000446	Log-linear	
Pneumonia	$PM_{2.5}$	Ito	2003	Detroit, MI	65+	All	All	None	24-hr avg	0.003979	0.001659	Log-linear	Reanalysis of Lippmann et al., 2000
Pneumonia	$PM_{2.5}$	Lippmann et al.	2000	Detroit, MI	65+	All	All	O_3	24-hr avg	0.004480	0.001918	Log-linear	
Pneumonia	PM_{10}	Moolgavkar et al.	1997	Minneapolis, MN	65+	All	All	NO_2, O_3, SO_2	24-hr avg	0.000498	0.000505	Log-linear	
Pneumonia	PM_{10}	Samet et al.	2000	14 cities	65+	All	All	None	24-hr avg	0.002049	0.000570	Log-linear	
Pneumonia	PM_{10}	Schwartz	1994	Minneapolis, MN	65+	All	All	None	24-hr avg	0.001570	0.000652	Log-linear	
Pneumonia	PM_{10}	Schwartz	1994	Minneapolis, MN	65+	All	All	O_3	24-hr avg	0.001655	0.000709	Log-linear	
Pneumonia	PM_{10}	Schwartz	1994	Detroit, MI	65+	All	All	O_3	24-hr avg	0.00115	0.00039	Log-linear	
All Cardiovascular	$PM_{2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	None	24-hr avg	0.002775	0.001542	Log-linear	
All Cardiovascular	$PM_{2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	O_3	24-hr avg	0.001264	0.001620	Log-linear	
All Cardiovascular	$PM_{2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	NO_2, O_3, SO_2	24-hr avg	-0.000639	0.001935	Log-linear	
All Cardiovascular	$PM_{10-2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	None	24-hr avg	0.007446	0.002183	Log-linear	

Appendix F. Particulate Matter C-R Functions

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time	Beta	Std Error	Functional Form	Notes
All Cardiovascular	$PM_{10-2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	O_3	24-hr avg	0.007039	0.002146	Log-linear	
All Cardiovascular	$PM_{10-2.5}$	Burnett et al.	1997	Toronto, CAN	All	All	All	NO_2, O_3, SO_2	24-hr avg	0.004581	0.002727	Log-linear	
All Cardiovascular	PM_{10}	Burnett et al.	1997	Toronto, CAN	All	All	All	None	24-hr avg	0.002278	0.001017	Log-linear	
All Cardiovascular	PM_{10}	Burnett et al.	1997	Toronto, CAN	All	All	All	O_3	24-hr avg	0.001733	0.001031	Log-linear	
All Cardiovascular	PM_{10}	Burnett et al.	1997	Toronto, CAN	All	All	All	NO_2, O_3, SO_2	24-hr avg	-0.000281	0.001223	Log-linear	
All Cardiovascular	$PM_{2.5}$	Moolgavkar	2003	Los Angeles, CA	65+	All	All	None	24-hr avg	0.001568	0.000342	Log-linear	Reanalysis of Moolgavkar, 2000a
All Cardiovascular	$PM_{2.5}$	Moolgavkar	2003	Los Angeles, CA	65+	All	All	CO	24-hr avg	0.000389	0.000423	Log-linear	Reanalysis of Moolgavkar, 2000a
All Cardiovascular	$PM_{2.5}$	Moolgavkar	2000	Los Angeles, CA	18-64	All	All	None	24-hr avg	0.0014	0.000341	Log-linear	
All Cardiovascular	$PM_{2.5}$	Moolgavkar	2000	Los Angeles, CA	18-64	All	All	CO	24-hr avg	0.0009	0.000500	Log-linear	
All Cardiovascular	PM_{10}	Samet et al.	2000	14 cities	65+	All	All	None	24-hr avg	0.001183	0.000111	Log-linear	
Dysrhythmia	$PM_{2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.002355	0.000809	Log-linear	
Dysrhythmia	$PM_{2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	CO, O_3	24-hr avg	0.001356	0.000910	Log-linear	
Dysrhythmia	$PM_{10-2.5}$	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.002000	0.001064	Log-linear	
Dysrhythmia	PM_{10}	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	0.001616	0.000533	Log-linear	
Dysrhythmia	$PM_{2.5}$	Ito	2003	Detroit, MI	65+	All	All	None	24-hr avg	0.001249	0.002033	Log-linear	Reanalysis of Lippmann et al., 2000
Dysrhythmia	$PM_{2.5}$	Lippmann et al.	2000	Detroit, MI	65+	All	All	O_3	24-hr avg	0.002138	0.002525	Log-linear	
Congestive Heart Failure	$PM_{2.5}$	Ito	2003	Detroit, MI	65+	All	All	None	24-hr avg	0.003074	0.001292	Log-linear	Reanalysis of Lippmann et al., 2000

Appendix F. Particulate Matter C-R Functions

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time	Beta	Std Error	Functional Form	Notes
Congestive Heart Failure	$PM_{2.5}$	Lippmann et al.	2000	Detroit, MI	65+	All	All	O_3	24-hr avg	0.004668	0.001650	Log-linear	
Ischemic Heart Disease	$PM_{2.5}$	Ito	2003	Detroit, MI	65+	All	All	None	24-hr avg	0.001435	0.001156	Log-linear	Reanalysis of Lippmann et al., 2000
Ischemic Heart Disease	$PM_{2.5}$	Lippmann et al.	2000	Detroit, MI	65+	All	All	O_3	24-hr avg	0.001116	0.001339	Log-linear	

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.

F .4 Hospitalizations

F .4.1 Hospital Admissions for All Respiratory (Burnett et al., 1997, Toronto)

Burnett et al. (1997) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada during the summers of 1992-1994. In a Poisson regression, all respiratory admissions (ICD codes 464-466,480-486,490-494,496) were linked to coefficient of haze (COH) and ozone; other PM measures were less strongly linked. In two pollutant models, they found that CO, NO_2, and SO_2 were not significant, controlling for COH. They found that ozone was still significant, controlling for COH. In multipollutant models with COH, O_3, NO_2, and SO_2, both ozone and COH remained signifcant. None of the other PM measures (PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$) were significant in four-pollutant models. The PM C-R functions are based on results from single and multipollutant models.

$PM_{2.5}$ Function(s)

Single Pollutant Model ($PM_{2.5}$)

In a single pollutant model with adjustment for temperature and dew point, the $PM_{2.5}$ coefficient and standard error are based on a relative risk of 1.037 (t-statistic 3.29) for an 11 $\mu g/m^3$ increase in four-day average $PM_{2.5}$ (Burnett et al., 1997, Table 2, p. 617).

Functional Form: Log-linear
Coefficient: 0.003303
Standard Error: 0.001004
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person (ICD codes 464, 466, 480-487, 490-496)
Population: population of all ages

Multipollutant Model ($PM_{2.5}$ and ozone)

In a model with ozone, the $PM_{2.5}$ coefficient and standard error are based on a relative risk of 1.027 (t-statistic 2.33) for an 11 $\mu g/m^3$ increase in four-day average $PM_{2.5}$ (Burnett et al., 1997, Table 4, p. 618).

Functional Form: Log-linear
Coefficient: 0.002422
Standard Error: 0.001039
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person (ICD codes 464, 466, 480-487, 490-496)
Population: population of all ages

Multipollutant Model ($PM_{2.5}$, NO_2, ozone, and SO_2)

In a four-pollutant model with NO_2, O_3, and SO_2, the $PM_{2.5}$ coefficient and standard error are based on a relative risk of 0.999 (t-statistic 0.10) for an 11 $\mu g/m^3$ increase in four-day average $PM_{2.5}$ (Burnett et al., 1997, Table 6, p. 618).

Functional Form: Log-linear
Coefficient: -0.000091
Standard Error: 0.000910
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person (ICD codes 464, 466, 480-487, 490-496)
Population: population of all ages

$PM_{10-2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model with adjustment for temperature and dew point, the $PM_{10-2.5}$ coefficient and standard error are based on a relative risk of 1.023 (t-statistic 3.41) for a 4.75 $\mu g/m^3$ increase in five-day average $PM_{10-2.5}$ (Burnett et al., 1997, Table 2, p. 617).

Functional Form: Log-linear
Coefficient: 0.004787
Standard Error: 0.001404
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person (ICD codes 464, 466, 480-487, 490-496)
Population: population of all ages

Multipollutant Model ($PM_{10-2.5}$ and ozone)

In a model with ozone, the $PM_{10-2.5}$ coefficient and standard error are based on a relative risk of 1.020 (t-statistic 3.04) for a 4.75 $\mu g/m^3$ increase in five-day average $PM_{10-2.5}$ (Burnett et al., 1997, Table 4, p. 618).

Functional Form: Log-linear
Coefficient: 0.004169
Standard Error: 0.001371
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person (ICD codes 464, 466, 480-487, 490-496)
Population: population of all ages

Multipollutant Model ($PM_{10-2.5}$, NO_2, ozone, and SO_2)

In a four-pollutant model with NO_2, ozone, and SO_2, the $PM_{10-2.5}$ coefficient and standard error are based on a relative risk of 1.007 (t-statistic 0.82) for a 4.75 $\mu g/m^3$ increase in five-day average $PM_{10-2.5}$ (Burnett et al., 1997, Table 6, p. 618).

Functional Form: Log-linear
Coefficient: 0.001469
Standard Error: 0.001791
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person (ICD codes 464, 466, 480-487, 490-496)
Population: population of all ages

PM_{10} Function(s)

Single Pollutant Model

In a single pollutant model with adjustment for temperature and dew point, the PM_{10} coefficient and standard error are based on a relative risk of 1.03 (t-statistic 3.42) for a 14.25 $\mu g/m^3$ increase in five-day average PM_{10} (Burnett et al., 1997, Table 2, p. 617).

Functional Form: Log-linear
Coefficient: 0.002074
Standard Error: 0.000607
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person (ICD codes 464, 466, 480-487, 490-496)
Population: population of all ages

Multipollutant Model (PM_{10} and ozone)

In a model with ozone, the PM_{10} coefficient and standard error are based on a relative risk of 1.027 (t-statistic 3.16) for a 14.25 $\mu g/m^3$ increase in five-day average PM_{10} (Burnett et al., 1997, Table 4, p. 618).

Functional Form: Log-linear
Coefficient: 0.001870
Standard Error: 0.000592
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person (ICD codes 464, 466, 480-487, 490-496)
Population: population of all ages

Multipollutant Model (PM_{10}, NO_2, ozone, and SO_2)

In a four-pollutant model with NO_2, ozone, and SO_2, the PM_{10} coefficient and standard error are based on a relative risk of 1.004 (t-statistic 0.36) for a 14.25 $\mu g/m^3$ increase in five-day average PM_{10} (Burnett et al., 1997, Table 6, p. 618).

Functional Form: Log-linear
Coefficient: 0.000280
Standard Error: 0.000778
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person (ICD codes 464, 466, 480-487, 490-496)
Population: population of all ages

F .4.2 Hospital Admissions for All Respiratory (Burnett et al., 2001, Toronto)

Burnett et al. (2001) studied the association between air pollution and acute respiratory hospital admissions (ICD codes 493, 466, 464.4, 480-486) in Toronto from 1980-1994, among children <2 years of age. They collected hourly concentrations of the gaseous pollutants, CO, NO_2, SO_2, and ozone. Daily measures of particulate matter were estimated for the May to August period of 1992-1994 using TSP, sulfates, and coefficient of haze data. The authors report a positive association between ozone in the May through August months and respiratory hospital admissions, for several single days after elevated ozone levels. The strongest association was found using a five-day moving average of ozone. No association was

found in the September through April months. In co-pollutant models with a particulate matter or another gaseous pollutant, the ozone effect was only slightly diminished. The effects for PM and gaseous pollutants were generally significant in single pollutant models but diminished in co-pollutant models with ozone, with the exception of CO. The C-R functions for $PM_{10-2.5}$ are based on a single pollutant and co-pollutant model, using the four-day moving average of $PM_{10-2.5}$. The C-R functions for $PM_{2.5}$ are based on a single pollutant and co-pollutant model, using the four-day moving average of $PM_{2.5}$.

$PM_{2.5}$ Function(s)

Single Pollutant Model

The single pollutant coefficient and standard error are based on a percent increase of 15.8 (t-stat 3.29) for an 18.0 $\mu g/m^3$ increase in four-day average $PM_{2.5}$ (Burnett et al., 2001, Table 3).

Functional Form: Log-linear
Coefficient: 0.008150
Standard Error: 0.002477
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person <2 (ICD codes 464, 466, 480-487, 493)
Population: population of ages under 2

Multipollutant Model ($PM_{2.5}$ and ozone)

In a model with ozone, the coefficient and standard error are based on a percent increase of 1.4 (t-stat 0.24) for an 18.0 $\mu g/m^3$ increase in four-day average $PM_{2.5}$ (Burnett et al., 2001, Table 3).

Functional Form: Log-linear
Coefficient: 0.000772
Standard Error: 0.003218
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person <2 (ICD codes 464, 466, 480-487, 493)
Population: population of ages under 2

$PM_{10-2.5}$ Function(s)

Single Pollutant Model

The single pollutant coefficient and standard error are based on a percent increase of 18.3 (t-stat 3.77) for a 16.2 $\mu g/m^3$ increase in four-day average $PM_{10-2.5}$ (Burnett et al., 2001, Table 3).

Functional Form: Log-linear
Coefficient: 0.010374
Standard Error: 0.002752
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person <2 (ICD codes 464, 466, 480-487, 493)
Population: population of ages under 2

Multipollutant Model ($PM_{10-2.5}$ and ozone)

In a model with ozone, the coefficient and standard error are based on a percent increase of 4.5 (t-stat 0.72) for a 16.2 μg/m^3 increase in four-day average $PM_{10-2.5}$ (Burnett et al., 2001, Table 3).

Functional Form: Log-linear
Coefficient: 0.002717
Standard Error: 0.003774
Incidence Rate: region-specific daily hospital admission rate for all respiratory disease per person <2 (ICD codes 464, 466, 480-487, 493)
Population: population of ages under 2

F .4.3 Hospital Admissions for All Respiratory (Schwartz, 1995, New Haven)

Schwartz (1995) examined the relationship between air pollution and respiratory hospital admissions (ICD codes 460-519) for individuals 65 and older in New Haven, Connecticut, from January 1988 to December 1990. In single-pollutant models, PM_{10} and SO_2 were significant, while ozone was marginally significant. In two-pollutant models, ozone was significant in a model with PM_{10} and not significant in a model with SO_2, but had relatively stable coefficient estimates. PM_{10} was significant in two-pollutant models with ozone and SO_2. SO_2 was significant only in the co-pollutant model with PM_{10}. The PM_{10} C-R functions are based on results from a single pollutant and two-pollutant model (PM_{10} and ozone).

Single Pollutant Model

In a single-pollutant model, the coefficient and standard error are calculated from the relative risk (1.06) and 95% confidence interval (1.00-1.13) for a 50 μg/m^3 increase in average daily PM_{10} levels (Schwartz, 1995, Table 3, p. 534).

Functional Form: Log-linear
Coefficient: 0.001165
Standard Error: 0.000624
Incidence Rate: region-specific daily hospital admission rate for respiratory admissions per person 65+ (ICD codes 460-519)
Population: population of ages 65 and older

Multipollutant Model (PM_{10} and ozone)

In a model with ozone, the coefficient and standard error are estimated from the relative risk (1.09) and 95% confidence interval (1.00-1.20) for a 50 μg/m^3 increase in average daily PM_{10} levels (Schwartz, 1995, Table 3, p. 534).

Functional Form: Log-linear
Coefficient: 0.001724
Standard Error: 0.000930
Incidence Rate: region-specific daily hospital admission rate for respiratory admissions per person 65+ (ICD codes 460-519)
Population: population of ages 65 and older

F .4.4 Hospital Admissions for All Respiratory (Schwartz, 1995, Tacoma)

Schwartz (1995) examined the relationship between air pollution and hospital admissions for individuals 65 and older in Tacoma, Washington, from January 1988 to December 1990. In single-pollutant models, PM_{10}, ozone, and SO_2 were all significant. Ozone remained significant in two-pollutant models with PM_{10} and SO_2, and had stable coefficient estimates. PM_{10} was significant in a two-pollutant model with SO_2, but not in a model with ozone, although the central estimate remained stable. SO_2 was not significant in two-pollutant models with ozone or PM_{10}. The PM_{10} C-R functions are based on results from a single pollutant and two-pollutant model (PM_{10} and ozone).

Single Pollutant Model

In a single-pollutant model, the coefficient and standard error are calculated from the relative risk (1.10) and 95% confidence interval (1.03-1.17) for a 50 $\mu g/m^3$ increase in average daily PM_{10} levels (Schwartz, 1995, Table 6, p. 535).

Functional Form: Log-linear
Coefficient: 0.001906
Standard Error: 0.000650
Incidence Rate: region-specific daily hospital admission rate for respiratory admissions per person 65+ (ICD codes 460-519)
Population: population of ages 65 and older

Multipollutant Model (PM_{10} and ozone)

In a model with PM_{10}, the coefficient and standard error are estimated from the relative risk (1.12) and 95% CI (0.97-1.29) for a 50 $\mu g/m^3$ increase in average daily PM_{10} levels (Schwartz, 1995, Table 6, p. 535).

Functional Form: Log-linear
Coefficient: 0.002267
Standard Error: 0.001455
Incidence Rate: region-specific daily hospital admission rate for respiratory admissions per person 65+ (ICD codes 460-519)
Population: population of ages 65 and older

F .4.5 Hospital Admissions for All Respiratory (Thurston et al., 1994, Toronto)

Thurston et al. (1994) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada, for six weeks in July and August 1986-1988. In single-pollutant linear regression models, ozone and various measures of PM were linked to all respiratory admissions (ICD codes 466, 480-482, 485, 490-493). In two-pollutant models, ozone was still significant, but measures of PM were often not significant; only H^+ was significant. The C-R functions for $PM_{2.5}$ and PM_{10} are based on results from the reported single pollutant models and co-pollutant models with ozone. For $PM_{10-2.5}$, results are reported only from a single pollutant model.

$PM_{2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model, the $PM_{2.5}$ coefficient (0.0828) and standard error (0.0367) are reported in Table 3 (Thurston et al., 1994, p. 281) for a unit $\mu g/m^3$ increase daily average $PM_{2.5}$ levels.

Functional Form: Linear
Coefficient: 0.0828
Standard Error: 0.0367
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

Multipollutant Model ($PM_{2.5}$ and ozone)

In a model with ozone, the $PM_{2.5}$ coefficient (0.0434) and standard error (0.0429) are reported in Table 3 (Thurston et al., 1994, p. 281) for a unit $\mu g/m^3$ increase daily average $PM_{2.5}$ levels.

Functional Form: Linear
Coefficient: 0.0434
Standard Error: 0.0429
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

$PM_{10-2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model, the $PM_{10-2.5}$ coefficient (0.1228) and standard error (0.0895) are reported in Table 3 (Thurston et al., 1994, p. 281) for a unit $\mu g/m^3$ increase daily average $PM_{10-2.5}$ levels.

Functional Form: Linear
Coefficient: 0.1228
Standard Error: 0.0895
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

PM_{10} Function(s)

Single Pollutant Model

In a single pollutant model, the PM_{10} coefficient (0.0642) and standard error (0.0290) are reported in Table 3 (Thurston et al., 1994, p. 281) for a unit $\mu g/m^3$ increase daily average PM_{10} levels.

Functional Form: Linear
Coefficient: 0.0642
Standard Error: 0.0290
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

Multipollutant Model (PM_{10} and ozone)

In a model with ozone, the PM_{10} coefficient (0.0339) and standard error (0.0344) are reported in Table 3 (Thurston et al., 1994, p. 281) for a unit $\mu g/m^3$ increase daily average PM_{10} levels.

Functional Form: Linear
Coefficient: 0.0339
Standard Error: 0.0344
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

F .4.6 Hospital Admissions for Asthma (Burnett et al., 1999, Toronto)

Burnett et al. (1999) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada from1980 to 1994. The authors examined single pollutant log-linear models for PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$, CO, NO_2, SO_2, and ozone and found all significantly associated with asthma except SO_2. They estimated multiple pollutant models, where pollutants for best fitting model were chosen using stepwise regression based on AIC criterion. Asthma admissions were linked to ozone, CO, and $PM_{10-2.5}$. The C-R functions for $PM_{10-2.5}$ are based on the results of a single pollutant model and three-pollutant model (O_3, CO, $PM_{10-2.5}$). The C-R functions for $PM_{2.5}$ and PM_{10} are based on the results of a single pollutant model.

$PM_{2.5}$ Function(s)

Single Pollutant Model

The coefficient and standard error are based on the percent increase (4.60) and t-statistic (3.22) reported in Table 3 (Burnett et al., 1999, p. 133) for an 18.0 $\mu g/m^3$ increase in three-day average $PM_{2.5}$ levels.

Functional Form: Log-linear
Coefficient: 0.002499
Standard Error: 0.000776
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per person (ICD code 493)
Population: population of all ages

$PM_{10-2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (5.25) and t-statistic (4.20) reported in Table 3 (Burnett et al., 1999, p. 133) for a 12.2 $\mu g/m^3$ increase in three-day average $PM_{10-2.5}$ levels.

Functional Form: Log-linear
Coefficient: 0.004194
Standard Error: 0.000999
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per person (ICD code 493)
Population: population of all ages

Multipollutant Model ($PM_{10-2.5}$, CO, and ozone)

In a model with ozone and CO, the $PM_{10-2.5}$ coefficient and standard error are based on the percent increase (4.00) reported in Table 5 (Burnett et al., 1999, p. 135) and the t-statistic obtained from the authors (3.04)[44] for a 12.2 $\mu g/m^3$ increase in three-day average $PM_{10-2.5}$ levels.

Functional Form: Log-linear
Coefficient: 0.003215
Standard Error: 0.001058
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per person (ICD code 493)
Population: population of all ages

PM_{10} Function(s)

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (5.27) and t-statistic (3.39) reported in Table 3 (Burnett et al., 1999, p. 133) for a 30.2 $\mu g/m^3$ increase in three-day average PM_{10} levels.

Functional Form: Log-linear
Coefficient: 0.001701
Standard Error: 0.000502
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per person (ICD code 493)
Population: population of all ages

[44] Rick Burnett (co-author), personal communication.

F .4.7 Hospital Admissions for Asthma (Lin et al., 2002, Toronto)

Lin et al. (2002) examined the association between ambient particulate matter in Toronto and asthma hospitalizations in children (ages 6-12) between 1981 and 1993. The authors collected PM data measured every six days for the period of 1984 to 1990.[45] The authors analyzed the PM-asthma hospitalization association using a case-crossover analysis (with unidirectional and bidirectional controls)[46] and a time series analysis with moving averages of PM ranging from 1 day to 7 days . They estimated the effects on boys and girls separately and found an increasing association between $PM_{10-2.5}$ and asthma hospitalizations as averaging time increased, with a leveling off around six or seven days. This effect remained significant in a model with CO, NO_2, SO_2, and ozone. Results for gaseous pollutants were not reported. They did not find a significant association for $PM_{2.5}$ or PM_{10} in models other than the unidirectional case-crossover analysis. The authors suggest that estimates from a unidirectional case-crossover analysis may be significantly biased when time trends are present. The considerable difference between the results from this model and the bidirectional and time series analyses suggest that this may be the case.

The C-R functions for PM are based on the time series analysis rather than the bidirectional case-crossover because the time series produces more stable estimates (i.e., the 95% confidence intervals are always narrower than those from the case-crossover design) and this design is more commonly used in air pollution epidemiology. The reported relative risks for $PM_{10-2.5}$ increase as the number of days included in the moving average increases – up through 7 days (the maximum number of days considered). This suggests that the multi-day averages are capturing to some extent what is essentially a distributed lag effect – that is, that $PM_{10-2.5}$ even 7 days earlier has some impact on asthma hospitalization rates. We therefore selected the model with the 7-day average for use in the single pollutant C-R functions. In multipollutant models, only results using 5- and 6-day averages were reported, so the C-R functions are based on 6-day averages.

$PM_{2.5}$ Function(s) – Boys

Single Pollutant Model

The single pollutant coefficient and standard error for boys are based on the relative risk (0.96) and 95% confidence interval (0.91-1.02) for a 9.3 $\mu g/m^3$ increase in 7-day average $PM_{2.5}$ (Lin et al., 2002, Table 3, p. 579)

Functional Form: Log-linear
Coefficient: -0.004389
Standard Error: 0.003130
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per male child ages 6-12 (ICD code 493)
Population: population of males, ages 6 through 12

[45] For the remaining days, they estimated PM using TSP, sulfate, and coefficient of haze data.

[46] In the case-crossover analysis, the same individual serves as a case and control. In the unidirectional model, the case period is during the hospital visit and the control period is at some point well in advance of the case period. In the bidirectional model, there are two control periods for each visit, one before the case period and one after the case period.

Multipollutant Model ($PM_{2.5}$, CO, NO_2, ozone, and SO_2)

In a model with CO, NO_2, ozone, and SO_2, the coefficient and standard error for boys are based on the relative risk (0.94) and 95% confidence interval (0.88-1.01) for a 9.3 μg/m^3 increase in 6-day average $PM_{2.5}$ (Lin et al., 2002, Table 5, p. 580).

Functional Form: Log-linear
Coefficient: -0.006653
Standard Error: 0.003779
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per male child ages 6-12 (ICD code 493)
Population: population of males, ages 6 through 12

$PM_{2.5}$ Function(s) – Girls

Single Pollutant Model

The single pollutant coefficient and standard error for girls are based on the relative risk (1.06) and 95% confidence interval (0.98-1.13) for a 9.3 μg/m^3 increase in 7-day average $PM_{2.5}$ (Lin et al., 2002, Table 4, p. 580)

Functional Form: Log-linear
Coefficient: 0.006265
Standard Error: 0.008377
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per female child ages 6-12 (ICD code 493)
Population: population of females, ages 6 through 12

Multipollutant Model ($PM_{2.5}$, CO, NO_2, ozone, and SO_2)

In a model with CO, NO_2, ozone, and SO_2, the coefficient and standard error for girls are based on the relative risk (0.98) and 95% confidence interval (0.90-1.08) for a 9.3 μg/m^3 increase in 6-day average $PM_{2.5}$ (Lin et al., 2002, Table 5, p. 580).

Functional Form: Log-linear
Coefficient: -0.002172
Standard Error: 0.005001
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per female child ages 6-12 (ICD code 493)
Population: population of females, ages 6 through 12

$PM_{10-2.5}$ Function(s) – Boys

Single Pollutant Model

The single pollutant coefficient and standard error for boys are based on the relative risk (1.12) and 95% confidence interval (1.04-1.20) for an 8.4 μg/m^3 increase in 7-day average $PM_{10-2.5}$ (Lin et al., 2002, Table 3, p. 579)

Functional Form: Log-linear
Coefficient: 0.013492
Standard Error: 0.004346
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per male child ages 6-12 (ICD code 493)
Population: population of males, ages 6 through 12

Multipollutant Model ($PM_{10-2.5}$, CO, NO_2, ozone, and SO_2)

In a model with CO, NO_2, ozone, and SO_2, the coefficient and standard error for boys are based on the relative risk (1.15) and 95% confidence interval (1.06-1.25) for an 8.4 $\mu g/m^3$ increase in 6-day average $PM_{10-2.5}$ (Lin et al., 2002, Table 5, p. 580).

Functional Form: Log-linear
Coefficient: 0.016638
Standard Error: 0.005007
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per male child ages 6-12 (ICD code 493)
Population: population of males, ages 6 through 12

$PM_{10-2.5}$ Function(s) – Girls

Single Pollutant Model

The single pollutant coefficient and standard error for girls are based on the relative risk (1.20) and 95% confidence interval (1.09-1.31) for an 8.4 $\mu g/m^3$ increase in 7-day average $PM_{10-2.5}$ (Lin et al., 2002, Table 4, p. 580)

Functional Form: Log-linear
Coefficient: 0.021705
Standard Error: 0.005583
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per female child ages 6-12 (ICD code 493)
Population: population of females, ages 6 through 12

Multipollutant Model ($PM_{10-2.5}$, CO, NO_2, ozone, and SO_2)

In a model with CO, NO_2, ozone, and SO_2, the coefficient and standard error for girls are based on the relative risk (1.15) and 95% confidence interval (1.03-1.29) for an 8.4 $\mu g/m^3$ increase in 6-day average $PM_{10-2.5}$ (Lin et al., 2002, Table 5, p. 580).

Functional Form: Log-linear
Coefficient: 0.016638
Standard Error: 0.006836
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per female child ages 6-12 (ICD code 493)
Population: population of females, ages 6 through 12

PM_{10} Function(s) – Boys

Single Pollutant Model

The single pollutant coefficient and standard error for boys are based on the relative risk (1.01) and 95% confidence interval (0.95-1.08) for a 14.8 μg/m^3 increase in 7-day average PM_{10} (Lin et al., 2002, Table 3, p. 579)

Functional Form: Log-linear
Coefficient: 0.000672
Standard Error: 0.002211
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per male child ages 6-12 (ICD code 493)
Population: population of males, ages 6 through 12

Multipollutant Model (PM_{10}, CO, NO_2, ozone, and SO_2)

In a model with CO, NO_2, ozone, and SO_2, the coefficient and standard error for boys are based on the relative risk (1.02) and 95% confidence interval (0.94-1.11) for a 14.8 μg/m^3 increase in 6-day average PM_{10} (Lin et al., 2002, Table 5, p. 580).

Functional Form: Log-linear
Coefficient: 0.001338
Standard Error: 0.002865
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per male child ages 6-12 (ICD code 493)
Population: population of males, ages 6 through 12

PM_{10} Function(s) – Girls

Single Pollutant Model

The single pollutant coefficient and standard error for girls are based on the relative risk (1.07) and 95% confidence interval (0.98-1.16) for a 14.8 μg/m^3 increase in 7-day average PM_{10} (Lin et al., 2002, Table 4, p. 580)

Functional Form: Log-linear
Coefficient: 0.004572
Standard Error: 0.002906
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per female child ages 6-12 (ICD code 493)
Population: population of females, ages 6 through 12

Multipollutant Model (PM_{10}, CO, NO_2, ozone, and SO_2)

In a model with CO, NO_2, ozone, and SO_2, the coefficient and standard error for girls are based on the relative risk (1.03) and 95% confidence interval (0.93-1.15) for a 14.8 μg/m^3 increase in 6-day average PM_{10} (Lin et al., 2002, Table 5, p. 580).

Functional Form: Log-linear
Coefficient: 0.001997
Standard Error: 0.003660
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per female child ages 6-12 (ICD code 493)
Population: population of females, ages 6 through 12

F .4.8 Hospital Admissions for Asthma (Sheppard et al., 1999; Sheppard, 2003)

Sheppard et al. (1999) studied the relation between air pollution in Seattle and nonelderly (<65) hospital admissions for asthma from 1987 to 1994. They used air quality data for PM_{10}, $PM_{2.5}$, coarse $PM10_{10\text{-}2.5}$, SO_2, ozone, and CO in a Poisson regression model with control for time trends, seasonal variations, and temperature-related weather effects.[47] They found asthma hospital admissions associated with PM_{10}, $PM_{2.5}$, $PM_{10\text{-}2.5}$, CO, and ozone. They did not observe an association for SO_2. They found PM and CO to be jointly associated with asthma admissions. The best fitting co-pollutant models were found using ozone. However, ozone data was only available April through October, so they did not consider ozone further. For the remaining pollutants, the best fitting models included $PM_{2.5}$ and CO. Results for other co-pollutant models were not reported.

In response to concerns that the work by Sheppard et al. (1999) may be biased because of the Splus issue (discussed in Appendix D of this User Manual), Sheppard (2003) reanalyzed some of this work, in particular Sheppard reanalyzed the original study's $PM_{2.5}$ single pollutant model.

$PM_{2.5}$ Function(s)

Single Pollutant Model (Sheppard, 2003)

The coefficient and standard error are based on the relative risk (1.04) and 95% confidence interval (1.01-1.06) for a 11.8 μg/m^3 increase in $PM_{2.5}$ in the 1-day lag GAM stringent model (Sheppard, 2003, pp. 228-229).

Functional Form: Log-linear
Coefficient: 0.003324
Standard Error: 0.001045
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per person <65 (ICD code 493)
Population: population of ages 65 and under

Multipollutant Model ($PM_{2.5}$ and CO) (Sheppard et al., 1999)

The coefficient and standard error for the co-pollutant model with CO are calculated from a relative risk of 1.03 (95% CI 1.01-1.06) for an 11.8 μg/m^3 increase[48] in $PM_{2.5}$ (Sheppard et al., 1999, p. 28).

[47] $PM_{2.5}$ levels were estimated from light scattering data.

[48] The reported IQR change in the abstract and text is smaller than reported in Table 3. We assume the change reported in the abstract and text to be correct because greater number of significant figures are reported.

Functional Form: Log-linear
Coefficient: 0.002505
Standard Error: 0.001045
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per person <65 (ICD code 493)
Population: population of ages 65 and under

$PM_{10-2.5}$ Function(s)

Single Pollutant Model (Sheppard et al., 1999)

The single pollutant coefficient and standard error are calculated from a relative risk of 1.04 (95% CI 1.01-1.07) for a 9.3 $\mu g/m^3$ increase[49] in $PM_{10-2.5}$ (Sheppard et al., 1999, p. 27).

Functional Form: Log-linear
Coefficient: 0.004217
Standard Error: 0.001583
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per person <65 (ICD code 493)
Population: population of ages 65 and under

PM_{10} Function(s)

Single Pollutant Model (Sheppard et al., 1999)

The single pollutant coefficient and standard error are calculated from a relative risk of 1.05 (95% CI 1.02-1.08) for a 19 $\mu g/m^3$ increase[50] in PM_{10} (Sheppard et al., 1999, p. 27).

Functional Form: Log-linear
Coefficient: 0.002568
Standard Error: 0.000767
Incidence Rate: region-specific daily hospital admission rate for asthma admissions per person <65 (ICD code 493)
Population: population of ages 65 and under

F .4.9 Hospital Admissions for Asthma (Thurston et al., 1994, Toronto)

Thurston et al. (1994) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada, for six weeks in July and August 1986-1988. In single-pollutant linear regression models, ozone was strongly associated with asthma admissions (ICD code 493) and various measures of PM were marginally significant. In two-pollutant models, ozone remained significant,

[49] The reported IQR change in the abstract and text is smaller than reported in Table 3. We assume the change reported in the abstract and text to be correct because greater number of significant figures are reported.

[50] The reported IQR change in the abstract and text is smaller than reported in Table 3. We assume the change reported in the abstract and text to be correct because greater number of significant figures are reported.

but measures of PM were often not significant. The C-R functions for $PM_{2.5}$ and PM_{10} are based on results from the reported single pollutant models and co-pollutant models with ozone. For $PM_{10-2.5}$, results are reported only from a single pollutant model.

$PM_{2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model, the $PM_{2.5}$ coefficient (0.0334) and standard error (0.0241) are reported in Table 4 (Thurston et al., 1994, p. 282) for a unit $\mu g/m^3$ increase daily average $PM_{2.5}$ levels.

Functional Form: Linear
Coefficient: 0.0334
Standard Error: 0.0241
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

Multipollutant Model ($PM_{2.5}$ and ozone)

In a model with ozone, the $PM_{2.5}$ coefficient (0.0132) and standard error (0.0273) are reported in Table 4 (Thurston et al., 1994, p. 282) for a unit $\mu g/m^3$ increase daily average $PM_{2.5}$ levels.

Functional Form: Linear
Coefficient: 0.0132
Standard Error: 0.0273
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

$PM_{10-2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model, the $PM_{10-2.5}$ coefficient (0.0670) and standard error (0.0571) are reported in Table 4 (Thurston et al., 1994, p. 282) for a unit $\mu g/m^3$ increase daily average $PM_{10-2.5}$ levels.

Functional Form: Linear
Coefficient: 0.0670
Standard Error: 0.0571
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

PM_{10} Function(s)

Single Pollutant Model

In a single pollutant model, the PM_{10} coefficient (0.0248) and standard error (0.0180) are reported in Table 4 (Thurston et al., 1994, p. 282) for a unit $\mu g/m^3$ increase daily average PM_{10} levels.

Functional Form: Linear
Coefficient: 0.0248
Standard Error: 0.0180
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

Multipollutant Model (PM_{10} and ozone)

In a model with ozone, the PM_{10} coefficient (0.0039) and standard error (0.0208) are reported in Table 4 (Thurston et al., 1994, p. 282) for a unit $\mu g/m^3$ increase daily average PM_{10} levels.

Functional Form: Linear
Coefficient: 0.0039
Standard Error: 0.0208
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

F .4.10 Hospital Admissions for Chronic Lung Disease (Lippmann et al., 2000; Ito, 2003)

Lippmann et al. (2000) studied the association between particulate matter and daily mortality and hospitalizations among the elderly in Detroit, MI. Data were analyzed for two separate study periods, 1985-1990 and 1992-1994. The 1992-1994 study period had a greater variety of data on PM size and was the main focus of the report. The authors collected hospitalization data for a variety of cardiovascular and respiratory endpoints. They used daily air quality data for PM_{10}, $PM_{2.5}$, and $PM_{10-2.5}$ in a Poisson regression model with generalized additive models (GAM) to adjust for nonlinear relationships and temporal trends. In single pollutant models, all PM metrics were statistically significant for pneumonia (ICD codes 480-486), $PM_{10-2.5}$ and PM_{10} were significant for ischemic heart disease (ICD code 410-414), and $PM_{2.5}$ and PM_{10} were significant for heart failure (ICD code 428). There were positive, but not statistically significant associations, between the PM metrics and COPD (ICD codes 490-496) and dysrhythmia (ICD code 427). In separate co-pollutant models with PM and either ozone, SO_2, NO_2, or CO, the results were generally comparable. The $PM_{2.5}$ C-R functions are based on results of the single pollutant model and co-pollutant model with ozone.

In response to concerns with the Splus issue, Ito (2003) reanalyzed the study by Lippmann et al. (2000). The reanalysis by Ito reported that more generalized additive models with stringent convergence criteria and generalized linear models resulted in smaller relative risk estimates. Not all of the original results were replicated, so we present here a mix of C-R functions from the reanalysis and from the original study (when the reanalyzed results were not available).

Single Pollutant Model (Ito, 2003)

The coefficient and standard error are based on the relative risk (1.043) and 95% confidence interval (0.902-1.207) for a 36 $\mu g/m^3$ increase in $PM_{2.5}$ in the 3-day lag GAM stringent model (Ito, 2003, Table 8).

Functional Form: Log-linear
Coefficient: 0.001169
Standard Error: 0.002064
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person 65+ (ICD codes 490-496)
Population: population of ages 65 and older

Multipollutant Model ($PM_{2.5}$ and ozone) (Lippmann et al., 2000)

The co-pollutant coefficient and standard error are calculated from a relative risk of 1.040 (95% CI 0.877-1.234) for a 36 $\mu g/m^3$ increase in $PM_{2.5}$ (Lippmann et al., 2000, Table 14, p. 26).

Functional Form: Log-linear
Coefficient: 0.001089
Standard Error: 0.002420
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person 65+ (ICD codes 490-496)
Population: population of ages 65 and older

F .4.11 Hospital Admissions for Chronic Lung Disease (Moolgavkar, 2000c; Moolgavkar, 2003)

Moolgavkar (2000c) examined the association between air pollution and COPD hospital admissions (ICD 490-496) in the Chicago, Los Angeles, and Phoenix metropolitan areas. He collected daily air pollution data for ozone, SO_2, NO_2, CO, and PM_{10} in all three areas. $PM_{2.5}$ data was available only in Los Angeles. The data were analyzed using a Poisson regression model with generalized additive models to adjust for temporal trends. Separate models were run for 0 to 5 day lags in each location. Among the 65+ age group in Chicago and Phoenix, weak associations were observed between the gaseous pollutants and admissions. No consistent associations were observed for PM_{10}. In Los Angeles, marginally significant associations were observed for $PM_{2.5}$, which were generally lower than for the gases. In co-pollutant models with CO, the $PM_{2.5}$ effect was reduced. Similar results were observed in the 0-19 and 20-64 year old age groups.

In response to concerns with the Splus issue, Moolgavkar (2003) reanalyzed his earlier study. In the reanalysis, he reported that more generalized additive models with stringent convergence criteria and generalized linear models resulted in smaller relative risk estimates. Not all of the original results were replicated, so we present here a mix of C-R functions from the reanalysis and from the original study (when the reanalyzed results were not available).

The $PM_{2.5}$ C-R functions for the 65+ age group are based on the reanalysis in Moolgavkar (Moolgavkar, 2003) of the single and co-pollutant models ($PM_{2.5}$ and CO). The $PM_{2.5}$ C-R functions for the 20-64 age group are based on the original study's single and co-pollutant models ($PM_{2.5}$ and CO). Since the true PM effect is most likely best represented by a distributed lag model, then any single lag model should underestimate the total PM effect. As a result, we selected the lag models with the greatest effect estimates for use in the C-R functions.

Ages 65 and older

Single Pollutant Model (Moolgavkar, 2003)

The coefficient and standard error are calculated from an estimated percentage change of 1.85[51] and t-statistic of 3.53 for a 10 μg/m^3 increase in $PM_{2.5}$ in the 2-day lag GAM-30df stringent (10^{-8}) model (Moolgavkar, 2003, Table 17).

Functional Form: Log-linear
Coefficient: 0.001833
Standard Error: 0.000519
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person 65+ (ICD codes 490-496)
Population: population of ages 65 and older

Multipollutant Model ($PM_{2.5}$ and NO_2) (Moolgavkar, 2003)

In a model with $PM_{2.5}$ and NO_2, the coefficient and standard error are calculated from the estimated percentage change of 0.42[40] and t-statistic of 0.62 for a 10 μg/m^3 increase in $PM_{2.5}$ in the 0-day lag GAM-100df stringent (10^{-8}) model (Moolgavkar, 2003, Table 19).

Functional Form: Log-linear
Coefficient: 0.000419
Standard Error: 0.000676
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person 65+ (ICD codes 490-496)
Population: population of ages 65 and older

Multipollutant Model ($PM_{2.5}$ and CO) (Moolgavkar, 2000c)

In a model with CO, the coefficient and standard error are calculated from an estimated percent change of 0.8[52] and t-statistic of 0.8 for a 10 μg/m^3 increase in $PM_{2.5}$ in the two-day lag model (Moolgavkar, 2000c, Table 3, p. 80).

Functional Form: Log-linear
Coefficient: 0.0008
Standard Error: 0.001000
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person 65+ (ICD codes 490-496)
Population: population of ages 65 and older

---

[51] In a log-linear model, the percent change is equal to (RR - 1) * 100. In this study, Moolgavkar defines and reports the "estimated" percent change as (log RR * 100). Because the relative risk is close to 1, RR-1 and log RR are essentially the same. For example, a true percent change of 2.0 would result in a relative risk of 1.020 and coefficient of 0.001980. The "estimated" percent change, as reported by Moolgavkar, of 2.0 results in a relative risk of 1.020201 and coefficient of 0.002.

[52] In a log-linear model, the percent change is equal to (RR - 1) * 100. In this study, Moolgavkar defines and reports the "estimated" percent change as (log RR * 100). Because the relative risk is close to 1, RR-1 and log RR are essentially the same. For example, a true percent change of 0.8 would result in a relative risk of 1.008 and coefficient of 0.000797. The "estimated" percent change, as reported by Moolgavkar, of 0.8 results in a relative risk of 1.008032 and coefficient of 0.0008.

Ages 18 to 64[53]

Single Pollutant Model (Moolgavkar, 2000c)

The single pollutant coefficient and standard error are calculated from an estimated percent change of 2.2[54] and t-statistic of 3.0 for a 10 μg/m^3 increase in $PM_{2.5}$ in the two-day lag model (Moolgavkar, 2000c, Table 4, p. 81).

Functional Form: Log-linear
Coefficient: 0.0022
Standard Error: 0.000733
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person 18-64 (ICD codes 490-492, 494-496)[55]
Population: population of ages 18 to 64

Multipollutant Model ($PM_{2.5}$ and CO) (Moolgavkar, 2000c)

In a model with CO, the coefficient and standard error are calculated from an estimated percent change of 2.0[56] and t-statistic of 2.2 for a 10 μg/m^3 increase in $PM_{2.5}$ in the two-day lag model (Moolgavkar, 2000c, Table 4, p. 81).

Functional Form: Log-linear
Coefficient: 0.0020
Standard Error: 0.000909
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person 18-64 (ICD codes 490-492, 494-496)[57]
Population: population of ages 18 to 64

[53] Although Moolgavkar (2000c) reports results for the 20-64 year old age range, for comparability to other studies, we apply the results to the population of ages 18 to 64.

[54] In a log-linear model, the percent change is equal to (RR - 1) * 100. In this study, Moolgavkar defines and reports the "estimated" percent change as (log RR * 100). Because the relative risk is close to 1, RR-1 and log RR are essentially the same. For example, a true percent change of 2.2 would result in a relative risk of 1.022 and coefficient of 0.002176. The "estimated" percent change, as reported by Moolgavkar, of 2.2 results in a relative risk of 1.022244 and coefficient of 0.0022.

[55] Moolgavkar (2000c) reports results for ICD codes 490-496. In order to avoid double counting non-elderly asthma hospitalizations (ICD code 493) with Sheppard et al. (1999) in a total benefits estimation, we have excluded ICD code 493 from the baseline incidence rate used in this function.

[56] In a log-linear model, the percent change is equal to (RR - 1) * 100. In this study, Moolgavkar defines and reports the "estimated" percent change as (log RR * 100). Because the relative risk is close to 1, RR-1 and log RR are essentially the same. For example, a true percent change of 2.0 would result in a relative risk of 1.020 and coefficient of 0.001980. The "estimated" percent change, as reported by Moolgavkar, of 2.0 results in a relative risk of 1.020201 and coefficient of 0.002.

[57] Moolgavkar (2000c) reports results for ICD codes 490-496. In order to avoid double counting non-elderly asthma hospitalizations (ICD code 493) with Sheppard et al. (1999) in a total benefits estimation, we have excluded ICD code 493 from the baseline incidence rate used in this function.

F .4.12 Hospital Admissions for Chronic Lung Disease (Moolgavkar et al., 1997, Minneapolis)

Moolgavkar et al. (1997) examined the relationship between air pollution and chronic lung disease hospital admissions (ICD codes 490-496) for individuals 65 and older in Minneapolis-St. Paul, Minnesota, from January 1986 to December 1991. In a Poisson regression, they found no significant effect for any of the pollutants (PM_{10}, ozone, or CO). The effect for ozone was marginally significant. The PM_{10} C-R function is based on the results from a three-pollutant model (ozone, CO, PM_{10}) to estimate chronic lung disease incidence. The model with a 100 df smoother was reported to be optimal (p. 368).

Multipollutant Model (PM_{10}, CO, and ozone)

In a model with ozone and CO, the estimated PM_{10} coefficient and standard error are based on a 1.77 percent increase in admissions (95% CI -1.3, 4.9) due to a PM_{10} change of 20 $\mu g/m^3$ (Moolgavkar et al., 1997, Table 4 and p. 366).

Functional Form: Log-linear
Coefficient: 0.000877
Standard Error: 0.000777
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person 65+ (ICD codes 490-496)
Population: population of ages 65 and older

F .4.13 Hospital Admissions for Chronic Lung Disease (Schwartz, 1994a, Minneapolis)

Schwartz (1994c) examined the relationship between air pollution and hospital admissions for individuals 65 and older in Minneapolis, MN, from January 1986 to December 1989. In single-pollutants models, PM_{10} was significantly related to chronic lung disease. Ozone was not significantly linked to chronic lung disease and the results were not reported. The PM_{10} C-R function is based on the results of the single-pollutant model with "spline" smoothing.

Single Pollutant Model

In a model with spline functions to adjust for time and weather, the coefficient and standard error are based on the relative risk (1.47) and 95% confidence interval (1.10-1.95) associated with a 100 $\mu g/m^3$ increase in two-day average PM_{10} levels (Schwartz, 1994c, Table 4, p. 369) .

Functional Form: Log-linear
Coefficient: 0.003853
Standard Error: 0.001461
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person 65+ (ICD codes 490-496)
Population: population of ages 65 and older

F .4.14 Hospital Admissions for Chronic Lung Disease (less Asthma) (Burnett et al., 1999, Toronto)

Burnett et al. (1999) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada from1980 to 1994. The authors examined single pollutant log-linear models for PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$, CO, NO_2, SO_2, and ozone and found $PM_{10-2.5}$, PM_{10}, and ozone significantly associated with chronic lung disease (ICD codes 490-492, 496). They estimated multiple pollutant models, where pollutants for the best fitting model were chosen using stepwise regression based on AIC criterion. In a three pollutant model, admissions for chronic obstructive pulmonary disease (COPD) were linked to ozone and $PM_{10-2.5}$. A non-significant association was found with CO. The C-R functions for $PM_{2.5}$ and PM_{10} are based on the results of a single pollutant model. The C-R functions for $PM_{10-2.5}$ are based on the results of a single pollutant model and three-pollutant model (O_3, CO, $PM_{10-2.5}$).

$PM_{2.5}$ Function(s)

Single Pollutant Model

The coefficient and standard error are based on the percent increase (3.42) and t-statistic (1.89) reported in Table 3 (Burnett et al., 1999, p. 133) for an 18.0 μg/m^3 increase in two-day average $PM_{2.5}$ levels.

Functional Form: Log-linear
Coefficient: 0.001868
Standard Error: 0.000988
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person (ICD codes 490-492, 494-496)
Population: population of all ages

$PM_{10-2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (6.07) and t-statistic (3.26) reported in Table 3 (Burnett et al., 1999, p. 133) for a 12.2 μg/m^3 increase in three-day average $PM_{10-2.5}$ levels.

Functional Form: Log-linear
Coefficient: 0.004830
Standard Error: 0.001482
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person (ICD codes 490-492, 494-496)
Population: population of all ages

Multipollutant Model ($PM_{10-2.5}$, CO, and ozone)

In a model with ozone and CO, the $PM_{10-2.5}$ coefficient and standard error are based on the percent increase (3.86) reported in Table 5 (Burnett et al., 1999, p. 135) and the t-statistic obtained from the authors (1.90)[58] for a 12.2 μg/m^3 increase in three-day average $PM_{10-2.5}$ levels.

Functional Form: Log-linear
Coefficient: 0.003104
Standard Error: 0.001634
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person (ICD codes 490-492, 494-496)
Population: population of all ages

PM_{10} Function(s)

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (4.11) and t-statistic (2.44) reported in Table 3 (Burnett et al., 1999, p. 133) for a 30.2 μg/m^3 increase in three-day average PM_{10} levels.

Functional Form: Log-linear
Coefficient: 0.001334
Standard Error: 0.000547
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person (ICD codes 490-492, 494-496)
Population: population of all ages

F .4.15 Hospital Admissions for Chronic Lung Disease (less Asthma) (Moolgavkar, 2000c)

Moolgavkar (2000c) examined the association between air pollution and COPD hospital admissions (ICD 490-496) in the Chicago, Los Angeles, and Phoenix metropolitan areas. He collected daily air pollution data for ozone, SO_2, NO_2, CO, and PM_{10} in all three areas. $PM_{2.5}$ data was available only in Los Angeles. The data were analyzed using a Poisson regression model with generalized additive models to adjust for temporal trends. Separate models were run for 0 to 5 day lags in each location. Among the 65+ age group in Chicago and Phoenix, weak associations were observed between the gaseous pollutants and admissions. No consistent associations were observed for PM_{10}. In Los Angeles, marginally significant associations were observed for $PM_{2.5}$, which were generally lower than for the gases. In co-pollutant models with CO, the $PM_{2.5}$ effect was reduced. Similar results were observed in the 0-19 and 20-64 year old age groups.

The $PM_{2.5}$ C-R functions are based on the single and co-pollutant models ($PM_{2.5}$ and CO) reported for the 20-64 and 65+ age groups. Since the true PM effect is most likely best represented by a distributed

[58] Rick Burnett (co-author), personal communication.

lag model, then any single lag model should underestimate the total PM effect. As a result, we selected the lag models with the greatest effect estimates for use in the C-R functions.

Ages 18 to 64[59]

Multipollutant Model ($PM_{2.5}$ and CO)

In a model with CO, the coefficient and standard error are calculated from an estimated percent change of 2.0[60] and t-statistic of 2.2 for a 10 $\mu g/m^3$ increase in $PM_{2.5}$ in the two-day lag model (Moolgavkar, 2000c, Table 4, p. 81).

Functional Form: Log-linear
Coefficient: 0.0020
Standard Error: 0.000909
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease admissions per person 18-64 (ICD codes 490-492, 494-496)[61]
Population: population of ages 18 to 64

F .4.16 Hospital Admissions for Chronic Lung Disease (less Asthma) (Samet et al., 2000, 14 Cities)

Samet et al. (2000) examined the relationship between air pollution and hospital admissions for individuals of ages 65 and over in 14 cities across the country.[62] Cities were selected on the basis of available air pollution data for at least four years between 1985 and 1994 during which at least 50% of days had observations between the city-specific start and end of measurements. Hospital admissions were obtained from the Health Care Financing Administration (HCFA) for the years 1992 and 1993. Poisson regression was used in the analysis with unconstrained distributed lag models to examine the possibility that air pollution affects hospital admissions on not only the same day but on later days as well. The use of unconstrained distributed lags has the advantages of (1) not inappropriately biasing down risk estimates due to tight constraints (e.g. one day lag) and (2) not leaving the often arbitrary choice of lag period to the investigator's discretion. The C-R functions are based on the pooled estimate across all 14 cities, using the unconstrained distributed lag model and fixed or random effects estimates, depending on the results of a test for heterogeneity.

For this analysis, the unadjusted, base models for the effect of PM_{10} on hospital admissions were used. The authors performed a second-stage regression to estimate the impact of SO_2 and O_3 on the PM_{10} -

[59] Although Moolgavkar (2000c) reports results for the 20-64 year old age range, for comparability to other studies, we apply the results to the population of ages 18 to 64.

[60] In a log-linear model, the percent change is equal to (RR - 1) * 100. In this study, Moolgavkar defines and reports the "estimated" percent change as (log RR * 100). Because the relative risk is close to 1, RR-1 and log RR are essentially the same. For example, a true percent change of 2.0 would result in a relative risk of 1.020 and coefficient of 0.001980. The "estimated" percent change, as reported by Moolgavkar, of 2.0 results in a relative risk of 1.020201 and coefficient of 0.002.

[61] Moolgavkar (2000c) reports results for ICD codes 490-496. In order to avoid double counting non-elderly asthma hospitalizations (ICD code 493) with Sheppard et al. (1999) in a total benefits estimation, we have excluded ICD code 493 from the baseline incidence rate used in this function.

[62] The cities under investigation include: Birmingham, Boulder, Canton, Chicago, Colorado Springs, Detroit, Minneapolis/St. Paul, Nashville, New Haven, Pittsburgh, Provo/Orem, Seattle, Spokane, Youngstown.

hospitalization effect. For ozone, the PM_{10} effect in each city was regressed on the correlation between ozone and particulate matter (the slope of a PM_{10} vs. O_3 regression) in that city. The fitted line for this regression will have a slope of zero if there is no relationship, meaning that the effect of PM_{10} is not dependent on the correlation between PM_{10} and O_3. The adjusted point estimate was obtained by determining the PM_{10} effect when the correlation between the pollutants is zero (i.e. the y-intercept of the fitted line). The effect of O_3 adjustment on the PM_{10} - hospitalization relationship appeared to be minimal except for the case of COPD. In this case, adjustment increased the point estimate of the independent particulate matter effect. The variance of this estimate, however, was quite large and the confidence intervals of the adjusted and unadjusted estimates overlapped substantially. For these reasons, there appeared to be little impact of O_3 adjustment.[63] Furthermore, the statistical power and robustness of this second-stage approach to co-pollutant adjustment are in question because of the small number of observations used in the regression (14 cities) and the potential for one or two observations to dramatically impact the results.[64] Finally, for the case of COPD, adjustment led to an increased PM_{10} independent effect, meaning that if the adjustment is valid, the impact on hospital admissions will be underestimated rather than overestimated.

Single Pollutant Model

The estimated PM_{10} coefficient is based on a 2.88 percent increase (RR = 1.0288) in admissions due to a PM_{10} change of 10.0 $\mu g/m^3$ (Samet et al., 2000, Part II - Table 14)[65]. The standard error is estimated from the reported lower (0.19 percent) and upper bounds (5.64 percent) of the percent increase (Samet et al., 2000, Part II - Table 14).

Functional Form: Log-linear
Coefficient: 0.002839
Standard Error: 0.001351
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease per person 65+ (ICD codes 490-492, 494-496)
Population: population of ages 65 and older

F .4.17 Hospital Admissions for Chronic Lung Disease (less Asthma) (Schwartz, 1994b, Detroit)

Schwartz (1994b) examined the relationship between air pollution and hospital admissions (ICD codes 491-492, 494-496) for individuals 65 and older in Detroit, Michigan, from January 1986 to December 1989. In a two-pollutant Poisson regression model, Schwartz found both PM_{10} and ozone significantly linked to pneumonia and COPD. The authors state that effect estimates were relatively unchanged compared to the unreported single pollutant models. No significant associations were found between either pollutant and asthma admissions. The C-R function for chronic lung disease incidence is based on the results of the "basic" co-pollutant model (PM_{10} and ozone) presented in Table 4 (p. 651).

[63] Joel Schwartz (co-author), personal communication.

[64] Commentary from the Health Review Committee (Samet et al., 2000, p.77) states that "[w]hile the approach used in the morbidity analysis is novel...the question arises as to the adequacy of statistical power for performing these analyses."

[65] The random effects estimate of the unconstrained distributed lag model was chosen for COPD admissions since the chi-square test of heterogeneity was significant (see Samet et al., 2000, Part II - Table 15).

Multipollutant Model (PM_{10} and ozone)

The PM_{10} coefficient and standard error are reported in Table 4 (Schwartz, 1994b, p. 651) for a one $\mu g/m^3$ increase in daily average PM_{10} levels.

Functional Form: Log-linear
Coefficient: 0.00202
Standard Error: 0.00059
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease per person 65+ (ICD codes 490-492, 494-496)
Population: population of ages 65 and older

F .4.18 Hospital Admissions for Pneumonia (Burnett et al., 1999, Toronto)

Burnett et al. (1999) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada from 1980 to 1994. The authors examined single pollutant log-linear models for PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$, CO, NO_2, SO_2, and ozone and found all significantly associated with pneumonia and other respiratory infections (ICD codes 464,466,480-487,494). They estimated multi-pollutant models, where pollutants for the best fitting model were chosen using stepwise regression based on AIC criterion. Respiratory infection admissions were linked to ozone, NO_2, and $PM_{2.5}$. The C-R functions for $PM_{10-2.5}$ and PM_{10} are based on the results of a single pollutant model. The C-R functions for $PM_{2.5}$ are based on the results of a single pollutant model and three-pollutant model (ozone, NO_2, $PM_{2.5}$).

$PM_{2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (7.64) and t-statistic (6.09) reported in Table 3 (Burnett et al., 1999, p. 133) for an 18.0 $\mu g/m^3$ increase in three-day average $PM_{2.5}$ levels.

Functional Form: Log-linear
Coefficient: 0.004090
Standard Error: 0.000672
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person (ICD codes 464, 466, 480-487)
Population: population of all ages

Multipollutant Model ($PM_{2.5}$, NO_2, and ozone)

In a model with ozone and NO_2, the $PM_{2.5}$ coefficient and standard error are based on the percent increase (6.08) reported in Table 5 (Burnett et al., 1999, p. 135) and the t-statistic obtained from the authors (4.46)[66] for an 18.0 $\mu g/m^3$ increase in three-day average $PM_{2.5}$ levels.

[66] Rick Burnett (co-author), personal communication.

Functional Form: Log-linear
Coefficient: 0.003279
Standard Error: 0.000735
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person (ICD codes 464, 466, 480-487)
Population: population of all ages

$PM_{10-2.5}$ Function(s)

Single Pollutant Model

The coefficient and standard error are based on the percent increase (4.44) and t-statistic (4.00) reported in Table 3 (Burnett et al., 1999, p. 133) for a 12.2 $\mu g/m^3$ increase in three-day average $PM_{10-2.5}$ levels.

Functional Form: Log-linear
Coefficient: 0.003561
Standard Error: 0.000890
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person (ICD codes 464, 466, 480-487)
Population: population of all ages

PM_{10} Function(s)

Single Pollutant Model

The coefficient and standard error are based on the percent increase (8.35) and t-statistic (5.96) reported in Table 3 (Burnett et al., 1999, p. 133) for a 30.2 $\mu g/m^3$ increase in three-day average PM_{10} levels.

Functional Form: Log-linear
Coefficient: 0.002656
Standard Error: 0.000446
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person (ICD codes 464, 466, 480-487)
Population: population of all ages

F .4.19 Hospital Admissions for Pneumonia (Lippmann et al., 2000; Ito, 2003)

Lippmann et al. (2000) studied the association between particulate matter and daily mortality and hospitalizations among the elderly in Detroit, MI. Data were analyzed for two separate study periods, 1985-1990 and 1992-1994. The 1992-1994 study period had a greater variety of data on PM size and was the main focus of the report. The authors collected hospitalization data for a variety of cardiovascular and respiratory endpoints. They used daily air quality data for PM_{10}, $PM_{2.5}$, and $PM_{10-2.5}$ in a Poisson regression model with generalized additive models (GAM) to adjust for nonlinear relationships and temporal trends. In single pollutant models, all PM metrics were statistically significant for pneumonia (ICD codes 480-

486), $PM_{10-2.5}$ and PM_{10} were significant for ischemic heart disease (ICD code 410-414), and $PM_{2.5}$ and PM_{10} were significant for heart failure (ICD code 428). There were positive, but not statistically significant associations, between the PM metrics and COPD (ICD codes 490-496) and dysrhythmia (ICD code 427). In separate co-pollutant models with PM and either ozone, SO_2, NO_2, or CO, the results were generally comparable.

In response to concerns with the Splus issue, Ito (2003) reanalyzed the study by Lippmann et al. (2000). The reanalysis by Ito reported that more generalized additive models with stringent convergence criteria and generalized linear models resulted in smaller relative risk estimates. Not all of the original results were replicated, so we present here a mix of C-R functions from the reanalysis and from the original study (when the reanalyzed results were not available). The $PM_{2.5}$ C-R functions are based on results of the single pollutant model and co-pollutant model with ozone.

Single Pollutant Model (Ito, 2003)

The estimated $PM_{2.5}$ coefficient and standard error are based on a relative risk of 1.154 (95% CI - 1.027, 1.298) due to a $PM_{2.5}$ change of 36 $\mu g/m^3$ in the 1-day lag GAM stringent model (Ito, 2003, Table 7).

Functional Form: Log-linear
Coefficient: 0.003979
Standard Error: 0.001659
Incidence Rate: region-specific daily hospital admission rate for pneumonia admissions per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older

Multipollutant Model ($PM_{2.5}$ and ozone) (Lippmann et al., 2000)

The co-pollutant coefficient and standard error are calculated from a relative risk of 1.175 (95% CI 1.026-1.345) for a 36 $\mu g/m^3$ increase in $PM_{2.5}$ (Lippmann et al., 2000, Table 14, p. 26).

Functional Form: Log-linear
Coefficient: 0.004480
Standard Error: 0.001918
Incidence Rate: region-specific daily hospital admission rate for pneumonia admissions per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older

F .4.20 Hospital Admissions for Pneumonia (Moolgavkar et al., 1997, Minneapolis)

Moolgavkar et al. (1997) examined the relationship between air pollution and hospital admissions for individuals 65 and older in Minneapolis-St. Paul, Minnesota, from January 1986 to December 1991. In a four pollutant Poisson model examining pneumonia admissions (ICD codes 480-487) in Minneapolis, ozone was significant, while NO_2, SO_2, and PM_{10} were not significant. The PM_{10} C-R function is based on the results from the four-pollutant model to estimate pneumonia incidence. The model with a 130 df smoother was reported to be optimal (p. 368).

Multipollutant (PM_{10}, NO_2, ozone, and SO_2)

In a model with NO_2 and ozone, the estimated PM_{10} coefficient and standard error are based on a 1.00 percent increase in admissions (95% CI -1.0, 3.0) due to a PM_{10} change of 20 $\mu g/m^3$ (Moolgavkar et al., 1997, Table 4, p. 366)

Functional Form: Log-linear
Coefficient: 0.000498
Standard Error: 0.000505
Incidence Rate: region-specific daily hospital admission rate for pneumonia admissions per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older

F .4.21 Hospital Admissions for Pneumonia (Samet et al., 2000, 14 Cities)

Samet et al. (2000) examined the relationship between air pollution and hospital admissions for individuals of ages 65 and over in 14 cities across the country.[67] Cities were selected on the basis of available air pollution data for at least four years between 1985 and 1994 during which at least 50% of days had observations between the city-specific start and end of measurements. Hospital admissions were obtained from the Health Care Financing Administration (HCFA) for the years 1992 and 1993. Poisson regression was used in the analysis with unconstrained distributed lag models to examine the possibility that air pollution affects hospital admissions on not only the same day but on later days as well. The use of unconstrained distributed lags has the advantages of (1) not inappropriately biasing down risk estimates due to tight constraints (e.g. one day lag) and (2) not leaving the often arbitrary choice of lag period to the investigator's discretion. The C-R functions are based on the pooled estimate across all 14 cities, using the unconstrained distributed lag model and fixed or random effects estimates, depending on the results of a test for heterogeneity.

For this analysis, the unadjusted, base models for the effect of PM_{10} on hospital admissions were used. The authors performed a second-stage regression to estimate the impact of SO_2 and O_3 on the PM_{10} - hospitalization effect. For ozone, the PM_{10} effect in each city was regressed on the correlation between ozone and particulate matter (the slope of a PM_{10} vs. O_3 regression) in that city. The fitted line for this regression will have a slope of zero if there is no relationship, meaning that the effect of PM_{10} is not dependent on the correlation between PM_{10} and O_3. The adjusted point estimate was obtained by determining the PM_{10} effect when the correlation between the pollutants is zero (i.e. the y-intercept of the fitted line). The effect of O_3 adjustment on the PM_{10} - hospitalization relationship appeared to be minimal except for the case of COPD. In this case, adjustment increased the point estimate of the independent particulate matter effect. The variance of this estimate, however, was quite large and the confidence intervals of the adjusted and unadjusted estimates overlapped substantially. For these reasons, there appeared to be little impact of O_3 adjustment.[68] Furthermore, the statistical power and robustness of this second-stage approach to co-pollutant adjustment are in question because of the small number of observations used in the regression (14 cities) and the potential for one or two observations to dramatically

[67] The cities under investigation include: Birmingham, Boulder, Canton, Chicago, Colorado Springs, Detroit, Minneapolis/St. Paul, Nashville, New Haven, Pittsburgh, Provo/Orem, Seattle, Spokane, Youngstown.

[68] Joel Schwartz (co-author), personal communication.

impact the results.[69] Finally, for the case of COPD, adjustment led to an increased PM_{10} independent effect, meaning that if the adjustment is valid, the impact on hospital admissions will be underestimated rather than overestimated.

Single Pollutant Model

The estimated PM_{10} coefficient is based on a 2.07 percent increase (RR = 1.0207) in admissions due to a PM_{10} change of 10.0 $\mu g/m^3$ (Samet et al., 2000, Part II - Table 14)[70]. The standard error is estimated from the reported lower (0.94 percent) and upper bounds (3.22 percent) of the percent increase (Samet et al., 2000, Part II - Table 14).

Functional Form: Log-linear
Coefficient: 0.002049
Standard Error: 0.000570
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older

F .4.22 Hospital Admissions for Pneumonia (Schwartz, 1994a, Minneapolis)

Schwartz (1994a) examined the relationship between air pollution and hospital admissions for individuals 65 and older in Minneapolis-St. Paul, Minnesota, from January 1986 to December 1989. In single-pollutant Poission regression models, both ozone and PM_{10} were significantly associated with pneumonia admissions. In a two-pollutant model, Schwartz found PM_{10} significantly related to pneumonia; ozone was weakly linked to pneumonia. The results were not sensitive to the methods used to control for seasonal patterns and weather. The PM_{10} C-R functions are based on the results of the single pollutant model and the two-pollutant model (PM_{10} and ozone) with "spline" smoothing.

Single Pollutant Model

The single pollutant coefficient and standard error are based on the relative risk (1.17) and 95% confidence interval (1.03-1.33) for a 100 $\mu g/m^3$ increase in daily average PM_{10} levels (Schwartz, 1994a, p. 369).

Functional Form: Log-linear
Coefficient: 0.001570
Standard Error: 0.000652
Incidence Rate: region-specific daily hospital admission rate for pneumonia admissions per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older.

[69] Commentary from the Health Review Committee (Samet et al., 2000, p.77) states that "[w]hile the approach used in the morbidity analysis is novel...the question arises as to the adequacy of statistical power for performing these analyses."

[70] The random effects estimate of the unconstrained distributed lag model was chosen for pneumonia admissions since the chi-square test of heterogeneity was significant (see Samet et al., 2000, Part II - Table 15).

Multipollutant Model (PM_{10} and ozone)

In a model with ozone and spline functions to adjust for time and weather, the coefficient and standard error are based on the relative risk (1.18) and 95% confidence interval (1.03, 1.36) for a 100 $\mu g/m^3$ increase in daily average PM_{10} levels (Schwartz, 1994a, Table 4).

Functional Form: Log-linear
Coefficient: 0.001655
Standard Error: 0.000709
Incidence Rate: region-specific daily hospital admission rate for pneumonia admissions per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older

F .4.23 Hospital Admissions for Pneumonia (Schwartz, 1994b, Detroit)

Schwartz (1994b) examined the relationship between air pollution and hospital admissions for individuals 65 and older in Detroit, Michigan, from January 1986 to December 1989. In a two-pollutant Poisson regression model, Schwartz found both PM_{10} and ozone significantly linked to pneumonia and COPD. The authors state that effect estimates were relatively unchanged compared to the unreported single pollutant models. No significant associations were found between either pollutant and asthma admissions. The PM_{10} C-R function for pneumonia incidence is based on results of the co-pollutant model (PM_{10} and ozone).

Multipollutant Model (PM_{10} and ozone)

The PM_{10} coefficient and standard error are reported in Table 4 (Schwartz, 1994b, p. 651) for a one $\mu g/m^3$ increase in daily average PM_{10} levels.

Functional Form: Log-linear
Coefficient: 0.00115
Standard Error: 0.00039
Incidence Rate: region-specific daily hospital admission rate for pneumonia admissions per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older

F .4.24 Hospital Admissions for All Cardiovascular (Burnett et al., 1997, Toronto)

Burnett et al. (1997) examined the relationship between air pollution and cardiac hospital admissions (ICD codes 410-414,427,428) for individuals of all ages in Toronto, Canada during the summers of 1992-1994. In a Poisson regression, cardiac admissions were linked to coefficient of haze (COH) and ozone; other PM measures were less strongly linked. In two pollutant models, they found that CO, NO_2, and SO_2 were not significant, controlling for COH. They found that ozone was still significant, controlling for COH. In multi-pollutant models with COH, ozone, NO_2, and SO_2, both ozone and COH remained significant. None of the other PM measures (PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$) were significant in four-pollutant models. PM C-R functions are based on the results of single and multipollutant models.

$PM_{2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model, the $PM_{2.5}$ coefficient and standard error are based on a relative risk of 1.031 (t-statistic 1.8) for an 11 $\mu g/m^3$ increase in four-day average $PM_{2.5}$ (Burnett et al., 1997, Table 2, p. 617).

Functional Form: Log-linear
Coefficient: 0.002775
Standard Error: 0.001542
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

Multipollutant Model ($PM_{2.5}$ and ozone)

In a model with ozone, the $PM_{2.5}$ coefficient and standard error are based on a relative risk of 1.014 (t-statistic 0.78) for an 11 $\mu g/m^3$ increase in four-day average $PM_{2.5}$ (Burnett et al., 1997, Table 5, p. 618).

Functional Form: Log-linear
Coefficient: 0.001264
Standard Error: 0.001620
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

Multipollutant Model ($PM_{2.5}$, NO_2, ozone, and SO_2)

In a four-pollutant model with NO_2, ozone, and SO_2, the $PM_{2.5}$ coefficient and standard error are based on a relative risk of 0.993 (t-statistic 0.33) for an 11 $\mu g/m^3$ increase in four-day average $PM_{2.5}$ (Burnett et al., 1997, Table 6, p. 618).

Functional Form: Log-linear
Coefficient: -0.000639
Standard Error: 0.001935
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

$PM_{10-2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model, the $PM_{10-2.5}$ coefficient and standard error are based on a relative risk of 1.036 (t-statistic 3.41) for a 4.75 $\mu g/m^3$ increase in four-day average $PM_{10-2.5}$ (Burnett et al., 1997, Table 2, p. 617).

Functional Form: Log-linear
Coefficient: 0.007446
Standard Error: 0.002183
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

Multipollutant Model ($PM_{10-2.5}$ and ozone)

In a model with ozone, the $PM_{10-2.5}$ coefficient and standard error are based on a relative risk of 1.034 (t-statistic 3.28) for a 4.75 $\mu g/m^3$ increase in four-day average $PM_{10-2.5}$ (Burnett et al., 1997, Table 5, p. 618).

Functional Form: Log-linear
Coefficient: 0.007039
Standard Error: 0.002146
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

Multipollutant Model ($PM_{10-2.5}$, NO_2, ozone, and SO_2)

In a four-pollutant model with NO_2, ozone, and SO_2, the $PM_{10-2.5}$ coefficient and standard error are based on a relative risk of 1.022 (t-statistic 1.68) for a 4.75 $\mu g/m^3$ increase in four-day average $PM_{10-2.5}$ (Burnett et al., 1997, Table 6, p. 618).

Functional Form: Log-linear
Coefficient: 0.004581
Standard Error: 0.002727
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

PM_{10} Function(s)

Single Pollutant Model

In a single pollutant model, the PM_{10} coefficient and standard error are based on a relative risk of 1.033 (t-statistic 2.24) for a 14.25 $\mu g/m^3$ increase in four-day average PM_{10} (Burnett et al., 1997, Table 2, p. 617).

Functional Form: Log-linear
Coefficient: 0.002278
Standard Error: 0.001017
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

Multipollutant Model (PM_{10} and ozone)

In a model with ozone, the PM_{10} coefficient and standard error are based on a relative risk of 1.025 (t-statistic 1.68) for a 14.25 $\mu g/m^3$ increase in four-day average PM_{10} (Burnett et al., 1997, Table 5, p. 618).

Functional Form: Log-linear
Coefficient: 0.001733
Standard Error: 0.001031
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

Multipollutant Model (PM_{10}, NO_2, ozone, and SO_2)

In a four-pollutant model with NO_2, ozone, and SO_2, the PM_{10} coefficient and standard error are based on a relative risk of 0.996 (t-statistic 0.23) for a 14.25 $\mu g/m^3$ increase in four-day average PM_{10} (Burnett et al., 1997, Table 6, p. 618).

Functional Form: Log-linear
Coefficient: -0.000281
Standard Error: 0.001223
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

F .4.25 Hospital Admissions for All Cardiovascular (Moolgavkar, 2000b; Moolgavkar, 2003)

Moolgavkar (2000a) examined the association between air pollution and cardiovascular hospital admissions (ICD 390-448) in the Chicago, Los Angeles, and Phoenix metropolitan areas. He collected daily air pollution data for ozone, SO_2, NO_2, CO, and PM_{10} in all three areas. $PM_{2.5}$ data was available only in Los Angeles. The data were analyzed using a Poisson regression model with generalized additive models to adjust for temporal trends. Separate models were run for 0 to 5 day lags in each location. Among the 65+ age group, the gaseous pollutants generally exhibited stronger effects than PM_{10} or $PM_{2.5}$. The strongest overall effects were observed for SO_2 and CO. In a single pollutant model, $PM_{2.5}$ was statistically significant for lag 0 and lag 1. In co-pollutant models with CO, the $PM_{2.5}$ effect dropped out and CO remained significant. For ages 20-64, SO_2 and CO exhibited the strongest effect and any $PM_{2.5}$ effect dropped out in co-pollutant models with CO.

In response to concerns with the Splus issue, Moolgavkar (2003) reanalyzed his earlier study. In the reanalysis, he reported that more generalized additive models with stringent convergence criteria and generalized linear models resulted in smaller relative risk estimates. Not all of the original results were replicated, so we present here a mix of C-R functions from the reanalysis and from the original study (when the reanalyzed results were not available). The $PM_{2.5}$ C-R functions are based on single pollutant and co-pollutant ($PM_{2.5}$ and CO) models.

Ages 65 and older

Single Pollutant Model (Moolgavkar, 2003)

The single pollutant coefficient and standard error are calculated from an estimated percent change of 1.58[71] and t-statistic of 4.59 for a 10 μg/m^3 increase in $PM_{2.5}$ in the 0-day lag GAM-30df stringent (10^{-8}) model (Moolgavkar, 2003, Table 12).

Functional Form: Log-linear
Coefficient: 0.001568
Standard Error: 0.000342
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular admissions per person 65+ (ICD codes 390-429)
Population: population of ages 65 and older.

Multipollutant Model ($PM_{2.5}$ and CO) (Moolgavkar, 2003)

In a model with $PM_{2.5}$ and CO, the single pollutant coefficient and standard error are calculated from an estimated percent change of 0.39[59] and t-statistic of 0.92 for a 10 μg/m^3 increase in $PM_{2.5}$ in the 0-day lag GAM-100df stringent (10^{-8}) model (Moolgavkar, 2003, Table 14).

Functional Form: Log-linear
Coefficient: 0.000389
Standard Error: 0.000423
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular admissions per person 65+ (ICD codes 390-429)
Population: population of ages 65 and older

Ages 18 to 64[72]

Single Pollutant Model (Moolgavkar, 2000b)

The single pollutant coefficient and standard error are calculated from an estimated percent change of 1.4[73] and t-statistic of 4.1 for a 10 μg/m^3 increase in $PM_{2.5}$ in the zero lag model (Moolgavkar, 2000a, Table 4, p. 1203).

---

[71] In a log-linear model, the percent change is equal to (RR - 1) * 100. In this study, Moolgavkar defines and reports the "estimated" percent change as (log RR * 100). Because the relative risk is close to 1, RR-1 and log RR are essentially the same. For example, a true percent change of 2.2 would result in a relative risk of 1.022 and coefficient of 0.002176. The "estimated" percent change, as reported by Moolgavkar, of 2.2 results in a relative risk of 1.022244 and coefficient of 0.0022.

[72] Although Moolgavkar (2000a) reports results for the 20-64 year old age range, for comparability to other studies, we apply the results to the population of ages 18 to 64.

[73] In a log-linear model, the percent change is equal to (RR - 1) * 100. In a similar hospitalization study by Moolgavkar (2000c), he defines and reports the "estimated" percent change as (log RR * 100). Because the relative risk is close to 1, RR-1 and log RR are essentially the same. For example, a true percent change of 1.4 would result in a relative risk of 1.014 and coefficient of 0.00139. Assuming that the 1.4 is the "estimated" percent change described previously would result in a relative risk of 1.014098 and coefficient of 0.0014. We assume that the "estimated" percent changes reported in this study reflect the definition from (Moolgavkar, 2000c).

Functional Form: Log-linear
Coefficient: 0.0014
Standard Error: 0.000341
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular admissions per person ages 18 to 64 (ICD codes 390-409, 411-459)[74]
Population: population of ages 18 to 64

Multipollutant Model ($PM_{2.5}$ and CO) (Moolgavkar, 2000b)

In a model with CO, the coefficient and standard error are calculated from an estimated percent change of 0.9[75] and t-statistic of 1.8 for a 10 $\mu g/m^3$ increase in $PM_{2.5}$ in the zero lag model (Moolgavkar, 2000a, Table 4, p. 1203).

Functional Form: Log-linear
Coefficient: 0.0009
Standard Error: 0.000500
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular admissions per person ages 18 to 64 (ICD codes 390-409, 411-459)[76]
Population: population of ages 18 to 64

F .4.26 Hospital Admissions for All Cardiovascular (Samet et al., 2000, 14 Cities)

Samet et al. (2000) examined the relationship between air pollution and hospital admissions for individuals of ages 65 and over in 14 cities across the country.[77] Cities were selected on the basis of available air pollution data for at least four years between 1985 and 1994 during which at least 50% of days had observations between the city-specific start and end of measurements. Hospital admissions were obtained from the Health Care Financing Administration (HCFA) for the years 1992 and 1993. Poisson regression was used in the analysis with unconstrained distributed lag models to examine the possibility that air pollution affects hospital admissions on not only the same day but on later days as well. The use of unconstrained distributed lags has the advantages of (1) not inappropriately biasing down risk estimates

[74] Moolgavkar (2000a) reports results that include ICD code 410 (heart attack). In the benefits analysis, avoided nonfatal heart attacks are estimated using the results reported by Peters et al. (2001). The baseline rate in the Peters et al. function is a modified heart attack hospitalization rate (ICD code 410), since most, if not all, nonfatal heart attacks will require hospitalization. In order to avoid double counting heart attack hospitalizations, we have excluded ICD code 410 from the baseline incidence rate used in this function.

[75] In a log-linear model, the percent change is equal to (RR - 1) * 100. In a similar hospitalization study by Moolgavkar (2000c), he defines and reports the "estimated" percent change as (log RR * 100). Because the relative risk is close to 1, RR-1 and log RR are essentially the same. For example, a true percent change of 0.9 would result in a relative risk of 1.009 and coefficient of 0.000896. Assuming that the 0.9 is the "estimated" percent change described previously would result in a relative risk of 1.009041 and coefficient of 0.0009. We assume that the "estimated" percent changes reported in this study reflect the definition from (Moolgavkar, 2000c).

[76] Moolgavkar (2000a) reports results that include ICD code 410 (heart attack). In the benefits analysis, avoided nonfatal heart attacks are estimated using the results reported by Peters et al. (2001). The baseline rate in the Peters et al. function is a modified heart attack hospitalization rate (ICD code 410), since most, if not all, nonfatal heart attacks will require hospitalization. In order to avoid double counting heart attack hospitalizations, we have excluded ICD code 410 from the baseline incidence rate used in this function.

[77] The cities under investigation include: Birmingham, Boulder, Canton, Chicago, Colorado Springs, Detroit, Minneapolis/St. Paul, Nashville, New Haven, Pittsburgh, Provo/Orem, Seattle, Spokane, Youngstown.

due to tight constraints (e.g. one day lag) and (2) not leaving the often arbitrary choice of lag period to the investigator's discretion. The C-R functions are based on the pooled estimate across all 14 cities, using the unconstrained distributed lag model and fixed or random effects estimates, depending on the results of a test for heterogeneity.

For this analysis, the unadjusted, base models for the effect of PM_{10} on hospital admissions were used. The authors performed a second-stage regression to estimate the impact of SO_2 and O_3 on the PM_{10} - hospitalization effect. For ozone, the PM_{10} effect in each city was regressed on the correlation between ozone and particulate matter (the slope of a PM_{10} vs. O_3 regression) in that city. The fitted line for this regression will have a slope of zero if there is no relationship, meaning that the effect of PM_{10} is not dependent on the correlation between PM_{10} and O_3. The adjusted point estimate was obtained by determining the PM_{10} effect when the correlation between the pollutants is zero (i.e. the y-intercept of the fitted line). The effect of O_3 adjustment on the PM_{10} - hospitalization relationship appeared to be minimal except for the case of COPD. In this case, adjustment increased the point estimate of the independent particulate matter effect. The variance of this estimate, however, was quite large and the confidence intervals of the adjusted and unadjusted estimates overlapped substantially. For these reasons, there appeared to be little impact of O_3 adjustment.[78] Furthermore, the statistical power and robustness of this second-stage approach to co-pollutant adjustment are in question because of the small number of observations used in the regression (14 cities) and the potential for one or two observations to dramatically impact the results.[79] Finally, for the case of COPD, adjustment led to an increased PM_{10} independent effect, meaning that if the adjustment is valid, the impact on hospital admissions will be underestimated rather than overestimated.

Single Pollutant Model

The estimated PM_{10} coefficient is based on a 1.19 percent increase (RR = 1.0119) in admissions due to a PM_{10} change of 10.0 $\mu g/m^3$ (Samet et al., 2000, Part II - Table 14)[80]. The standard error is estimated from the reported lower (0.97 percent) and upper bounds (1.41 percent) of the percent increase (Samet et al., 2000, Part II - Table 14).

Functional Form: Log-linear
Coefficient: 0.001183
Standard Error: 0.000111
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person 65+ (ICD codes 390-459)
Population: population of ages 65 and older

F .4.27 Hospital Admissions for Dysrhythmias (Burnett et al., 1999, Toronto)

Burnett et al. (1999) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada from1980 to 1994. The authors examined single pollutant log-linear models for PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$, CO, NO_2, SO_2, and O_3 and found $PM_{2.5}$, PM_{10}, and CO significantly

[78] Joel Schwartz (co-author), personal communication.

[79] Commentary from the Health Review Committee (Samet et al., 2000, p.77) states that "[w]hile the approach used in the morbidity analysis is novel...the question arises as to the adequacy of statistical power for performing these analyses."

[80] The fixed effects estimate of the unconstrained distributed lag model was chosen for CVD admissions since the chi-square test of heterogeneity was non-significant (see Samet et al., 2000, Part II - Table 15).

associated with admissions. They estimated multiple pollutant models, where pollutants for best fitting model were chosen using stepwise regression based on AIC criterion. The final model for dysrhythmias admissions included O_3, CO, and $PM_{2.5}$. CO was significantly associated with admissions, while O_3 and $PM_{2.5}$ were marginally significant. The C-R functions are based on the reported single pollutant and multipollutant models for $PM_{2.5}$ and single pollutant models for other PM measures.

$PM_{2.5}$ Function(s)

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (4.33) and t-statistic (2.91) reported in Table 3 (Burnett et al., 1999, p. 133) for an 18.0 μg/m^3 increase in daily average $PM_{2.5}$ concentration.

Functional Form: Log-linear
Coefficient: 0.002355
Standard Error: 0.000809
Incidence Rate: region-specific daily hospital admission rate for dysrhythmia disease per person (ICD code 427)
Population: population of all ages

Multipollutant Model ($PM_{2.5}$, CO, and ozone)

In a model with ozone and CO, the $PM_{2.5}$ coefficient and standard error are based on the percent increase (2.47) reported in Table 5 (Burnett et al., 1999, p. 135) and the t-statistic obtained from the authors (1.49)[81] for an 18.0 μg/m^3 increase in daily average $PM_{2.5}$ concentration.

Functional Form: Log-linear
Coefficient: 0.001356
Standard Error: 0.000910
Incidence Rate: region-specific daily hospital admission rate for dysrhythmia disease per person (ICD code 427)
Population: population of all ages

$PM_{10-2.5}$ Function(s)

Single Pollutant Model

The coefficient and standard error are based on the percent increase (2.47) and t-statistic (1.88) reported in Table 3 (Burnett et al., 1999, p. 133) for a 12.2 μg/m^3 increase in daily average $PM_{10-2.5}$ concentration.

[81] Rick Burnett (co-author), personal communication.

Functional Form: Log-linear
Coefficient: 0.002000
Standard Error: 0.001064
Incidence Rate: region-specific daily hospital admission rate for dysrhythmia disease per person (ICD code 427)
Population: population of all ages

PM_{10} Function(s)

Single Pollutant Model

The coefficient and standard error are based on the percent increase (5.00) and t-statistic (4.25) reported in Table 3 (Burnett et al., 1999, p. 133) for a 30.2 $\mu g/m^3$ increase in daily average PM_{10} concentration.

Functional Form: Log-linear
Coefficient: 0.001616
Standard Error: 0.000533
Incidence Rate: region-specific daily hospital admission rate for dysrhythmia disease per person (ICD code 427)
Population: population of all ages

F .4.28 Hospital Admissions for Dysrhythmia (Lippmann et al., 2000; Ito, 2003)

Lippmann et al. (2000) studied the association between particulate matter and daily mortality and hospitalizations among the elderly in Detroit, MI. Data were analyzed for two separate study periods, 1985-1990 and 1992-1994. The 1992-1994 study period had a greater variety of data on PM size and was the main focus of the report. The authors collected hospitalization data for a variety of cardiovascular and respiratory endpoints. They used daily air quality data for PM_{10}, $PM_{2.5}$, and $PM_{10-2.5}$ in a Poisson regression model with generalized additive models (GAM) to adjust for nonlinear relationships and temporal trends. In single pollutant models, all PM metrics were statistically significant for pneumonia (ICD codes 480-486), $PM_{10-2.5}$ and PM_{10} were significant for ischemic heart disease (ICD code 410-414), and $PM_{2.5}$ and PM_{10} were significant for heart failure (ICD code 428). There were positive, but not statistically significant associations, between the PM metrics and COPD (ICD codes 490-496) and dysrhythmia (ICD code 427). In separate co-pollutant models with PM and either ozone, SO_2, NO_2, or CO, the results were generally comparable.

In response to concerns with the Splus issue, Ito (2003) reanalyzed the study by Lippmann et al. (2000). The reanalysis by Ito reported that more generalized additive models with stringent convergence criteria and generalized linear models resulted in smaller relative risk estimates. Not all of the original results were replicated, so we present here a mix of C-R functions from the reanalysis and from the original study (when the reanalyzed results were not available).

Single Pollutant Model (Ito, 2003)

The co-pollutant coefficient and standard error are calculated from a relative risk of 1.046 (95% CI 0.906-1.207) for a 36 $\mu g/m^3$ increase in $PM_{2.5}$ in the 1-day lag GAM stringent model (Ito, 2003, Table 10).

Functional Form: Log-linear
Coefficient: 0.001249
Standard Error: 0.002033
Incidence Rate: region-specific daily hospital admission rate for dysrhythmia admissions per person 65+ (ICD code 427)
Population: population of ages 65 and older

Multipollutant Model ($PM_{2.5}$ and ozone) (Lippmann et al., 2000)

The co-pollutant coefficient and standard error are calculated from a relative risk of 1.080 (95% CI 0.904-1.291) for a 36 $\mu g/m^3$ increase in $PM_{2.5}$ (Lippmann et al., 2000, Table 14, p. 27).

Functional Form: Log-linear
Coefficient: 0.002138
Standard Error: 0.002525
Incidence Rate: region-specific daily hospital admission rate for dysrhythmia admissions per person 65+ (ICD code 427)
Population: population of ages 65 and older

F .4.29 Hospital Admissions for Congestive Heart Failure (Lippmann et al., 2000; Ito, 2003)

Lippmann et al. (2000) studied the association between particulate matter and daily mortality and hospitalizations among the elderly in Detroit, MI. Data were analyzed for two separate study periods, 1985-1990 and 1992-1994. The 1992-1994 study period had a greater variety of data on PM size and was the main focus of the report. The authors collected hospitalization data for a variety of cardiovascular and respiratory endpoints. They used daily air quality data for PM_{10}, $PM_{2.5}$, and $PM_{10-2.5}$ in a Poisson regression model with generalized additive models (GAM) to adjust for nonlinear relationships and temporal trends. In single pollutant models, all PM metrics were statistically significant for pneumonia (ICD codes 480-486), $PM_{10-2.5}$ and PM_{10} were significant for ischemic heart disease (ICD code 410-414), and $PM_{2.5}$ and PM_{10} were significant for congestive heart failure (ICD code 428). There were positive, but not statistically significant associations, between the PM metrics and COPD (ICD codes 490-496) and dysrhythmia (ICD code 427). In separate co-pollutant models with PM and either ozone, SO_2, NO_2, or CO, the results were generally comparable.

In response to concerns with the Splus issue, Ito (2003) reanalyzed the study by Lippmann et al. (2000). The reanalysis by Ito reported that more generalized additive models with stringent convergence criteria and generalized linear models resulted in smaller relative risk estimates. Not all of the original results were replicated, so we present here a mix of C-R functions from the reanalysis and from the original study (when the reanalyzed results were not available).

Single Pollutant Model (Ito, 2003)

The co-pollutant coefficient and standard error are calculated from a relative risk of 1.117 (95% CI 1.020-1.224) for a 36 $\mu g/m^3$ increase in $PM_{2.5}$ in the 1-day lag GAM stringent model (Ito, 2003, Table 11).

Functional Form: Log-linear
Coefficient: 0.003469
Standard Error: 0.001293
Incidence Rate: region-specific daily hospital admission rate for congestive heart failure admissions per person 65+ (ICD code 428)
Population: population of ages 65 and older

Multipollutant Model ($PM_{2.5}$ and ozone) (Lippmann et al., 2000)

The co-pollutant coefficient and standard error are calculated from a relative risk of 1.183 (95% CI 1.053-1.329) for a 36 $\mu g/m^3$ increase in $PM_{2.5}$ (Lippmann et al., 2000, Table 14, p. 27).

Functional Form: Log-linear
Coefficient: 0.004668
Standard Error: 0.001650
Incidence Rate: region-specific daily hospital admission rate for congestive heart failure admissions per person 65+ (ICD code 428)
Population: population of ages 65 and older

F .4.30 Hospital Admissions for Ischemic Heart Disease (Lippmann et al., 2000; Ito, 2003)

Lippmann et al. (2000) studied the association between particulate matter and daily mortality and hospitalizations among the elderly in Detroit, MI. Data were analyzed for two separate study periods, 1985-1990 and 1992-1994. The 1992-1994 study period had a greater variety of data on PM size and was the main focus of the report. The authors collected hospitalization data for a variety of cardiovascular and respiratory endpoints. They used daily air quality data for PM_{10}, $PM_{2.5}$, and $PM_{10-2.5}$ in a Poisson regression model with generalized additive models (GAM) to adjust for nonlinear relationships and temporal trends. In single pollutant models, all PM metrics were statistically significant for pneumonia (ICD codes 480-486), $PM_{10-2.5}$ and PM_{10} were significant for ischemic heart disease (ICD code 410-414), and $PM_{2.5}$ and PM_{10} were significant for heart failure (ICD code 428). There were positive, but not statistically significant associations, between the PM metrics and COPD (ICD codes 490-496) and dysrhythmia (ICD code 427). In separate co-pollutant models with PM and either ozone, SO_2, NO_2, or CO, the results were generally comparable.

In response to concerns with the Splus issue, Ito (2003) reanalyzed the study by Lippmann et al. (2000). The reanalysis by Ito reported that more generalized additive models with stringent convergence criteria and generalized linear models resulted in smaller relative risk estimates. Not all of the original results were replicated, so we present here a mix of C-R functions from the reanalysis and from the original study (when the reanalyzed results were not available).

Single Pollutant Model (Ito, 2003)

The co-pollutant coefficient and standard error are calculated from a relative risk of 1.053 (95% CI 0.971-1.143) for a 36 $\mu g/m^3$ increase in $PM_{2.5}$ in the 1-day lag GAM stringent model (Ito, 2003, Table 9)

Functional Form: Log-linear
Coefficient: 0.001435
Standard Error: 0.001156
Incidence Rate: region-specific daily hospital admission rate for ischemic heart disease admissions per person 65+ (ICD codes 411-414)[82]
Population: population of ages 65 and older

Multipollutant Model ($PM_{2.5}$ and ozone) (Lippmann et al., 2000)

The co-pollutant coefficient and standard error are calculated from a relative risk of 1.041 (95% CI 0.947-1.144) for a 36 $\mu g/m^3$ increase in $PM_{2.5}$ (Lippmann et al., 2000, Table 14, p. 27).

Functional Form: Log-linear
Coefficient: 0.001116
Standard Error: 0.001339
Incidence Rate: region-specific daily hospital admission rate for ischemic heart disease admissions per person 65+ (ICD codes 411-414)[83]
Population: population of ages 65 and older

[82] Lippmann et al. (2000) reports results for ICD codes 410-414. In the benefits analysis, avoided nonfatal heart attacks are estimated using the results reported by Peters et al. (2001). The baseline rate in the Peters et al. function is a modified heart attack hospitalization rate (ICD code 410), since most, if not all, nonfatal heart attacks will require hospitalization. In order to avoid double counting heart attack hospitalizations, we have excluded ICD code 410 from the baseline incidence rate used in this function.

[83] Lippmann et al. (2000) reports results for ICD codes 410-414. In the benefits analysis, avoided nonfatal heart attacks are estimated using the results reported by Peters et al. (2001). The baseline rate in the Peters et al. function is a modified heart attack hospitalization rate (ICD code 410), since most, if not all, nonfatal heart attacks will require hospitalization. In order to avoid double counting heart attack hospitalizations, we have excluded ICD code 410 from the baseline incidence rate used in this function.

Appendix F. Particulate Matter C-R Functions

Exhibit F-5. Concentration-Response (C-R) Functions for Particulate Matter and Emergency Room Visits

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Beta	Std Error	Functional Form
Asthma	$PM_{2.5}$	Norris et al.	1999	Seattle, WA	<18	All	All	None	24-hr avg	0.014712	0.003492	Log-linear
Asthma	$PM_{2.5}$	Norris et al.	1999	Seattle, WA	<18	All	All	NO_2, SO_2	24-hr avg	0.016527	0.004139	Log-linear
Asthma	PM_{10}	Norris et al.	1999	Seattle, WA	<18	All	All	None	24-hr avg	0.011296	0.003480	Log-linear
Asthma	PM_{10}	Norris et al.	1999	Seattle, WA	<18	All	All	NO_2, SO_2	24-hr avg	0.011296	0.004220	Log-linear
Asthma	PM_{10}	Schwartz et al.	1993	Seattle, WA	<65	All	All	None	24-hr avg	0.00367	0.00126	Log-linear

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.

F .5 Emergency Room Visits

F .5.1 Emergency Room Visits for Asthma (Norris et al., 1999)

Norris et al. (1999) examined the relation between air pollution in Seattle and childhood (<18) hospital admissions for asthma from 1995 to 1996. The authors used air quality data for PM_{10}, light scattering (used to estimate fine PM), CO, SO_2, NO_2, and O_3 in a Poisson regression model with adjustments for day of the week, time trends, temperature, and dew point. They found significant associations between asthma ER visits and light scattering (converted to $PM_{2.5}$), PM_{10}, and CO. No association was found between O_3, NO_2, or SO_2 and asthma ER visits, although O_3 had a significant amount of missing data. In multipollutant models with either PM metric (light scattering or PM_{10}) and NO_2 and SO_2, the PM coefficients remained significant while the gaseous pollutants were not associated with increased asthma ER visits. The PM C-R functions are based on results of the single and multipollutant models reported.

$PM_{2.5}$ Function(s)

Single Pollutant Model

The single pollutant coefficient and standard error are calculated from a relative risk of 1.15 (95% CI 1.08-1.23) for a 9.5 μg/m^3 increase in $PM_{2.5}$ (Norris et al., 1999, Table 4, p. 492).

Functional Form: Log-linear
Coefficient: 0.014712
Standard Error: 0.003492
Incidence Rate: region-specific daily emergency room rate for asthma admissions per person <18 (ICD code 493)
Population: population of ages under 18

Multipollutant Model ($PM_{2.5}$, NO_2, and SO_2)

In a model with NO_2 and SO_2, the $PM_{2.5}$ coefficient and standard error are calculated from a relative risk of 1.17 (95% CI 1.08-1.26) for a 9.5 μg/m^3 increase in $PM_{2.5}$ (Norris et al., 1999, p. 491).

Functional Form: Log-linear
Coefficient: 0.016527
Standard Error: 0.004139
Incidence Rate: region-specific daily emergency room rate for asthma admissions per person <18 (ICD code 493)
Population: population of ages under 18

PM_{10} Function(s)

Single Pollutant Model

The single pollutant coefficient and standard error are calculated from a relative risk of 1.14 (95% CI 1.05-1.23) for an 11.6 μg/m^3 increase in PM_{10} (Norris et al., 1999, Table 4, p. 492).

Functional Form: Log-linear
Coefficient: 0.011296
Standard Error: 0.003480
Incidence Rate: region-specific daily emergency room rate for asthma admissions per person <18 (ICD code 493)
Population: population of ages under 18

Multipollutant Model (PM_{10}, NO_2, and SO_2)

In a model with NO_2 and SO_2, the PM_{10} coefficient and standard error are calculated from a relative risk of 1.14 (95% CI 1.04-1.26) for an 11.6 $\mu g/m^3$ increase in PM_{10} (Norris et al., 1999, p. 491).

Functional Form: Log-linear
Coefficient: 0.011296
Standard Error: 0.004220
Incidence Rate: region-specific daily emergency room rate for asthma admissions per person <18 (ICD code 493)
Population: population of ages under 18

F .5.2 Emergency Room Visits for Asthma (Schwartz et al., 1993, Seattle)

Schwartz et al. (1993) examined the relationship between air quality and emergency room visits for asthma (ICD codes 493,493.01,493.10,493.90,493.91) in persons under 65 and 65 and over, living in Seattle from September 1989 to September 1990. Using single-pollutant models they found daily levels of PM_{10} linked to ER visits in individuals ages under 65, and they found no effect in individuals ages 65 and over. They did not find a significant effect for SO_2 and ozone in either age group. The results of the single pollutant model for PM_{10} are used in this analysis.

Single Pollutant Model

The PM_{10} coefficient and standard error are reported by Schwartz et al. (1993, p. 829) for a unit $\mu g/m^3$ increase in four-day average PM_{10} levels.

Functional Form: Log-linear
Coefficient: 0.00367
Standard Error: 0.00126
Incidence Rate: region-specific daily emergency room rate for asthma admissions per person <65 (ICD code 493)
Population: population of ages under 65

Appendix F. Particulate Matter C-R Functions

Exhibit F-6. Concentration-Response (C-R) Functions for Particulate Matter and Acute Effects

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Beta	Std Error	Functional Form
Acute Bronchitis	$PM_{2.5}$	Dockery et al.	1996	24 communities	8-12	All	All	None	Annual Avg	0.027212	0.017096	Logistic
Acute Myocardial Infarction, Nonfatal	$PM_{2.5}$	Peters et al.	2001	Boston, MA	18+	All	All	None	24-hr avg	0.024121	0.009285	Logistic
Acute Myocardial Infarction, Nonfatal	$PM_{10-2.5}$	Peters et al.	2001	Boston, MA	18+	All	All	None	24-hr avg	0.021954	0.015000	Logistic
Acute Myocardial Infarction, Nonfatal	PM_{10}	Peters et al.	2001	Boston, MA	18+	All	All	None	24-hr avg	0.016894	0.006870	Logistic
Any of 19 Respiratory Symptoms	PM_{10}	Krupnick	1990	Los Angeles, CA	18-64	All	All	O_3	24-hr avg	0.000461	0.000239	Linear
Lower Respiratory Symptoms	$PM_{2.5}$	Schwartz and Neas	2000	6 cities	7-14	All	All	None	24-hr avg	0.019012	0.006005	Logistic
Lower Respiratory Symptoms	$PM_{2.5}$	Schwartz and Neas	2000	6 cities	7-14	All	All	$PM_{10-2.5}$	24-hr avg	0.016976	0.006680	Logistic
Lower Respiratory Symptoms	$PM_{2.5}$	Schwartz et al.	1994	6 cities	7-14	All	All	None	24-hr avg	0.01823	0.00586	Logistic
Minor Restricted Activity Days	$PM_{2.5}$	Ostro and Rothschild	1989	nationwide	18-64	All	All	O_3	24-hr avg	0.00741	0.00070	Log-linear
School Loss Days, All Cause	PM_{10}	Chen et al.	2000	Washoe Co, NV	6-11	All	All	CO, O_3	24-hr avg	-0.015400	0.004400	Linear
School Loss Days, All Cause	PM_{10}	Gilliland et al.	2001	Southern California	9-10	All	All	None	24-hr avg	0.020539	0.004894	Log-linear
School Loss Days, All Cause	PM_{10}	Ransom and Pope	1992	Provo, UT	6-11	All	All	None	24-hr avg	0.021921	0.00461	Linear
School Loss Days, All Cause	PM_{10}	Ransom and Pope	1992	Orem, UT	6-11	All	All	None	24-hr avg	0.02115	0.00460	Linear
School Loss Days, Illness-Related	PM_{10}	Gilliland et al.	2001	Southern California	9-10	All	All	None	24-hr avg	0.005543	0.009387	Log-linear
School Loss Days, Respiratory-Related	PM_{10}	Gilliland et al.	2001	Southern California	9-10	All	All	None	24-hr avg	-0.004395	0.017569	Log-linear
Work Loss Days	$PM_{2.5}$	Ostro	1987	nationwide	18-64	All	All	None	24-hr avg	0.0046	0.00036	Log-linear

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.

F .6 Acute Effects

F .6.1 Acute Bronchitis (Dockery et al., 1996)

Dockery et al. (1996) examined the relationship between PM and other pollutants on the reported rates of asthma, persistent wheeze, chronic cough, and bronchitis, in a study of 13,369 children ages 8-12 living in 24 communities in U.S. and Canada. Health data were collected in 1988-1991, and single-pollutant models were used in the analysis to test a number of measures of particulate air pollution. Dockery et al. found that annual level of sulfates and particle acidity were significantly related to bronchitis, and $PM_{2.1}$ and PM_{10} were marginally significantly related to bronchitis.[84] They also found nitrates were linked to asthma, and sulfates linked to chronic phlegm. It is important to note that thestudy examined annual pollution exposures, and the authors did not rule out that acute (daily) exposures could be related to asthma attacks and other acute episodes. Earlier work, by Dockery et al. (1989), based on six U.S. cities, found acute bronchitis and chronic cough significantly related to PM_{15}. Because it is based on a larger sample, the Dockery et al. (1996) study is the better study to develop a C-R function linking $PM_{2.5}$ with bronchitis.

Bronchitis was counted in the study only if there were "reports of symptoms in the past 12 months" (Dockery et al., 1996, p. 501). It is unclear, however, if the cases of bronchitis are acute and temporary, or if the bronchitis is a chronic condition. Dockery et al. found no relationship between PM and chronic cough and chronic phlegm, which are important indicators of chronic bronchitis. For this analysis, we assumed that the C-R function based on Dockery et al. is measuring acute bronchitis. The C-R function is based on results of the single pollutant model reported in Table 1.

Single Pollutant Model

The estimated logistic coefficient and standard error are based on the odds ratio (1.50) and 95% confidence interval (0.91-2.47) associated with being in the most polluted city ($PM_{2.1}$ = 20.7 $\mu g/m^3$) versus the least polluted city ($PM_{2.1}$ = 5.8 $\mu g/m^3$) (Dockery et al., 1996, Tables 1 and 4). The original study used $PM_{2.1}$, however, we use the $PM_{2.1}$ coefficient and apply it to $PM_{2.5}$ data.

Functional Form: Logistic
Coefficient: 0.027212
Standard Error: 0.017096
Incidence Rate: annual bronchitis incidence rate per person = 0.043 (American Lung Association, 2002a, Table 11)
Population: population of ages 8-12

F .6.2 Acute Myocardial Infarction (Heart Attacks), Nonfatal (Peters et al., 2001)

Peters et al. (2001) studied the relationship between increased particulate air pollution and onset of heart attacks in the Boston area from 1995 to 1996. The authors used air quality data for PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$,"black carbon", O_3, CO, NO_2, and SO_2 in a case-crossover analysis. For each subject, the case period was matched to three control periods, each 24 hours apart. In univariate analyses, the authors observed a positive association between heart attack occurrence and $PM_{2.5}$ levels hours before and days

[84] The original study measured $PM_{2.1}$, however when using the study's results we use $PM_{2.5}$. This makes only a negligible difference, assuming that the adverse effects of $PM_{2.1}$ and $PM_{2.5}$ are comparable.

before onset. The authors estimated multivariate conditional logistic models including two-hour and twenty-four hour pollutant concentrations for each pollutant. They found significant and independent associations between heart attack occurrence and both two-hour and twenty-four hour $PM_{2.5}$ concentrations before onset. Significant associations were observed for PM_{10} as well. None of the other particle measures or gaseous pollutants were significantly associated with acute myocardial infarction for the two hour or twenty-four hour period before onset.

The patient population for this study was selected from health centers across the United States. The mean age of participants was 62 years old, with 21% of the study population under the age of 50. In order to capture the full magnitude of heart attack occurrence potentially associated with air pollution and because age was not listed as an inclusion criteria for sample selection, we apply an age range of 18 and over in the C-R function. According to the National Hospital Discharge Survey, there were no hospitalizations for heart attacks among children <15 years of age in 1999 and only 5.5% of all hospitalizations occurred in 15-44 year olds (Popovic, 2001, Table 10).

$PM_{2.5}$ Function(s)

Single Pollutant Model

The coefficient and standard error are calculated from an odds ratio of 1.62 (95% CI 1.13-2.34) for a 20 $\mu g/m^3$ increase in twenty-four hour average $PM_{2.5}$ (Peters et al., 2001, Table 4, p. 2813).

Functional Form: Logistic
Coefficient: 0.024121
Standard Error: 0.009285
Incidence Rate: region-specific daily nonfatal heart attack rate per person 18+ = 93% of region-specific daily heart attack hospitalization rate (ICD code 410) [85]
Population: population of ages 18 and older

$PM_{10-2.5}$ Function(s)

Single Pollutant Model

The coefficient and standard error are calculated from an odds ratio of 1.39 (95% CI 0.89-2.15) for a 15 $\mu g/m^3$ increase in twenty-four hour average $PM_{10-2.5}$ (Peters et al., 2001, Table 4, p. 2813).

Functional Form: Logistic
Coefficient: 0.021954
Standard Error: 0.015000
Incidence Rate: region-specific daily nonfatal heart attack rate = 93% of region-specific daily heart attack hospitalization rate (ICD code 410) [86]
Population: population of all ages

[85] This estimate assumes that all heart attacks that are not instantly fatal will result in a hospitalization. In addition, Rosamond et al. (1999) report that approximately six percent of male and eight percent of female hospitalized heart attack patients die within 28 days (either in or outside of the hospital). We applied a factor of 0.93 to the number of hospitalizations to estimate the number of nonfatal heart attacks per year.

[86] This estimate assumes that all heart attacks that are not instantly fatal will result in a hospitalization. In addition, Rosamond et al. (1999) report that approximately six percent of male and eight percent of female hospitalized heart attack patients die within 28 days (either in or outside of the hospital). We applied a factor of 0.93 to the number of hospitalizations to estimate the number of nonfatal heart attacks per year.

PM_{10} Function(s)

Single Pollutant Model

The coefficient and standard error are calculated from an odds ratio of 1.66 (95% CI 1.11-2.49) for a 30 μg/m^3 increase in twenty-four hour average PM_{10} (Peters et al., 2001, Table 4, p. 2813).

Functional Form: Logistic
Coefficient: 0.016894
Standard Error: 0.006870
Incidence Rate: region-specific daily nonfatal heart attack rate = 93% of region-specific daily heart attack hospitalization rate (ICD code 410) [87]
Population: population of all ages

F .6.3 Any of 19 Respiratory Symptoms (Krupnick et al., 1990)

Krupnick et al. (1990) estimated the impact of air pollution on the incidence of any of 19 respiratory symptoms or conditions in 570 adults and 756 children living in three communities in Los Angeles, California from September 1978 to March 1979. Krupnick et al. (1990) listed 13 specific "symptoms or conditions": head cold, chest cold, sinus trouble, croup, cough with phlegm, sore throat, asthma, hay fever, doctor-diagnosed ear infection, flu, pneumonia, bronchitis, and bronchiolitis. The other six symptoms or conditions are not specified.

In their analysis, they included COH, ozone, NO_2, and SO_2, and they used a logistic regression model that takes into account whether a respondent was well or not the previous day. A key difference between this and the usual logistic model, is that the model they used includes a lagged value of the dependent variable. In single-pollutant models, daily O_3, COH, and SO_2 were significantly related to respiratory symptoms in adults. Controlling for other pollutants, they found that ozone was still significant. The results were more variable for COH and SO_2, perhaps due to collinearity. NO_2 had no significant effect. No effect was seen in children for any pollutant. The results from the two-pollutant model with COH and ozone are used to develop a C-R function.

Multipollutant Model (PM_{10} and ozone)

The C-R function used to estimate the change in ARD2 associated with a change in daily average PM_{10} concentration is based on Krupnick et al. (1990, p. 12):[88]

$$\Delta ARD2 \cong \beta^*_{PM_{10}} \cdot \Delta PM_{10} \cdot pop \,,$$

[87] This estimate assumes that all heart attacks that are not instantly fatal will result in a hospitalization. In addition, Rosamond et al. (1999) report that approximately six percent of male and eight percent of female hospitalized heart attack patients die within 28 days (either in or outside of the hospital). We applied a factor of 0.93 to the number of hospitalizations to estimate the number of nonfatal heart attacks per year.

[88] Krupnick and Kopp (1988, p. 2-24) and ESEERCO (1994, p. V-32) used the same C-R functional form as that used here.

Functional Form: Linear
Coefficient: first derivative of the stationary probability = 0.000461
Standard Error: 0.000239
Population: population of ages 18-64 years[89]

The logistic regression model used by Krupnick et al. (1990) takes into account whether a respondent was well or not the previous day. Following Krupnick et al. (p. 12), the probability that one is sick is on a given day is:

$$probability(ARD2) = \frac{p_0}{1 - p_1 + p_0}$$

$$probability(ARD2 \mid sickness\ or\ not_{t-1}) = p_i = \frac{1}{1 - e^{\beta_0 + \beta_1 \cdot ARD2_{t-1} + X \cdot \beta}}, for\ i = 0,1.$$

where:

X	=	the matrix of explanatory variables
p_0	=	the probability of sickness on day t, given wellness on day t-1, and
p_1	=	the probability of sickness on day t, given sickness on day t-1.

In other words, the transition probabilities are estimated using a logistic function; the key difference between this and the usual logistic model, is that the model includes a lagged value of the dependent variable.

To calculate the impact of COH (or other pollutants) on the probability of ARD2, it is possible, in principle, to estimate ARD2 before the change in COH and after the change:

$$\Delta ARD2 = ARD2_{after} - ARD2_{before}.$$

However the full suite of coefficient estimates are not available.[90] Rather than use the full suite of coefficient values, the impact of COH on the probability of probability of ARD2 may be approximated by the derivative of ARD2 with respect to COH:

[89] Krupnick et al. (1990, Table 1) reported the age distribution in their complete data, but they did not report the ages of individuals that were considered "adult." This analysis assumes that individuals 18 and older were considered adult. Only a small percentage (0.6%) of the study population is above the age of 60, so the C-R function was limited to the adult population. up through the age of 65.

[90] The model without NO_2 (Krupnick et al., 1990, Table V equation 3) was used in this analysis, but the full suite of coefficient estimates for this model were not reported. Krupnick et al. (1990, Table IV) reported all of the estimated coefficients for a model of children and for a model of adults when four pollutants were included (ozone, COH, SO_2, and NO_2). However, because of high collinearity between NO_2 and COH, NO_2 was dropped from some of the reported analyses (Krupnick et al., p. 10), and the resulting coefficient estimates changed substantially (see Krupnick et al., 1990, Table IV). Both the ozone and COH coefficients dropped by about a factor of two or more.

$$\frac{\partial probability(ARD2)}{\partial COH} = \frac{p_0 \cdot (1 - p_1) \cdot \beta_{COH} \cdot \left[p_1 + (1 - p_0) \right]}{(1 - p_1 + p_0)^2} = \beta^*_{COH},$$

where β_{COH} is the reported logistic regression coefficient for COH. Since COH data are not available for the benefits analysis, an estimated PM_{10} logistic regression coefficient is used based on the following assumed relationship between PM_{10}, COH, and TSP:

$$COH = 0.116 \cdot TSP$$

$$PM_{10} = 0.55 \cdot TSP$$

$$\Rightarrow COH = 0.2109 \cdot PM_{10}$$

$$\Rightarrow \beta_{PM_{10}} = 0.2109 \cdot \beta_{COH} = 0.2109 \cdot 0.0088 = 0.001856.$$

This analysis uses $\beta_{COH} = 0.0088$ (Krupnick et al., 1990, Table V equation 3). The conversion from COH to TSP is based on study-specific information provided to ESEERCO (1994, p. V-32). The conversion of TSP to PM_{10} is from also from ESEERCO (1994, p. V-5), which cited studies by EPA (1986) and the California Air Resources Board (1982).

The change in the incidence of ARD2 associated with a given change in COH is then estimated by:

$$\frac{\partial ARD2}{\partial PM_{10}} \cong \frac{\Delta ARD2}{\Delta PM_{10}}$$

$$\Rightarrow \frac{\Delta ARD2}{\Delta PM_{10}} \cong \beta^*_{PM_{10}}$$

$$\Rightarrow \Delta ARD2 \cong \beta^*_{PM_{10}} \cdot \Delta PM_{10}.$$

This analysis uses transition probabilities obtained from Krupnick et al. as reported by ESEERCO (1994, p. V-32), for the adult population: $p_1 = 0.7775$ and $p_0 = 0.0468$. This implies:

$$\beta^*_{PM_{10}} = \frac{0.0468 \cdot (1 - 0.7775) \cdot 0.001856 \cdot \left[0.7775_1 + (1 - 0.0468) \right]}{(1 - 0.7775 + 0.0468)^2} = 0.000461.$$

The *standard error* for the coefficient is derived using the reported standard error of the logistic regression coefficient in Krupnick et al. (1990, Table V):

$$\Rightarrow \beta_{PM_{10},high} = 0.2109 \cdot \beta_{COH,high} = 0.2109 \cdot (0.0088 + (1.96 \cdot 0.0046)) = 0.003757$$

$$\Rightarrow \beta^{*}_{PM_{10},high} = \frac{0.0468 \cdot (1 - 0.7775) \cdot 0.003757 \cdot [0.7775 + (1 - 0.0468)]}{(1 - 0.7775 + 0.0468)^2} = 0.000934$$

$$\sigma_{\beta,high} = \frac{\beta_{PM_{10},high} - \beta_{PM_{10}}}{1.96} = \frac{(0.000934 - 0.000461)}{1.96} = 0.000236$$

$$\beta_{PM_{10},low} = 0.2109 \cdot \beta_{COH,low} = 0.2109 \cdot (0.0088 - (1.96 \cdot 0.0046)) = -4.555 \cdot 10^{-5}$$

$$\Rightarrow \beta^{*}_{PM_{10},low} = \frac{0.0468 \cdot (1 - 0.7775) \cdot (-4.555 \cdot 10^{-5}) \cdot [0.7775 + (1 - 0.0468)]}{(1 - 0.7775 + 0.0468)^2} = -1.132 \cdot 10^{-5}$$

$$\Rightarrow \sigma_{\beta,low} = \frac{\beta - \beta_{low}}{1.96} = \frac{(0.000461 + 1.132 \cdot 10^{-5})}{1.96} = 0.000241$$

$$\sigma_{\beta} = \frac{\sigma_{\beta,high} + \sigma_{\beta,low}}{2} = 0.000239 .$$

F .6.4 Lower Respiratory Symptoms (Schwartz and Neas, 2000)

Schwartz et al. (2000) replicated a previous analysis (Schwartz et al., 1994) linking PM levels to lower respiratory symptoms in children in six cities in the U.S. The original study enrolled 1,844 children into a year-long study that was conducted in different years (1984 to 1988) in six cities. The students were in grades two through five at the time of enrollment in 1984. By the completion of the final study, the cohort would then be in the eighth grade (ages 13-14); this suggests an age range of 7 to 14. The previous study focused on PM_{10}, acid aerosols, and gaseous pollutants, although single-pollutant $PM_{2.5}$ results were reported. Schwartz et al. (2000) focused more on the associations between $PM_{2.5}$ and $PM_{10-2.5}$ and lower respiratory symptoms. In single and co-pollutant models, $PM_{2.5}$ was significantly associated with lower respiratory symptoms, while $PM_{10-2.5}$ was not. $PM_{10-2.5}$ exhibited a stronger association with cough than did $PM_{2.5}$. The $PM_{2.5}$ C-R functions for lower respiratory symptoms are based on the results of the reported single pollutant and co-pollutant model ($PM_{2.5}$ and $PM_{10-2.5}$).

Single Pollutant Model

The coefficient and standard error are calculated from the reported odds ratio (1.33) and 95% confidence interval (1.11-1.58) associated with a 15 $\mu g/m^3$ change in $PM_{2.5}$ (Schwartz and Neas, 2000, Table 2).

Functional Form: Logistic
Coefficient: 0.019012
Standard Error: 0.006005
Incidence Rate: daily lower respiratory symptom incidence rate per person = 0.0012 (Schwartz et al., 1994, Table 2)
Population: population of ages 7 to 14

Multipollutant Model ($PM_{2.5}$ and $PM_{10-2.5}$)

In a model with $PM_{10-2.5}$, the $PM_{2.5}$ coefficient and standard error are calculated from the reported odds ratio (1.29) and 95% confidence interval (1.06-1.57) associated with a 15 $\mu g/m^3$ change in $PM_{2.5}$ (Schwartz and Neas, 2000, Table 2).

Functional Form: Logistic
Coefficient: 0.016976
Standard Error: 0.006680
Incidence Rate: daily lower respiratory symptom incidence rate per person = 0.0012 (Schwartz et al., 1994, Table 2)
Population: population of ages 7 to 14

F .6.5 Lower Respiratory Symptoms (Schwartz et al., 1994)

Schwartz et al. (1994) used logistic regression to link lower respiratory symptoms in children with SO_2, NO_2, ozone, PM_{10}, $PM_{2.5}$, sulfate and H^+ (hydrogen ion). Children were selected for the study if they were exposed to indoor sources of air pollution: gas stoves and parental smoking. The study enrolled 1,844 children into a year-long study that was conducted in different years (1984 to 1988) in six cities. The students were in grades two through five at the time of enrollment in 1984. By the completion of the final study, the cohort would then be in the eighth grade (ages 13-14); this suggests an age range of 7 to 14.

In single pollutant models SO_2, NO_2, $PM_{2.5}$, and PM_{10} were significantly linked to cough. In two-pollutant models, PM_{10} had the most consistent relationship with cough; ozone was marginally significant, controlling for PM_{10}. In models for upper respiratory symptoms, they reported a marginally significant association for PM_{10}. In models for lower respiratory symptoms, they reported significant single-pollutant models, using SO_2, O_3, $PM_{2.5}$, PM_{10}, SO_4, and H^+. The $PM_{2.5}$ C-R function is based on the single pollutant model reported in Table 5.

Single Pollutant Model

The coefficient and standard error are calculated from the reported odds ratio (1.44) and 95% confidence interval (1.15-1.82) associated with a 20 $\mu g/m^3$ change in $PM_{2.5}$ (Schwartz et al., 1994, Table 5).

Functional Form: Logistic
Coefficient: 0.018232
Standard Error: 0.005856
Incidence Rate: daily lower respiratory symptom incidence rate per person = 0.0012 (Schwartz et al., 1994, Table 2)
Population: population of ages 7 to 14

F .6.6 Minor Restricted Activity Days: Ostro and Rothschild (1989)

Ostro and Rothschild (1989) estimated the impact of $PM_{2.5}$ and ozone on the incidence of minor restricted activity days (MRADs) and respiratory-related restricted activity days (RRADs) in a national sample of the adult working population, ages 18 to 65, living in metropolitan areas.[91] The annual national survey results used in this analysis were conducted in 1976-1981. Controlling for $PM_{2.5}$, two-week average ozone has highly variable association with RRADs and MRADs. Controlling for ozone, two-week average $PM_{2.5}$ was significantly linked to both health endpoints in most years. The C-R function for PM is based on this co-pollutant model.

The study is based on a "convenience" sample of non-elderly individuals. Applying the C-R function to this age group is likely a slight underestimate, as it seems likely that elderly are at least as susceptible to PM as individuals under 65. The elderly appear more likely to die due to PM exposure than other age groups (e.g., Schwartz, 1994d, p. 30) and a number of studies have found that hospital admissions for the elderly are related to PM exposures (e.g., Schwartz, 1994b; Schwartz, 1994c).

Multipollutant Model ($PM_{2.5}$ and ozone)

Using the results of the two-pollutant model, we developed separate coefficients for each year in the analysis, which were then combined for use in this analysis. The coefficient is a weighted average of the coefficients in Ostro and Rothschild (1989, Table 4) using the inverse of the variance as the weight:

$$\beta = \left(\frac{\sum_{i=1976}^{1981} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\sum_{i=1976}^{1981} \frac{1}{\sigma_{\beta_i}^2}} \right) = 0.00741 .$$

The standard error of the coefficient is calculated as follows, assuming that the estimated year-specific coefficients are independent:

$$\sigma_\beta^2 = \text{var}\left(\frac{\sum_{i=1976}^{1981} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\sum_{i=1976}^{1981} \frac{1}{\sigma_{\beta_i}^2}} \right) = \left(\frac{\sum_{i=1976}^{1981} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\gamma} \right) = \sum_{i=1976}^{1981} \text{var}\left(\frac{\beta_i}{\sigma_{\beta_i}^2 \cdot \gamma} \right) .$$

This reduces down to:

$$\sigma_\beta^2 = \frac{1}{\gamma} \Rightarrow \sigma_\beta = \sqrt{\frac{1}{\gamma}} = 0.00070 .$$

[91] The study population is based on the Health Interview Survey (HIS), conducted by the National Center for Health Statistics. In publications from this ongoing survey, non-elderly adult populations are generally reported as ages 18-64. From the study, it is not clear if the age range stops at 65 or includes 65 year olds. We apply the C-R function to individuals ages 18-64 for consistency with other studies estimating impacts to non-elderly adult populations.

Functional Form: Log-linear
Coefficient: 0.00741
Standard Error: 0.00070
Incidence Rate: daily incidence rate for minor restricted activity days (MRAD) = 0.02137 (Ostro and Rothschild, 1989, p. 243)
Population: adult population ages 18 to 64

F .6.7 School Loss Days, All Cause (Chen et al., 2000)

Chen et al. (2000) studied the association between air pollution and elementary school absenteeism (grades 1-6)[92] in Washoe County, Nevada. Daily absence data were available for all elementary schools in the Washoe Country School District. The authors regressed daily total absence rate on the three air pollutants, meteorological variables, and indicators for day of the week, month, and holidays. They reported statistically significant associations between both ozone and CO and daily total absence rate for grades one through six. PM_{10} was negatively associated with absence rate, after adjustment for ozone, CO, and meteorological and temporal variables. The C-R function for PM is based on the results from a multiple linear regression model with CO, ozone, and PM_{10}.

Multipollutant Model (PM_{10}, CO, and ozone)

The coefficient and standard error are presented in Table 3 (Chen et al., 2000, p. 1008) for a unit $\mu g/m^3$ increase in daily PM_{10} concentration.

The reported coefficient represents an *absolute* increase in absenteeism rate for a unit increase in PM_{10}. If we apply this study to other locations, we assume that the same absolute increase will occur for a unit increase in PM_{10}, regardless of the baseline rate. If the study location has a particularly high baseline rate, we may be overestimating decreases in absenteeism nationally, and vice-versa. As an example, consider if the baseline absenteeism rate were 10% in the study and 5% nationally. An absolute increase in absence rate of 2% associated with a given increase in PM_{10} reflects a relative increase in absence rate of 20% for the study population. However, in the national estimate, we would assume the same absolute increase of 2%, but this would reflect a relative increase in the absenteeism rate of 40%.

An alternative approach is to estimate apply the *relative* increase in absenteeism rate in the C-R function by adjusting the results by the ratio of the national absenteeism rate to the study-specific rate. As a result, the percent increase in absenteeism rate associated with an increase in PM_{10} is extrapolated nationally rather than the absolute increase in absenteeism rate. The incidence derivation section above describes the data used to estimate national and study-specific absence rates.

In addition to this scaling factor, there are two other scaling factors which are applied to the function. A scaling factor of 0.01 is used to convert the beta from a percentage (x 100) per unit increase of PM_{10} to a proportion per unit increase of PM_{10}. As a result it can be applied directly to the national population of school children ages 6 through 11 to estimate the number of absences avoided.

The final scaling factor is used to adjust for the proportion of school days in the full year. In the modeling program, the function is applied to every day in the year, however, in reality, school absences will be avoided only on school days. Using an estimate of 180 school days per year, we estimate that

[92] Assuming that most children start kindergarten at age 5, the corresponding ages for grades 1 through 6 would be 6 through 11.

49.3% of the days in a given year are school days (180/365). The C-R function parameters are shown below.

Functional Form: Linear
Coefficient: -0.015400
Standard Error: 0.004400
Population: population of children ages 6-11
Scaling Factor 1: Ratio of national school absence rate to study-specific school absence rate[93] = 1.081
Scaling Factor 2: Convert beta in percentage terms to a proportion = 0.01
Scaling Factor 3: Proportion of days in the year that are school days[94] = 0.493

F .6.8 School Loss Days, All Cause (Gilliland et al., 2001)

Gilliland et al. (2001) examined the association between air pollution and school absenteeism among 4th grade school children (ages 9-10) in 12 southern Californian communities. The study was conducted from January through June 1996. The authors used school records to collect daily absence data and parental telephone interviews to identify causes. They defined illness-related absences as respiratory or non-respiratory. A respiratory illness was defined as an illness that included at least one of the following: runny nose/sneezing, sore throat, cough, earache, wheezing, or asthma attack. The authors used 15 and 30 day distributed lag models to quantify the association between ozone, PM_{10}, and NO_2 and incident school absences. Ozone levels were positively associated with all school absence measures and significantly associated with all illness-related school absences (non-respiratory illness, respiratory illness, URI and LRI). Neither PM_{10} nor NO_2 was significantly associated with illness-related school absences, but PM_{10} was associated with non-illness related absences. The C-R function for PM_{10} is based on the results of the single pollutant model.

Gilliland et al. (2001) defines an incident absence as an absence that followed attendance on the previous day and the incidence rate as the number of incident absences on a given day over the population at risk for an absence on a given day (i.e. those children who were not absent on the previous day). Since school absences due to air pollution may last longer than one day, an estimate of the average duration of school absences could be used to calculated the total avoided school loss days from an estimate of avoided new absences. A simple ratio of the total absence rate divided by the new absence rate would provide an estimate of the average duration of school absences, which could be applied to the estimate of avoided new absences as follows:

$$Duration = \frac{totalAbsences}{newAbsences}$$

$$\Delta\, TotalAbsences = -\left[incidence \cdot \left(e^{-\beta \cdot \Delta PM_{10}} - 1\right)\right] \cdot duration \cdot pop$$

[93] National school absence rate of 5.50% obtained from the U.S. Department of Education (1996, Table 42-1). Study-specific school absence rate of 5.09% obtained from Chen et al. (2000, Table 1).

[94] Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365).

Since the function is log-linear, the baseline incidence rate (in this case, the rate of new absences) is multiplied by duration, which reduces to the total school absence rate. Therefore, the same result would be obtained by using a single estimate of the total school absence rate in the C-R function. Using this approach, we assume that the same relationship observed between pollutant and new school absences in the study would be observed for total absences on a given day. As a result, the total school absence rate is used in the function below. The derivation of this rate is described in the section on baseline incidence rate estimation.

Single Pollutant Model

For all absences, the coefficient and standard error are based on a percent increase of 22.8 percent (95% CI 11.6 percent, 35.2 percent) associated with a 10 $\mu g/m^3$ increase in daily average PM_{10} concentration (2001, Table 6, p. 52).

A scaling factor is used to adjust for the number of school days in the ozone season. In the modeling program, the function is applied to every day in the year, however, in reality, school absences will be avoided only on school days. Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365).

In addition, not all children are at-risk for a new school absence, as defined by the study. On average, 5.5% of school children are absent from school on a given day (U.S. Department of Education, 1996, Table 42-1). Only those who are in school on the previous day are at risk for a new absence (1-0.055 = 94.5%). As a result, a factor of 94.5% is used in the function to estimate the population of school children at-risk for a new absence.

Functional Form: Log-linear
Coefficient: 0.020539
Standard Error: 0.004894
Incidence Rate: daily school absence rate = 0.055 (U.S. Department of Education, 1996, Table 42-1)
Population: population of children ages 9-10 not absent from school on a given day[95] = 94.5% of children ages 9-10
Scaling Factor: Proportion of school days in a year[96] = 0.493

F .6.9 School Loss Days, All Cause (Ransom and Pope, 1992, Provo)

Ransom and Pope (1992) studied the relationship between particulate air pollution and elementary school absenteeism (grades 1-6)[97] in Utah Valley from 1985 to 1990. The authors identified school absences using weekly attendance data from the Provo School District and daily attendance data from an elementary school in Orem, Utah. The authors regressed school absence rates on PM_{10},weather variables, day of the week, month of the year, and indicators for holidays or extended weekends. The authors report that a four week moving average of PM_{10} provided the best model fit. They found a statistically significant

[95] The proportion of children not absent from school on a given day (5.5%) is based on 1996 data from the U.S. Department of Education (1996, Table 42-1).

[96] Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365).

[97] Assuming that most children start kindergarten at age 5, the corresponding ages for grades 1 through 6 would be 6 through 11.

association between increases in PM_{10} and absence rates in Provo and Orem, after adjustment for weather variables and temporal trends. The C-R function for PM_{10} is based on results of the linear regression model in the Provo School District for grades 1-6 (Ransom and Pope, 1992, Table 3, p. 211).

Single Pollutant Model

For Provo, the coefficient and standard error for a 100 $\mu g/m^3$ increase in four-week average PM_{10} concentration are reported as 2.1921 and 0.4610, respectively (Ransom and Pope, 1992, Table 3, p. 211).

The reported coefficient represents an *absolute* increase in absenteeism rate for a unit increase in PM_{10}. If we apply this study to other locations, we assume that the same absolute increase will occur for a unit increase in PM_{10}, regardless of the baseline rate. If the study location has a particularly high baseline rate, we may be overestimating decreases in absenteeism nationally, and vice-versa. As an example, consider if the baseline absenteeism rate were 10% in the study and 5% nationally. An absolute increase in absence rate of 2% associated with a given increase in PM_{10} reflects a relative increase in absence rate of 20% for the study population. However, in the national estimate, we would assume the same absolute increase of 2%, but this would reflect a relative increase in the absenteeism rate of 40%.

An alternative approach is to estimate apply the *relative* increase in absenteeism rate in the C-R function by adjusting the results by the ratio of the national absenteeism rate to the study-specific rate. As a result, the percent increase in absenteeism rate associated with an increase in PM_{10} is extrapolated nationally rather than the absolute increase in absenteeism rate. The incidence derivation section above describes the data used to estimate national and study-specific absence rates.

An additional scaling factor is used to adjust for the proportion of school days in the full year. In the modeling program, the function is applied to every day in the year, however, in reality, school absences will be avoided only on school days. Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365). The C-R function parameters are shown below.

Functional Form: Linear
Coefficient: 0.021921
Standard Error: 0.00461
Population: population of children ages 6-11
Scaling Factor 1: Ratio of national school absence rate to study-specific school absence rate[98] = 1.211
Scaling Factor 2: Proportion of school days in a year[99] = 0.493

F .6.10 School Loss Days, All Cause (Ransom and Pope, 1992, Orem)

Ransom and Pope (1992) studied the relationship between particulate air pollution and elementary school absenteeism (grades 1-6)[100] in Utah Valley from 1985 to 1990. The authors identified school

[98] National school absence rate of 5.5% obtained from the U.S. Department of Education (1996, Table 42-1). Study-specific school absence rate of 4.54% obtained from Ransom and Pope (1992, Table 2).

[99] Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365).

[100] Assuming that most children start kindergarten at age 5, the corresponding ages for grades 1 through 6 would be 6 through 11.

absences using weekly attendance data from the Provo School District and daily attendance data from an elementary school in Orem, Utah. The authors regressed school absence rates on PM_{10},weather variables, day of the week, month of the year, and indicators for holidays or extended weekends. The authors report that a four week moving average of PM_{10} provided the best model fit. They found a statistically significant association between increases in PM_{10} and absence rates in Provo and Orem, after adjustment for weather variables and temporal trends. The C-R function for PM_{10} is based on results of the linear regression model for grades 1-6 in Orem, Utah (Ransom and Pope, 1992, Table 4, p. 212).

Single Pollutant Model

For Orem, the coefficient and standard error for a 100 $\mu g/m^3$ increase in four-week average PM_{10} concentration are reported as 2.115 and 0.4600, respectively (Ransom and Pope, 1992, Table 4, p. 212).

The reported coefficient represents an *absolute* increase in absenteeism rate for a unit increase in PM_{10}. If we apply this study to other locations, we assume that the same absolute increase will occur for a unit increase in PM_{10}, regardless of the baseline rate. If the study location has a particularly high baseline rate, we may be overestimating decreases in absenteeism nationally, and vice-versa. As an example, consider if the baseline absenteeism rate were 10% in the study and 5% nationally. An absolute increase in absence rate of 2% associated with a given increase in PM_{10} reflects a relative increase in absence rate of 20% for the study population. However, in the national estimate, we would assume the same absolute increase of 2%, but this would reflect a relative increase in the absenteeism rate of 40%.

An alternative approach is to estimate apply the *relative* increase in absenteeism rate in the C-R function by adjusting the results by the ratio of the national absenteeism rate to the study-specific rate. As a result, the percent increase in absenteeism rate associated with an increase in PM_{10} is extrapolated nationally rather than the absolute increase in absenteeism rate. The incidence derivation section above describes the data used to estimate national and study-specific absence rates.

An additional scaling factor is used to adjust for the proportion of school days in the full year. In the modeling program, the function is applied to every day in the year, however, in reality, school absences will be avoided only on school days. Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365). The C-R function parameters are shown below.

Functional Form: Linear
Coefficient: 0.02115
Standard Error: 0.00460
Population: population of children ages 6-11
Scaling Factor 1: Ratio of national school absence rate to study-specific school absence rate[101] = 1.076
Scaling Factor 2: Proportion of school days in a year[102] = 0.493

[101] National school absence rate of 5.50% obtained from the U.S. Department of Education (1996, Table 42-1). Study-specific school absence rate of 5.11% obtained from Ransom and Pope (1992, Table 1).

[102] Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365).

F .6.11 School Loss Days, Illness-Related (Gilliland et al., 2001)

Gilliland et al. (2001) examined the association between air pollution and school absenteeism among 4th grade school children (ages 9-10) in 12 southern Californian communities. The study was conducted from January through June 1996. The authors used school records to collect daily absence data and parental telephone interviews to identify causes. They defined illness-related absences as respiratory or non-respiratory. A respiratory illness was defined as an illness that included at least one of the following: runny nose/sneezing, sore throat, cough, earache, wheezing, or asthma attack. The authors used 15 and 30 day distributed lag models to quantify the association between ozone, PM_{10}, and NO_2 and incident school absences. Ozone levels were positively associated with all school absence measures and significantly associated with all illness-related school absences (non-respiratory illness, respiratory illness, URI and LRI). Neither PM_{10} nor NO_2 was significantly associated with illness-related school absences, but PM_{10} was associated with non-illness related absences. The C-R function for PM_{10} is based on the results of the single pollutant model.

Gilliland et al. (2001) defines an incident absence as an absence that followed attendance on the previous day and the incidence rate as the number of incident absences on a given day over the population at risk for an absence on a given day (i.e. those children who were not absent on the previous day). Since school absences due to air pollution may last longer than one day, an estimate of the average duration of school absences could be used to calculated the total avoided school loss days from an estimate of avoided new absences. A simple ratio of the total absence rate divided by the new absence rate would provide an estimate of the average duration of school absences, which could be applied to the estimate of avoided new absences as follows:

$$Duration = \frac{totalAbsences}{newAbsences}$$

$$\Delta\, TotalAbsences = -\left[incidence \cdot (e^{-\beta \cdot \Delta PM_{10}} - 1)\right] \cdot duration \cdot pop$$

Since the function is log-linear, the baseline incidence rate (in this case, the rate of new absences) is multiplied by duration, which reduces to the total school absence rate. Therefore, the same result would be obtained by using a single estimate of the total school absence rate in the C-R function. Using this approach, we assume that the same relationship observed between pollutant and new school absences in the study would be observed for total absences on a given day. As a result, the total school absence rate is used in the function below. The derivation of this rate is described in the section on baseline incidence rate estimation.

Single Pollutant Model

For total illness-related absences, the coefficient and standard error are based on a percent increase of 5.7 percent (95% CI -12.1 percent, 27.0 percent) associated with a 10 $\mu g/m^3$ increase in daily average PM_{10} concentration (2001, Table 6, p. 52).

A scaling factor is used to adjust for the number of school days in the ozone season. In the modeling program, the function is applied to every day in the year, however, in reality, school absences will be avoided only on school days. Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365).

In addition, not all children are at-risk for a new school absence, as defined by the study. On average, 5.5% of school children are absent from school on a given day (U.S. Department of Education, 1996, Table 42-1). Only those who are in school on the previous day are at risk for a new absence (1-0.055 = 94.5%). As a result, a factor of 94.5% is used in the function to estimate the population of school children at-risk for a new absence.

Functional Form: Log-linear
Coefficient: 0.005543
Standard Error: 0.009387
Incidence Rate: region-specific daily illness-related school absence rate (Adams et al., 1999, Table 47), assuming 180 school days per year.
Population: population of children ages 9-10 not absent from school on a given day[103] = 94.5% of children ages 9-10
Scaling Factor: Proportion of school days in a year[104] = 0.493

F .6.12 School Loss Days, Respiratory Illness-Related (Gilliland et al., 2001)

Gilliland et al. (2001) examined the association between air pollution and school absenteeism among 4th grade school children (ages 9-10) in 12 southern Californian communities. The study was conducted from January through June 1996. The authors used school records to collect daily absence data and parental telephone interviews to identify causes. They defined illness-related absences as respiratory or non-respiratory. A respiratory illness was defined as an illness that included at least one of the following: runny nose/sneezing, sore throat, cough, earache, wheezing, or asthma attack. The authors used 15 and 30 day distributed lag models to quantify the association between ozone, PM_{10}, and NO_2 and incident school absences. Ozone levels were positively associated with all school absence measures and significantly associated with all illness-related school absences (non-respiratory illness, respiratory illness, URI and LRI). Neither PM_{10} nor NO_2 was significantly associated with illness-related school absences, but PM_{10} was associated with non-illness related absences.

Gilliland et al. (2001) defines an incident absence as an absence that followed attendance on the previous day and the incidence rate as the number of incident absences on a given day over the population at risk for an absence on a given day (i.e. those children who were not absent on the previous day). Since school absences due to air pollution may last longer than one day, an estimate of the average duration of school absences could be used to calculated the total avoided school loss days from an estimate of avoided new absences. A simple ratio of the total absence rate divided by the "incident" rate would provide an estimate of the average duration of school absences, which could be applied to the estimate of avoided new absences as follows:

$$Duration = \frac{totalAbsences}{newAbsences}$$

$$\Delta\, TotalAbsences = -\left[incidence \cdot \left(e^{-\beta \cdot \Delta PM_{10}} - 1\right)\right] \cdot duration \cdot pop$$

[103] The proportion of children not absent from school on a given day (5.5%) is based on 1996 data from the U.S. Department of Education (1996, Table 42-1).

[104] Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365).

Since the function is log-linear, the baseline incidence rate (in this case, the rate of new absences) is multiplied by duration, which reduces to the total school absence rate. Therefore, the same result would be obtained by using a single estimate of the total school absence rate in the C-R function. Using this approach, we assume that the same relationship observed between pollutant and new school absences in the study would be observed for total absences on a given day. As a result, the total school absence rate is used in the function below. The derivation of this rate is described in the section on baseline incidence rate estimation.

Single Pollutant Model

For respiratory illness-related absences, the coefficient and standard error are based on a percent increase of -4.3 percent (95% CI -32.2 percent, 35.0 percent) associated with a 10 $\mu g/m^3$ increase in daily average PM_{10} concentration (2001, Table 6, p. 52).

A scaling factor is used to adjust for the number of school days in the ozone season. In the modeling program, the function is applied to every day in the year, however, in reality, school absences will be avoided only on school days. Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365).

In addition, not all children are at-risk for a new school absence, as defined by the study. On average, 5.5% of school children are absent from school on a given day (U.S. Department of Education, 1996, Table 42-1). Only those who are in school on the previous day are at risk for a new absence (1-0.055 = 94.5%). As a result, a factor of 94.5% is used in the function to estimate the population of school children at-risk for a new absence.

Functional Form: Log-linear
Coefficient: -0.004395
Standard Error: 0.017569
Incidence Rate: region-specific daily respiratory illness-related school absence rate (Adams et al., 1999, Table 47), assuming 180 school days per year.
Population: population of children ages 9-10 not absent from school on a given day[105] = 94.5% of children ages 9-10
Scaling Factor: Proportion of school days in a year[106] = 0.493

F .6.13 Work Loss Days (Ostro, 1987)

Ostro (1987) estimated the impact of $PM_{2.5}$ on the incidence of work-loss days (WLDs), restricted activity days (RADs), and respiratory-related RADs (RRADs) in a national sample of the adult working population, ages 18 to 65, living in metropolitan areas.[107] The annual national survey results used in this

[105] The proportion of children not absent from school on a given day (5.5%) is based on 1996 data from the U.S. Department of Education (1996, Table 42-1).

[106] Using an estimate of 180 school days per year, we estimate that 49.3% of the days in a given year are school days (180/365).

[107] The study population is based on the Health Interview Survey (HIS), conducted by the National Center for Health Statistics. In publications from this ongoing survey, non-elderly adult populations are generally reported as ages 18-64. From the study, it is not clear if the age range stops at 65 or includes 65 year olds. We apply the C-R function to individuals ages 18-64 for consistency with other studies estimating impacts to non-elderly adult populations.

analysis were conducted in 1976-1981. Ostro reported that two-week average $PM_{2.5}$ levels[108] were significantly linked to work-loss days, RADs, and RRADs, however there was some year-to-year variability in the results. Separate coefficients were developed for each year in the analysis (1976-1981); these coefficients were pooled. The coefficient used in the concentration-response function presented here is a weighted average of the coefficients in Ostro (1987, Table III) using the inverse of the variance as the weight.

The study is based on a "convenience" sample of non-elderly individuals. Applying the C-R function to this age group is likely a slight underestimate, as it seems likely that elderly are at least as susceptible to PM as individuals under 65. The elderly appear more likely to die due to PM exposure than other age groups (e.g., Schwartz, 1994d, p. 30) and a number of studies have found that hospital admissions for the elderly are related to PM exposures (e.g., Schwartz, 1994b; Schwartz, 1994c). On the other hand, the number of workers over the age of 65 is relatively small; it was approximately 3% of the total workforce in 2001(U.S. Bureau of the Census, 2002, Table 561).

Single Pollutant Model

The coefficient used in the C-R function is a weighted average of the coefficients in Ostro (1987, Table III) using the inverse of the variance as the weight:

$$\beta = \left(\frac{\sum_{i=1976}^{1981} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\sum_{i=1976}^{1981} \frac{1}{\sigma_{\beta_i}^2}} \right) = 0.0046 .$$

The standard error of the coefficient is calculated as follows, assuming that the estimated year-specific coefficients are independent:

$$\sigma_\beta^2 = \text{var}\left(\frac{\sum_{i=1976}^{1981} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\sum_{i=1976}^{1981} \frac{1}{\sigma_{\beta_i}^2}} \right) = \left(\frac{\sum_{i=1976}^{1981} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\gamma} \right) = \sum_{i=1976}^{1981} \text{var}\left(\frac{\beta_i}{\sigma_{\beta_i}^2 \cdot \gamma} \right) .$$

This eventually reduces down to:

$$\sigma_\beta^2 = \frac{1}{\gamma} \Rightarrow \sigma_\beta = \sqrt{\frac{1}{\gamma}} = 0.00036 .$$

[108] The study used a two-week average pollution concentration; the C-R function uses a daily average, which is assumed to be a reasonable approximation.

Functional Form: Log-linear
Coefficient: 0.0046
Standard Error: 0.00036
Incidence Rate: daily work-loss-day incidence rate per person ages 18 to 64 = 0.00595 (U.S. Bureau of the Census, 1997, No. 22; Adams et al., 1999, Table 41)
Population: adult population ages 18 to 64

Appendix F. Particulate Matter C-R Functions

Exhibit F-7. Concentration-Response (C-R) Functions for Particulate Matter and Asthma-Related Effects

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Beta	Std Error	Functional Form	Notes
Acute Bronchitis	$PM_{2.5}$	McConnell et al.	1999	Southern California	9-15	All	All	None	Annual Avg	0.022431	0.015957	Logistic	
Acute Bronchitis	PM_{10}	McConnell et al.	1999	Southern California	9-15	All	All	None	Annual Avg	0.017709	0.006612	Logistic	
Asthma Exacerbation, Asthma Attacks	PM_{10}	Whittemore and Korn	1980	Los Angeles, CA	All	All	All	O_3	24-hr avg	0.001436	0.000558	Logistic	
Asthma Exacerbation, Cough	$PM_{2.5}$	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.000985	0.000747	Logistic	Day with symptoms
Asthma Exacerbation, Cough	$PM_{2.5}$	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.003177	0.001156	Logistic	New onset of symptoms
Asthma Exacerbation, Cough	PM_{10}	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.005606	0.001639	Logistic	Day with symptoms
Asthma Exacerbation, Cough	PM_{10}	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.013126	0.003241	Logistic	New onset of symptoms
Asthma Exacerbation, Cough	PM_{10}	Vedal et al.	1998	Vancouver, CAN	6-13	All	All	None	24-hr avg	0.007696	0.003786	Logistic	
Asthma Exacerbation, Moderate or Worse	$PM_{2.5}$	Ostro et al.	1991	Denver, CO	All	All	All	None	24-hr avg	0.0006	0.0003	Linear (log of pollutant)	
Asthma Exacerbation, One or More Symptoms	PM_{10}	Yu et al.	2000	Seattle, WA	5-13	All	All	None	24-hr avg	0.009531	0.003032	Logistic	
Asthma Exacerbation, One or More Symptoms	PM_{10}	Yu et al.	2000	Seattle, WA	5-13	All	All	CO, SO_2	24-hr avg	0.004879	0.005095	Logistic	
Asthma Exacerbation, Shortness of Breath	PM_{10}	Ostro et al.	1995	Los Angeles, CA	7-12	Black	All	None	24-hr avg	0.008412	0.003631	Logistic	
Asthma Exacerbation, Shortness of Breath	$PM_{2.5}$	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.002565	0.001335	Logistic	Day with symptoms
Asthma Exacerbation, Shortness of Breath	$PM_{2.5}$	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.003177	0.001550	Logistic	New onset of symptoms
Asthma Exacerbation, Shortness of Breath	PM_{10}	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.007708	0.002639	Logistic	Day with symptoms

Appendix F. Particulate Matter C-R Functions

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Beta	Std Error	Functional Form	Notes
Asthma Exacerbation, Shortness of Breath	PM_{10}	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.010725	0.003850	Logistic	New onset of symptoms
Asthma Exacerbation, Wheeze	$PM_{2.5}$	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.001942	0.000803	Logistic	Day with symptoms
Asthma Exacerbation, Wheeze	$PM_{2.5}$	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.002565	0.001030	Logistic	New onset of symptoms
Asthma Exacerbation, Wheeze	PM_{10}	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.002307	0.001733	Logistic	Day with symptoms
Asthma Exacerbation, Wheeze	PM_{10}	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	24-hr avg	0.006666	0.002957	Logistic	New onset of symptoms
Chronic Phlegm	$PM_{2.5}$	McConnell et al.	1999	Southern California	9-15	All	All	None	Annual Avg	0.063701	0.025580	Logistic	
Chronic Phlegm	PM_{10}	McConnell et al.	1999	Southern California	9-15	All	All	None	Annual Avg	0.039049	0.011512	Logistic	
Upper Respiratory Symptoms	PM_{10}	Pope et al.	1991	Utah Valley	9-11	All	All	None	24-hr avg	0.0036	0.0015	Logistic	

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.

F .7 Asthma-Related Effects

F .7.1 Acute Bronchitis (McConnell et al., 1999)

McConnell et al. (1999) examined the relationship between air pollution and bronchitic symptoms among asthmatic 4th, 7th, and 10th grade children in southern California.[109] The authors collected information on the prevalence of bronchitis, chronic cough, and chronic phlegm among children with and without a history of asthma and/or wheeze. They used annual measurements of ozone, PM_{10}, $PM_{2.5}$, NO_2, and acids in a logistic regression model with adjustments for personal covariates. Neither bronchitis, cough, or phlegm were associated with any of the pollutants among children with no history of wheeze or asthma or a history of wheeze without diagnosed asthma. Among asthmatics, PM_{10} was significantly associated with bronchitis and phlegm; $PM_{2.5}$ was significantly associated with phlegm and marginally associated with bronchitis; NO_2 and acids were both significantly associated with phlegm; and ozone was not significantly associated with any of the endpoints.

Bronchitis was defined in the study by the question: "How many times in the past 12 months did your child have bronchitis?" (McConnell et al., 1999, p. 757). It is unclear, however, if the cases of bronchitis are acute and temporary, or if the bronchitis is a chronic condition. McConnell et al. found a relationship between PM and chronic phlegm but none with chronic cough, each of which may be indicators of chronic bronchitis. For this analysis, we assumed that the C-R function based on McConnell et al. is measuring acute bronchitis. The PM C-R functions for bronchitis among asthmatics are based on the results of the single pollutant model reported in Table 3.

$PM_{2.5}$

Single Pollutant Model

The estimated logistic coefficient and standard error are based on the odds ratio (1.4) and 95% confidence interval (0.9-2.3) associated with an increase in yearly mean 2-week average $PM_{2.5}$ of 15 $\mu g/m^3$. (McConnell et al., 1999, Table 3)

Functional Form: Logistic
Coefficient: 0.022431
Standard Error: 0.015957
Incidence Rate: annual incidence rate of one or more episodes of bronchitis per asthmatic = 0.326 (McConnell et al., 1999, Table 2)
Population: population of asthmatics ages 9 to 15 = 5.67%[110] of population ages 9 to 15

[109] Assuming that a child enters kindergarten at age 5, 4th grade corresponds to age 9 and 10th grade corresponds to age 15. We therefore applied the results of this study to children ages 9 to 15.

[110] The American Lung Association (2002c, Table 7) estimates asthma prevalence for children ages 5 to 17 at 5.67% (based on data from the 1999 National Health Interview Survey).

PM_{10}

Single Pollutant Model

The estimated logistic coefficient and standard error are based on the odds ratio (1.4) and 95% confidence interval (1.1-1.8) associated with an increase in annual average PM_{10} of 19 $\mu g/m^3$. (McConnell et al., 1999, Table 3)

Functional Form: Logistic
Coefficient: 0.017709
Standard Error: 0.006612
Incidence Rate: annual incidence rate of one or more episodes of bronchitis per asthmatic = 0.326 (McConnell et al., 1999, Table 2)
Population: population of asthmatics ages 9 to 15 = 5.67%[111] of population ages 9 to 15

F .7.2 Asthma Attacks (Whittemore and Korn, 1980)

Whittemore and Korn (1980) examined the relationship between air pollution and asthma attacks in a survey of 443 children and adults, living in six communities in southern California during three 34-week periods in 1972-1975. The analysis focused on TSP and oxidants (O_x). Respirable PM, NO_2, SO_2 were highly correlated with TSP and excluded from the analysis. In a two pollutant model, daily levels of both TSP and oxidants were significantly related to reported asthma attacks. The results from this model were used, and the oxidant result was adjusted so it may be used with ozone data.

Multipollutant Model (PM_{10} and ozone)

The PM_{10} C-R function is based on the results of a co-pollutant model of TSP and ozone (Whittemore and Korn, 1980, Table 5). Assuming that PM_{10} is 55 percent of TSP[112] and that particulates greater than ten micrometers are harmless, the coefficient is calculated by dividing the TSP coefficient (0.00079) by 0.55. The standard error is calculated from the two-tailed p-value (<0.01) reported by Whittemore and Korn (1980, Table 5), which implies a t-value of at least 2.576 (assuming a large number of degrees of freedom).

Functional Form: Logistic
Coefficient: 0.001436
Standard Error: 0.000558
Incidence Rate: daily incidence of asthma attacks = 0.0550[113]
Population: population of asthmatics of all ages = 3.86% of the population of all ages (American Lung Association, 2002c, Table 7)

---

[111] The American Lung Association (2002c, Table 7) estimates asthma prevalence for children ages 5 to 17 at 5.67% (based on data from the 1999 National Health Interview Survey).

[112] The conversion of TSP to PM_{10} is from ESEERCO (1994, p. V-5), who cited studies by EPA (1986) and the California Air Resources Board (1982).

[113] Based on an analysis of the 1999 National Health Interview Survey, the daily incidence of wheezing attacks for adult asthmatics is estimated to be 0.0550. In the same survey, wheezing attacks for children were examined, however, the number of wheezing attacks per year were censored at 12 (compared to censoring at 95 for adults). Due to the potential for underestimation of the number of children's wheezing attacks, we used the adult rate for all individuals.

F .7.3 Asthma Exacerbation, Cough (Ostro et al., 2001)

Ostro et al. (2001) studied the relation between air pollution in Los Angeles and asthma exacerbation in African-American children (8 to 13 years old) from August to November 1993. They used air quality data for PM_{10}, $PM_{2.5}$, NO_2, and O_3 in a logistic regression model with control for age, income, time trends, and temperature-related weather effects.[114] Asthma symptom endpoints were defined in two ways: "probability of a day with symptoms" and "onset of symptom episodes". New onset of a symptom episode was defined as a day with symptoms followed by a symptom-free day. The authors found cough prevalence associated with PM_{10} and $PM_{2.5}$ and cough incidence associated with $PM_{2.5}$ PM_{10}, and NO_2. Ozone was not significantly associated with cough among asthmatics. The PM C-R functions are based on the results of single pollutant models looking at both the probability of symptoms and the onset of new symptoms.

$PM_{2.5}$ Function(s)

Single Pollutant Model (probability of symptoms)

The coefficient and standard error are based on an odds ratio of 1.03 (95% CI 0.98-1.07) for a 30 $\mu g/m^3$ increase in 12-hour average $PM_{2.5}$ concentration (Ostro et al., 2001, Table 4, p.204).

Functional Form: Logistic
Coefficient: 0.000985
Standard Error: 0.000747
Incidence Rate: daily cough rate per person (Ostro et al., 2001, p.202) = 0.145
Population: asthmatic African-American population ages 8 to 13 = 7.26%[115] of African-American population ages 8 to 13

Single Pollutant Model (probability of a new onset of symptoms)

The coefficient and standard error are based on an odds ratio of 1.10 (95% CI 1.03-1.18) for a 30 $\mu g/m^3$ increase in 12-hour average $PM_{2.5}$ concentration.

The C-R function based on this model will estimate the number of new onset episodes of cough avoided. In order to convert this estimate to the total number of episodes avoided, the results are adjusted by an estimate of the duration of symptom episodes. The average duration can be estimated from Ostro et al. (2001) using the ratio of the probability of a symptom episode to the probability of a new onset episode. For cough, this ratio is 2.2 (14.5% divided by 6.7%) (Ostro et al., 2001, p.202).

In addition, not all children are at-risk for a new onset of cough, as defined by the study. On average, 14.5% of African-American asthmatics have cough on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day are at-risk for a new onset episode (1-0.145 =

[114] The authors note that there were 26 days in which $PM_{2.5}$ concentrations were reported higher than PM_{10} concentrations. The majority of results the authors reported were based on the full dataset. These results were used for the basis for the C-R functions.

[115] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

85.5%). As a result, a factor of 85.5% is used in the function to estimate the population of African-American 8 to 13 year old children at-risk for a new cough episode.

Functional Form: Logistic
Coefficient: 0.003177
Standard Error: 0.001156
Incidence Rate: daily new onset cough (incidence) rate per person (Ostro et al., 2001, p.202) = 0.067
Population: asthmatic African-American population ages 8 to 13 at-risk for a new episode of cough = 6.21% of African-American population ages 8 to 13 multiplied (85.5% at-risk[116] times 7.26% asthmatic[117])
Adjustment Factor: average number of consecutive days with a cough episode (days) = 2.2

PM_{10} Function(s)

Single Pollutant Model (probability of symptoms)

The coefficient and standard error are based on an odds ratio of 1.10 (95% CI 1.04-1.16) for a 17 $\mu g/m^3$ increase in daily average PM_{10} concentration (Ostro et al., 2001, Table 4, p.204).

Functional Form: Logistic
Coefficient: 0.005606
Standard Error: 0.001639
Incidence Rate: daily cough rate per person (Ostro et al., 2001, p.202) = 0.145
Population: asthmatic African-American population ages 8 to 13 = 7.26%[118] of African-American population ages 8 to 13

Single Pollutant Model (probability of a new onset of symptoms)

The coefficient and standard error are based on an odds ratio of 1.25 (95% CI 1.12-1.39) for a 17 $\mu g/m^3$ increase in daily average PM_{10} concentration (Ostro et al., 2001, Table 5, p.204).

The C-R function based on this model will estimate the number of new onset episodes of cough avoided. In order to convert this estimate to the total number of episodes avoided, the results are adjusted by an estimate of the duration of symptom episodes. The average duration can be estimated from Ostro et al. (2001) using the ratio of the probability of a symptom episode to the probability of a new onset episode. For cough, this ratio is 2.2 (14.5% divided by 6.7%) (Ostro et al., 2001, p.202).

In addition, not all children are at-risk for a new onset of cough, as defined by the study. On average, 14.5% of African-American asthmatics have cough on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day are at-risk for a new onset episode (1-0.145 = 85.5%). As a result, a factor of 85.5% is used in the function to estimate the population of African-American 8 to 13 year old children at-risk for a new cough episode.

---

[116] On average, 17.3% of African-American asthmatics have cough episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day (1-0.145 = 85.5%) are at-risk for a new onset episode.

[117] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

[118] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

Functional Form: Logistic
Coefficient: 0.013126
Standard Error: 0.003241
Incidence Rate: daily new onset cough (incidence) rate per person (Ostro et al., 2001, p.202) = 0.067
Population: asthmatic African-American population ages 8 to 13 at-risk for a new episode of cough = 6.21% of African-American population ages 8 to 13 multiplied (85.5% at-risk[119] times 7.26% asthmatic[120])
Adjustment Factor: average number of consecutive days with a cough episode (days) = 2.2

F .7.4 Asthma Exacerbation, Cough (Vedal et al., 1998)

Vedal et al. (1998) studied the relationship between air pollution and respiratory symptoms among asthmatics and non-asthmatic children (ages 6 to 13) in Port Alberni, British Columbia, Canada. Four groups of elementary school children were sampled from a prior cross-sectional study: (1) all children with current asthma, (2) children without doctor diagnosed asthma who experienced a drop in FEV after exercise, (3) children not in groups 1 or 2 who had evidence of airway obstruction, and (4) a control group of children with matched by classroom. The authors used logistic regression and generalized estimating equations to examine the association between daily PM_{10} levels and daily increases in various respiratory symptoms among these groups. In the entire sample of children, PM_{10} was significantly associated with cough, phlegm, nose symptoms, and throat soreness. Among children with diagnosed asthma, the authors report a significant association between PM_{10} and cough symptoms, while no consistent effects were observed in the other groups. Since the study population has an over-representation of asthmatics, due to the sampling strategy, the results from the full sample of children are not generalizeable to the entire population. The C-R function presented below is based on results among asthmatics only.

Single Pollutant Model

The PM_{10} coefficient and standard error are based on an increase in odds of 8% (95% CI 0-16%) reported in the abstract for a 10 $\mu g/m^3$ increase in daily average PM_{10}.

Functional Form: Logistic
Coefficient: 0.007696
Standard Error: 0.003786
Incidence Rate: daily cough rate per person (Vedal et al., 1998, Table 1, p. 1038) = 0.086
Population: asthmatic population ages 6 to 13 = 5.67%[121] of population ages 6 to 13

F .7.5 Asthma Exacerbation, Moderate or Worse (Ostro et al., 1991)

Ostro et al. (1991) examined the effect of air pollution on asthmatics, ages 18 to 70, living in Denver, Colorado from December 1987 to February 1988. The respondents in this study were asked to

[119] On average, 17.3% of African-American asthmatics have cough episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day (1-0.145 = 85.5%) are at-risk for a new onset episode.

[120] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

[121] The American Lung Association (2002c, Table 7) estimates asthma prevalence for children 5-17 at 5.67% (based on data from the 1999 National Health Interview Survey).

record daily a subjective rating of their overall asthma status each day (0=none, 1=mild, 2=moderate, 3=severe, 4=incapacitating). Ostro et al. then examined the relationship between moderate (or worse) asthma and H^+, sulfate, SO_2, $PM_{2.5}$, estimated $PM_{2.5}$, PM_{10}, nitrate, and nitric acid. Daily levels of H^+ were linked to cough, asthma, and shortness of breath. $PM_{2.5}$ was linked to asthma. Sulfate was linked to shortness of breath. No effects seen for other pollutants. The C-R function is based on a single-pollutant linear regression model where the log of the pollutant is used.

Single Pollutant Model

Two $PM_{2.5}$ coefficients are presented, both equal 0.0006, however only one is significant. The coefficient based on data that does not include estimates of missing $PM_{2.5}$ values is not significant (std error = 0.0053); the coefficient that includes estimates of missing $PM_{2.5}$ values (estimated using a function of sulfate and nitrate) is significant at $p < 0.5$ (std error = 0.0003). The latter coefficient is used here. The C-R function to estimate the change in the number of days with moderate (or worse) asthma is as follows:

$$\Delta Days\ Moderate/Worse\ Asthma = -\beta \ln\left(\frac{PM_{2.5,\,after}}{PM_{2.5,\,before}}\right)\cdot pop\ ,$$

Functional Form: Linear (using log of the pollutant)
Coefficient: 0.0006
Standard Error: 0.0003
Population: population of asthmatics of all ages[122] = 3.86% of the population of all ages (American Lung Association, 2002c, Table 7)

F .7.6 Asthma Exacerbation, One or More Symptoms (Yu et al., 2000)

Yu et al. (2000) examined the association between air pollution and asthmatic symptoms among mild to moderate asthmatic children ages 5-13 in Seattle. They collected air quality data for CO, SO_2, PM_{10}, and $PM_{1.0}$ and asked study subjects to record symptoms daily. They used logistic regression models with generalized estimating equations in two different approaches. A "marginal approach" was used to estimate the impact of air pollution on asthma symptoms and a "transition approach" was used to estimate the association conditioned on the previous day's outcome. The primary endpoint, odds of at least one asthma symptom, was significantly associated with CO, PM_{10}, and $PM_{1.0}$ in single pollutant models. In multipollutant models, CO remained significant while PM effects declined slightly. The magnitude of the effects were similar between the "marginal" and "transition" approaches. The C-R function is based on the results of the "transition approach," where the previous day's symptoms is an explanatory variable.

Single Pollutant Model

The single pollutant PM_{10} coefficient and standard error are based on the odds ratio (1.10) and 95% confidence interval (1.03-1.16) for a 10 $\mu g/m^3$ increase in one-day lagged daily average PM_{10} (Yu et al., 2000, Table 4, p. 1212).

[122] The C-R function is applied to asthmatics of all ages, although the study population consists of asthmatics between the ages of 18 and 70. It seems reasonable to assume that individuals over the age of 70 are at least as susceptible as individuals in the study population. It also seems reasonable to assume that individuals under the age of 18 are also susceptible. For example, controlling for oxidant levels, Whittemore and Korn (1980) found TSP significantly related to asthma attacks in a study population comprised primarily (59 percent) of individuals less than 16 years of age.

Functional Form: Logistic
Coefficient: 0.009531
Standard Error: 0.003032
Incidence Rate: daily rate of at least one asthma episode per person (Yu et al., 2000, Table 2, p. 1212) = 0.60
Population: asthmatic population ages 5 to 13 = 5.67%[123] of population ages 5 to 13

Multipollutant Model (PM_{10}, CO, SO_2)

The C-R function is based on the results of the "transition approach," where the previous day's symptoms is an explanatory variable. The multipollutant PM_{10} coefficient and standard error are based on the odds ratio (1.05) and 95% confidence interval (0.95-1.16) for a 10 μg/m^3 increase in one-day lagged daily average PM_{10} (Yu et al., 2000, Table 4, p. 1212).

Functional Form: Logistic
Coefficient: 0.004879
Standard Error: 0.005095
Incidence Rate: daily rate of at least one asthma episode per person (Yu et al., 2000, Table 2, p. 1212) = 0.60
Population: asthmatic population ages 5 to 13 = 5.67%[124] of population ages 5 to 13

F .7.7 Asthma Exacerbation, Shortness of Breath (Ostro et al., 1995)

Using a logistic regression estimation, Ostro et al. (1995) estimated the impact of PM_{10}, ozone, NO_2, and SO_2 on the incidence of coughing, shortness of breath, and wheezing in 83 African-American asthmatic children ages 7-12 living in Los Angeles from August through September 1992. Regression results show both PM_{10} and ozone significantly linked to shortness of breath; the beginning of an asthma episode was also significantly linked to ozone. No effect was seen for NO_2 and SO_2. Results for single-pollutant models only were presented in the published paper. The C-R function is based on the model with adjustment for respiratory infection, temperature, and outdoor mold levels.

Single Pollutant Model

The PM_{10} coefficient and standard error are based on the odds ratio (1.60) and 95% confidence interval (1.07-2.37) (Ostro et al., 1995, Table 3) associated with a change in daily mean PM_{10} of 55.87 μg/m^3 (Ostro et al., 1995, Table 2).

[123] The American Lung Association (2002c, Table 7) estimates asthma prevalence for children 5 to 17 at 5.67% (based on data from the 1999 National Health Interview Survey).

[124] The American Lung Association (2002c, Table 7) estimates asthma prevalence for children 5 to 17 at 5.67% (based on data from the 1999 National Health Interview Survey).

Functional Form: Logistic
Coefficient: 0.008412
Standard Error: 0.003631
Incidence Rate: daily shortness of breath incidence rate per person (Ostro et al., 1995, p. 715) = 0.056
Population: asthmatic African-American population ages 7 to 12 = 7.26%[125] of African-American population ages 7 to 12

F .7.8 Asthma Exacerbation, Shortness of Breath (Ostro et al., 2001)

Ostro et al. (2001) studied the relationship between air pollution in Los Angeles and asthma exacerbation in African-American children (8 to 13 years old) from August to November 1993. They used air quality data for PM_{10}, $PM_{2.5}$, NO_2, and ozone in a logistic regression model with control for age, income, time trends, and temperature-related weather effects. Asthma symptom endpoints were defined in two ways: "probability of a day with symptoms" and "new onset of a symptom episode". New onset of a symptom episode was defined as a day with symptoms followed by a symptom-free day. The authors found that both the prevalent and incident episodes of shortness of breath were associated with $PM_{2.5}$ and PM_{10}. Neither ozone nor NO_2 were significantly associated with shortness of breath among asthmatics. The PM C-R functions are based on the results of single pollutant models looking at both the probability of symptoms and the onset of new symptoms.

$PM_{2.5}$ Function(s)

Single Pollutant Model (probability of symptoms)

The coefficient and standard error are based on an odds ratio of 1.08 (95% CI 1.00-1.17) for a 30 $\mu g/m^3$ increase in 12-hour average $PM_{2.5}$ concentration (Ostro et al., 2001, Table 4, p.204).

Functional Form: Logistic
Coefficient: 0.002565
Standard Error: 0.001335
Incidence Rate: daily shortness of breath rate per person (Ostro et al., 2001, p.202) = 0.074
Population: asthmatic African-American population ages 8 to 13 = 7.26%[126] of African-American population ages 8 to 13

Single Pollutant Model (probability of a new onset of symptoms)

The coefficient and standard error are based on an odds ratio of 1.10 (95% CI 1.00-1.20) for a 30 $\mu g/m^3$ increase in 12-hour average $PM_{2.5}$ concentration (Ostro et al., 2001, Table 5, p.204).

The C-R function based on this model will estimate the number of new onset episodes of shortness of breath avoided. In order to convert this estimate to the total number of episodes avoided, the results are adjusted by an estimate of the duration of symptom episodes. The average duration can be estimated from Ostro et al. (2001) using the ratio of the probability of a symptom episode to the probability of a new onset episode. For shortness of breath, this ratio is 2.0 (7.4% divided by 3.7%) (Ostro et al., 2001, p.202).

[125] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

[126] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

In addition, not all children are at-risk for a new onset of shortness of breath, as defined by the study. On average, 7.4% of African-American asthmatics have shortness of breath episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day are at-risk for a new onset episode (1-0.074 = 92.6%). As a result, a factor of 92.6% is used in the function to estimate the population of African-American 8 to 13 year old children at-risk for a new shortness of breath episode.

Functional Form: Logistic
Coefficient: 0.003177
Standard Error: 0.001550
Incidence Rate: daily new onset shortness of breath (incidence) rate per person (Ostro et al., 2001, p.202) = 0.037
Population: asthmatic African-American population ages 8 to 13 at-risk for a new episode of shortness of breath = 6.72% of African-American population ages 8 to 13 multiplied (92.6% at-risk[127] times 7.26% asthmatic[128])
Adjustment Factor: average number of consecutive days with a shortness of breath episode (days) = 2.0

PM_{10} Function(s)

Single Pollutant Model (probability of symptoms)

The coefficient and standard error are based on an odds ratio of 1.14 (95% CI 1.04-1.24) for a 17 $\mu g/m^3$ increase in daily average PM_{10} concentration (Ostro et al., 2001, Table 4, p.204).

Functional Form: Logistic
Coefficient: 0.007708
Standard Error: 0.002639
Incidence Rate: daily shortness of breath rate per person (Ostro et al., 2001, p.202) = 0.074
Population: asthmatic African-American population ages 8 to 13 = 7.26%[129] of African-American population ages 8 to 13

Single Pollutant Model (probability of a new onset of symptoms)

The coefficient and standard error are based on an odds ratio of 1.20 (95% CI 1.06-1.37) for a 17 $\mu g/m^3$ increase in daily average PM_{10} concentration (Ostro et al., 2001, Table 5, p.204).

The C-R function based on this model will estimate the number of new onset episodes of shortness of breath avoided. In order to convert this estimate to the total number of episodes avoided, the results are adjusted by an estimate of the duration of symptom episodes. The average duration can be estimated from Ostro et al. (2001) using the ratio of the probability of a symptom episode to the probability of a new onset episode. For shortness of breath, this ratio is 2.0 (7.4% divided by 3.7%) (Ostro et al., 2001, p.202).

[127] On average, 7.4% of African-American asthmatics have shortness of breath episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day (1-0.074 = 92.6%) are at-risk for a new onset episode.

[128] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

[129] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

In addition, not all children are at-risk for a new onset of shortness of breath, as defined by the study. On average, 7.4% of African-American asthmatics have shortness of breath episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day are at-risk for a new onset episode (1-0.074 = 92.6%). As a result, a factor of 92.6% is used in the function to estimate the population of African-American 8 to 13 year old children at-risk for a new shortness of breath episode.

Functional Form: Logistic
Coefficient: 0.010725
Standard Error: 0.003850
Incidence Rate: daily new onset shortness of breath (incidence) rate per person (Ostro et al., 2001, p.202) = 0.037
Population: asthmatic African-American population ages 8 to 13 at-risk for a new episode of shortness of breath = 6.72% of African-American population ages 8 to 13 multiplied (92.6% at-risk[130] times 7.26% asthmatic[131])
Adjustment Factor: average number of consecutive days with a shortness of breath episode (days) = 2.0

F .7.9 Asthma Exacerbation, Wheeze (Ostro et al., 2001)

Ostro et al. (2001) studied the relation between air pollution in Los Angeles and asthma exacerbation in African-American children (8 to 13 years old) from August to November 1993. They used air quality data for PM_{10}, $PM_{2.5}$, NO_2, and O_3 in a logistic regression model with control for age, income, time trends, and temperature-related weather effects. Asthma symptom endpoints were defined in two ways: "probability of a day with symptoms" and "onset of symptom episodes". New onset of a symptom episode was defined as a day with symptoms followed by a symptom-free day. The authors found both the prevalence and incidence of wheeze associated with $PM_{2.5}$ PM_{10}, and NO_2. Ozone was not significantly associated with wheeze among asthmatics. The PM C-R functions are based on the results of single pollutant models looking at both the probability of symptoms and the onset of new symptoms.

$PM_{2.5}$ Function(s)

Single Pollutant Model (probability of symptoms)

The coefficient and standard error are based on an odds ratio of 1.06 (95% CI 1.01-1.11) for a 30 $\mu g/m^3$ increase in 12-hour average $PM_{2.5}$ concentration (Ostro et al., 2001, Table 4, p.204).

Functional Form: Logistic
Coefficient: 0.001942
Standard Error: 0.000803
Incidence Rate: daily wheeze rate per person (Ostro et al., 2001, p.202) = 0.173
Population: asthmatic African-American population ages 8 to 13 = 7.26%[132] of African-American population ages 8 to 13

[130] On average, 7.4% of African-American asthmatics have shortness of breath episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day (1-0.074 = 92.6%) are at-risk for a new onset episode.

[131] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

[132] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

Single Pollutant Model (probability of a new onset of symptoms)

The coefficient and standard error are based on an odds ratio of 1.08 (95% CI 1.01-1.14) for a 30 $\mu g/m^3$ increase in 12-hour average $PM_{2.5}$ concentration (Ostro et al., 2001, Table 5, p.204).

The C-R function based on this model will estimate the number of new onset episodes of wheeze avoided. In order to convert this estimate to the total number of episodes avoided, the results are adjusted by an estimate of the duration of symptom episodes. The average duration can be estimated from Ostro et al. (2001) using the ratio of the probability of a symptom episode to the probability of a new onset episode. For wheeze, this ratio is 2.3 (17.3% divided by 7.6%) (Ostro et al., 2001, p.202).

In addition, not all children are at-risk for a new onset of wheeze, as defined by the study. On average, 17.3% of African-American asthmatics have wheeze on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day are at-risk for a new onset episode (1-0.173 = 82.7%). As a result, a factor of 82.7% is used in the function to estimate the population of African-American 8 to 13 year old children at-risk for a new wheeze episode.

Functional Form: Logistic
Coefficient: 0.002565
Standard Error: 0.001030
Incidence Rate: daily new onset wheeze (incidence) rate per person (Ostro et al., 2001, p.202) = 0.076
Population: asthmatic African-American population ages 8 to 13 at-risk for a new episode of wheeze = 6.00% of African-American population ages 8 to 13 multiplied (82.7% at-risk[133] times 7.26% asthmatic[134])
Adjustment Factor: average number of consecutive days with a wheeze episode (days) = 2.3

PM_{10} Function(s)

Single Pollutant Model (probability of symptoms)

The coefficient and standard error are based on an odds ratio of 1.04 (95% CI 0.98-1.10) for a 17 $\mu g/m^3$ increase in daily average PM_{10} concentration (Ostro et al., 2001, Table 4, p.204).

Functional Form: Logistic
Coefficient: 0.002307
Standard Error: 0.001733
Incidence Rate: daily wheeze rate per person (Ostro et al., 2001, p.202) = 0.173
Population: asthmatic African-American population ages 8 to 13 = 7.26%[135] of African-American population ages 8 to 13

[133] On average, 17.3% of African-American asthmatics have wheeze episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day (1-0.173 = 82.7%) are at-risk for a new onset episode.

[134] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

[135] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

Single Pollutant Model (probability of a new onset of symptoms)

The coefficient and standard error are based on an odds ratio of 1.12 (95% CI 1.01-1.23) for a 17 $\mu g/m^3$ increase in daily average PM_{10} concentration (Ostro et al., 2001, Table 5, p.204).

The C-R function based on this model will estimate the number of new onset episodes of wheeze avoided. In order to convert this estimate to the total number of episodes avoided, the results are adjusted by an estimate of the duration of symptom episodes. The average duration can be estimated from Ostro et al. (2001) using the ratio of the probability of a symptom episode to the probability of a new onset episode. For wheeze, this ratio is 2.3 (17.3% divided by 7.6%) (Ostro et al., 2001, p.202).

In addition, not all children are at-risk for a new onset of wheeze, as defined by the study. On average, 17.3% of African-American asthmatics have wheeze on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day are at-risk for a new onset episode (1-0.173 = 82.7%). As a result, a factor of 82.7% is used in the function to estimate the population of African-American 8 to 13 year old children at-risk for a new wheeze episode.

Functional Form: Logistic
Coefficient: 0.006666
Standard Error: 0.002957
Incidence Rate: daily new onset wheeze (incidence) rate per person (Ostro et al., 2001, p.202) = 0.076
Population: asthmatic African-American population ages 8 to 13 at-risk for a new episode of wheeze = 6.00% of African-American population ages 8 to 13 multiplied (82.7% at-risk[136] times 7.26% asthmatic[137])
Adjustment Factor: average number of consecutive days with a wheeze episode (days) = 2.3

F .7.10 Chronic Phlegm (McConnell et al., 1999)

McConnell et al. (1999) examined the relationship between air pollution and bronchitic symptoms among asthmatic 4th, 7th, and 10th grade children in southern California.[138] The authors collected information on the prevalence of bronchitis, chronic cough, and chronic phlegm among children with and without a history of asthma and/or wheeze. They used annual measurements of ozone, PM_{10}, $PM_{2.5}$, NO_2, and acids in a logistic regression model with adjustments for personal covariates. Neither bronchitis, cough, or phlegm were associated with any of the pollutants among children with no history of wheeze or asthma or a history of wheeze without diagnosed asthma. Among asthmatics, PM_{10} was significantly associated with bronchitis and phlegm; $PM_{2.5}$ was significantly associated with phlegm and marginally associated with bronchitis; NO_2 and acids were both significantly associated with phlegm; and ozone was not significantly associated with any of the endpoints.

Phlegm was defined in the study by the question: "Other than with colds, does this child usually seem congested in the chest or bring up phlegm?" (McConnell et al., 1999, p. 757). The authors refer to this definition as "chronic phlegm" and we also assume that the term "usually" refers to chronic, rather

[136] On average, 17.3% of African-American asthmatics have wheeze episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day (1-0.173 = 82.7%) are at-risk for a new onset episode.

[137] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

[138] Assuming that a child enters kindergarten at age 5, 4th grade corresponds to age 9 and 10th grade corresponds to age 15. We therefore applied the results of this study to children ages 9 to 15.

than acute, phlegm. The PM C-R functions for chronic phlegm among asthmatics are based on the results of the single pollutant model reported in Table 3.

$PM_{2.5}$

Single Pollutant Model

The estimated logistic coefficient and standard error are based on the odds ratio (2.6) and 95% confidence interval (1.2-5.4) associated with an increase in yearly mean 2-week average $PM_{2.5}$ of 15 $\mu g/m^3$. (McConnell et al., 1999, Table 3)

Functional Form: Logistic
Coefficient: 0.063701
Standard Error: 0.025580
Incidence Rate: annual incidence rate of phlegm per asthmatic = 0.257 (McConnell et al., 1999, Table 2)
Population: population of asthmatics ages 9 to 15 = 5.67%[139] of population ages 9 to 15

PM_{10}

Single Pollutant Model

The estimated logistic coefficient and standard error are based on the odds ratio (2.1) and 95% confidence interval (1.4-3.3) associated with an increase in annual average PM_{10} of 19 $\mu g/m^3$. (McConnell et al., 1999, Table 3)

Functional Form: Logistic
Coefficient: 0.039049
Standard Error: 0.011512
Incidence Rate: annual incidence rate of phlegm per asthmatic = 0.257 (McConnell et al., 1999, Table 2)
Population: population of asthmatics ages 9 to 15 = 5.67%[140] of population ages 9 to 15

F .7.11 Upper Respiratory Symptoms (Pope et al., 1991)

Using logistic regression, Pope et al. (1991) estimated the impact of PM_{10} on the incidence of a variety of minor symptoms in 55 subjects (34 "school-based" and 21 "patient-based") living in the Utah Valley from December 1989 through March 1990. The children in the Pope et al. study were asked to record respiratory symptoms in a daily diary. With this information, the daily occurrences of upper respiratory symptoms (URS) and lower respiratory symptoms (LRS) were related to daily PM_{10} concentrations. Pope et al. describe URS as consisting of one or more of the following symptoms: runny or stuffy nose; wet cough; and burning, aching, or red eyes. Levels of ozone, NO_2, and SO_2 were reported low during this period, and were not included in the analysis. The sample in this study is relatively small and is most representative of the asthmatic population, rather than the general population. The school-based subjects (ranging in age from 9 to 11) were chosen based on "a positive response to one or more of

[139] The American Lung Association (2002c, Table 7) estimates asthma prevalence for children ages 5 to 17 at 5.67% (based on data from the 1999 National Health Interview Survey).

[140] The American Lung Association (2002c, Table 7) estimates asthma prevalence for children ages 5 to 17 at 5.67% (based on data from the 1999 National Health Interview Survey).

three questions: ever wheezed without a cold, wheezed for 3 days or more out of the week for a month or longer, and/or had a doctor say the 'child has asthma' (Pope et al., 1991, p. 669)." The patient-based subjects (ranging in age from 8 to 72) were receiving treatment for asthma and were referred by local physicians. Regression results for the school-based sample (Pope et al., 1991, Table 5) show PM_{10} significantly associated with both upper and lower respiratory symptoms. The patient-based sample did not find a significant PM_{10} effect. The results from the school-based sample are used here.

Single Pollutant Model

The coefficient and standard error for a one $\mu g/m^3$ change in PM_{10} is reported in Table 5.

Functional Form: Logistic
Coefficient: 0.0036
Standard Error: 0.0015
Incidence Rate: daily upper respiratory symptom incidence rate per person = 0.3419 (Pope et al., 1991, Table 2)
Population: asthmatic population ages 9 to 11 = 5.67%[141] of population ages 9 to 11

[141] The American Lung Association (2002c, Table 7) estimates asthma prevalence for children ages 5 to 17 at 5.67% (based on data from the 1999 National Health Interview Survey).

Appendix F. Particulate Matter C-R Functions

Exhibit F-8. Concentration-Response (C-R) Functions for Particulate Matter and Welfare Effects

Endpoint Name	Pollutant	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time	Beta	Std Error	Functional Form
Household Soiling Damage	PM_{10}	ESEERCO	1994	nationwide	All	All	All	None	Annual avg	–	--	--

F .8 Welfare Effects

F .8.1 Household Soiling Damage (ESEERCO, 1994)

Particulate matter air pollution has been shown to result in dirtier clothes, which in turn results in higher annual cleaning costs for consumers. One benefit of reduced particulate matter, then, is the consequent reduction in cleaning costs for consumers. Several studies have provided estimates of the cost to households of PM soiling. The study that is cited by ESEERCO (1994) as one of the most sophisticated and is relied upon by EPA in its 1988 Regulatory Impact Analysis for SO_2 is Manuel et al. (1982). Using a household production function approach and household expenditure data from the 1972-73 Bureau of Labor Statistics Consumer Expenditure Survey for over twenty cities in the United States, Manuel et al. estimated the annual cost of cleaning per $\mu g/m^3$ PM per household as $1.55 ($0.59 per person times 2.63 persons per household). This estimate is low compared with others (e.g., estimates provided by Cummings et al. (1985) and Watson and Jaksch (1982) are about eight times and five times greater, respectively). The ESEERCO report notes, however, that the Manuel estimate is probably downward biased because it does not include the time cost of do-it-yourselfers. Estimating that these costs may comprise at least half the cost of PM-related cleaning costs, they double the Manuel estimate to obtain a point estimate of $3.10 (reported by ESEERCO in 1992 dollars as $2.70).

The Manuel et al. (1982) study measured particulate matter as TSP rather than PM_{10} or $PM_{2.5}$. If a one $\mu g/m^3$ increase in TSP causes $1.55 worth of cleaning expenses per household, the same unit dollar value can be used for PM_{10} (or $PM_{2.5}$) only if particle size doesn't matter -- i.e., only if particles of all sizes are equally soiling. Suppose, for example, that PM_{10} is 75% of TSP and that all particles are equally soiling. Then 75% of the damage caused by a one $\mu g/m^3$ increase in TSP is due to PM_{10}. This is (0.75)($1.55) = $1.16. However, this corresponds to a 0.75 $\mu g/m^3$ increase in PM_{10}. A one $\mu g/m^3$ increase in PM_{10} would therefore yield a dollar soiling damage of $1.16/0.75 = $1.55.

Suppose, however, that only PM_{10} matters. Then the $1.55 underestimates the impact of a one $\mu g/m^3$ increase in PM_{10}, because it corresponds to a less than one $\mu g/m^3$ increase in PM_{10} (e.g., a 0.75 $\mu g/m^3$ increase in PM_{10}). In this case, the correct unit value per unit of PM_{10} would be ($1.55)/0.75 = $2.07. If only PM_{10} matters, then either (1) the dollar value can be adjusted by dividing it by the percentage of TSP that is PM_{10} and PM_{10} can be used in the soiling damage function, or (2) the dollar value can be left unadjusted and TSP, rather than PM_{10}, can be used in the soiling damage function.

Finally, it is possible that, while both PM_{10} and $PM_{2.5}$ are components of TSP that cause consumer cleaning costs, the remaining portion of TSP has a greater soiling capability than either the PM_{10} or $PM_{2.5}$ component. In this case, using either PM_{10} or $PM_{2.5}$ air quality data with a household soiling function based on TSP would yield overestimates of the PM_{10}- or $PM_{2.5}$-related consumer cleaning costs avoided by reductions in concentration of these pollutants.

There is, however, insufficient information on the relative soiling capabilities of the different components of TSP. We have assumed that all components of TSP have an equivalent soiling capacity.

Appendix G: Ozone Concentration-Response Functions

In this Appendix, we present the concentration-response (C-R) functions used to estimate ozone-related adverse health effects. Each sub-section has an Exhibit with a brief description of the C-R function and the underlying parameters. Following each Exhibit, we present a brief summary of each of the studies and any items that are unique to the study.

Note that the main text describes the methods that we used to choose these C-R functions from the wide range available in the literature. In addition, Appendix D mathematically derives the standard types of C-R functions that we encountered in the epidemiological literature, such as, log-linear, logistic and linear, so we simply note here the type of functional form. Finally, Appendix E presents a detailed description of the sources for the incidence and prevalence data used in these C-R functions.

Appendix G. Ozone C-R Functions

Exhibit G-1. Concentration-Response (C-R) Functions for Ozone and Short-Term Mortality

Endpoint Name	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Functional Form	Beta	Std Error	Notes
Non-Accidental	Fairley	2003	Santa Clara County	All	All	All	None	8-hr max	Log-linear	0.001558	0.000871	Reanalysis of Fairley, 1999
Non-Accidental	Fairley	2003	Santa Clara County	All	All	All	$PM_{2.5}$	8-hr max	Log-linear	0.002828	0.002668	Reanalysis of Fairley, 1999
Non-Accidental	Ito and Thurston	1996	Chicago, IL	All	All	All	None	1-hr max	Log-linear	0.000953	0.000208	
Non-Accidental	Ito and Thurston	1996	Chicago, IL	All	All	All	PM_{10}	1-hr max	Log-linear	0.000634	0.000251	
Non-Accidental	Kinney et al.	1995	Los Angeles, CA	All	All	All	None	1-hr max	Log-linear	0.000138	0.000087	
Non-Accidental	Kinney et al.	1995	Los Angeles, CA	All	All	All	PM_{10}	1-hr max	Log-linear	0	0.000214	
Non-Accidental	Moolgavkar et al.	1995	Philadelphia, PA	All	All	All	SO_2, TSP	24-hr avg	Log-linear	0.000611	0.000216	
Non-Accidental	Samet et al.	1997	Philadelphia, PA	All	All	All	None	24-hr avg	Log-linear	0.001115	0.000372	
Non-Accidental	Samet et al.	1997	Philadelphia, PA	All	All	All	CO, NO_2, SO_2, TSP	24-hr avg	Log-linear	0.000936	0.000312	
Non-Accidental	WHO Working Group	2003	Europe	All	All	All	–[2]	1-hr max	Log-linear	0.000784	0.000250	
Non-Accidental	WHO Working Group	2003	Europe	All	All	All	–[2]	8-hr avg	Log-linear	0.001174	0.000299	

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.
2. The WHO Working Group meta-study is unclear as to whether some of the studies used in the meta-study included other pollutants in their ozone models.

G .1 Short-term Mortality

Exhibit G-1 summarizes the C-R functions used to estimate the relationship between ozone and short-term mortality. Detailed summaries of each of the studies used to generate the functions are described below, along with the parameters used in each of the functions.

G .1.1 Short-Term Mortality, Non-Accidental (Fairley, 2003)

Using data from 1989-1996 in Santa Clara County, California, Fairley et al. (1999) examined the relationship between daily non-accidental mortality and fluctuations in a variety of pollutants, including $PM_{2.5}$, coarse PM_{10} (i.e., $PM_{2.5-10}$), nitrate (NO_3), SO_4, coefficient of haze (COH), ozone, CO, and NO_2. They reported that $PM_{2.5}$ and NO_3 were significant in single-pollutant models, as well as two-pollutant models. $PM_{2.5}$ was only insignificant when paired with PM_{10} and NO_3, and NO_3 was only insignificant when paired with $PM_{2.5}$. The other pollutants were insignificant when paired with either $PM_{2.5}$ or NO_3.

The analysis by Fairly et al. (1999) relied on a generalized additive model based on the Splus software. Because of potential bias from using Splus, Fairley (2003) conducted a reanalysis, and reported that the conclusions of the original study were unchanged. Both $PM_{2.5}$ and NO_3 appear significantly related to non-accidental mortality.

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the relative risk (1.031) and 95% confidence interval (0.997-1.066) reported for a 19.6 ppb increase in daily 8-hour maximum ozone concentration in the 0-day lag GAM stringent ('New GAM') model (Fairley, 2003, Table 1a).

Functional Form: Log-linear
Coefficient: 0.001558
Standard Error: 0.000871
Incidence Rate: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

Multipollutant Model (ozone and $PM_{2.5}$)

In a co-pollutant model with $PM_{2.5}$, the coefficient and standard error are based on the relative risk (1.057) and 95% confidence interval (0.954-1.171) reported for a 19.6 ppb increase in daily 8-hour maximum ozone concentration in the 0-day lag GAM stringent ('New GAM') model (Fairley, 2003, Table 1b).

Functional Form: Log-linear
Coefficient: 0.002828
Standard Error: 0.002668
Incidence Rate: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

G .1.2 Short-Term Mortality, Non-Accidental (Ito and Thurston, 1996, Chicago)

Ito and Thurston (1996) examined the relationship between daily non-accidental mortality and air pollution levels in Cook County, Illinois from 1985 to 1990. They examined daily levels of ozone, PM_{10}, SO_2, and CO, and found a significant relationship for ozone and PM_{10} with both pollutants in the model; no significant effects were found for SO_2 and CO. In single pollutant models the effects were slightly larger. The C-R functions for ozone are based on results from both the single and co-pollutant models.

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the relative risk (1.10) and 95% confidence interval (1.06-1.15) reported for a 100 ppb increase in daily one-hour maximum ozone concentration (Ito and Thurston, 1996, p. 87).

Functional Form: Log-linear
Coefficient: 0.000953
Standard Error: 0.000208
Incidence Rate: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

Multipollutant Model (ozone and PM_{10})

In a co-pollutant model with PM_{10}, the coefficient (0.000634) and standard error (0.000251) were obtained directly from the author because the published paper reported incorrect information.

Functional Form: Log-linear
Coefficient: 0.000634
Standard Error: 0.000251
Incidence Rate: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

G .1.3 Short-Term Mortality, Non-Accidental (Kinney et al., 1995, Los Angeles)

Kinney et al. (1995) examined the relationship between daily non-accidental mortality and air pollution levels in Los Angeles, California from 1985 to 1990. They examined ozone, PM_{10}, and CO, and found a significant relationship for each pollutant in single pollutant models. The effect for ozone dropped to zero with the inclusion of PM_{10} in the model, while the effect for CO and PM_{10} appeared co-pollutant ozone models.

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the relative risk (1.02) and 95% confidence interval (1.00-1.05) reported for a 143 ppb increase in daily one-hour maximum ozone levels (Kinney et al., 1995, Table 2, p. 64).

Functional Form: Log-linear
Coefficient: 0.000138
Standard Error: 0.000087
Incidence Rate: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

Multipollutant Model (ozone and PM_{10})

In a model with PM_{10}, the coefficient and standard error are based on the relative risk (1.00) and 95% confidence interval (0.94-1.06) reported for a 143 ppb increase in daily one-hour maximum ozone concentration (Kinney et al., 1995, Table 2, p. 64).

Functional Form: Log-linear
Coefficient: 0
Standard Error: 0.000214
Incidence: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

G .1.4 Short-Term Mortality, Non-Accidental (Moolgavkar et al., 1995, Philadelphia)

Moolgavkar et al. (1995) examined the relationship between daily non-accidental mortality and air pollution levels in Philadelphia, Pennsylvania from 1973 to 1988. They examined ozone, TSP, and SO_2 in a three-pollutant model, and found a significant relationship for ozone and SO_2; TSP was not significant. In season-specific models, ozone was significantly associated with mortality only in the summer months. The C-R function for ozone is based on the full-year three-pollutant model reported in Table 5 (Moolgavkar et al., 1995, p. 482).

Multipollutant Model (ozone, SO_2, TSP)

The coefficient and standard error are based on the relative risk (1.063) and 95% confidence interval (1.018-1.108) associated with a 100 ppb increase in daily average ozone (Moolgavkar et al., 1995, p. 482, Table 5).

Functional Form: Log-linear
Coefficient: 0.000611
Standard Error: 0.000216
Incidence Rate: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

G .1.5 Short-Term Mortality, Non-Accidental (Samet et al., 1997, Philadelphia)

Samet et al. (1997) examined the relationship between daily non-accidental mortality and air pollution levels in Philadelphia, Pennsylvania from 1974 to 1988. They examined ozone, TSP, SO_2, NO_2, and CO in a Poisson regression model. In single pollutant models, ozone, SO_2, TSP, and CO were significantly associated with mortality. In a five-pollutant model, they found a positive statistically significant relationship for each pollutant except NO_2. The C-R functions for ozone are based on the single pollutant and five-pollutant model (ozone, CO, NO_2, SO_2, and TSP) reported in Table 9 (Samet et al., 1997, p. 20).

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (2.28) and t-statistic (3) associated with a 20.219 ppb increase in two-day average ozone (Samet et al., 1997, p. 20, Table 9).

Functional Form: Log-linear
Coefficient: 0.001115
Standard Error: 0.000372
Incidence Rate: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

Multipollutant Model (ozone, CO, NO_2, SO_2, and TSP)

In a model with CO, NO_2, SO_2, and TSP, the ozone coefficient and standard error are based on the percent increase (1.91) and t-statistic (3) associated with a 20.219 ppb increase in two-day average ozone (Samet et al., 1997, p. 20, Table 9).

Functional Form: Log-linear
Coefficient: 0.000936
Standard Error: 0.000312
Incidence Rate: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages

G .1.6 Short-Term Mortality, Non-Accidental (World Health Organization (WHO) Working Group, 2003, Europe)

The World Health Organization (WHO 2003, p. 44) conducted a meta-analysis of time-series studies conducted between 1996 and 2001. The results of the analysis are preliminary, so we have presented it as an alternative estimate of the relationship between ozone and premature mortality. We consider two sets of results: one based on the 1-hour maximum and the other based on the 8-hour average.

1-Hour Maximum Model

In a model with the daily 1-hour maximum the relative risk is 1.004 associated with a 10 ug/m^3 change in ozone, with a 5^{th} and 95^{th} estimate of 1.001 and 1.006 (WHO 2003, p. 44). In calculating the coefficient and standard error for the C-R function, we assume a conversion of 1.963 ug/m^3 per ppb. This is the standard conversion at 25° C and one atmosphere. We have used this function with a population of all ages, as well as just the population ages 65 and up.

Functional Form: Log-linear
Coefficient: 0.000784
Standard Error: 0.000250
Incidence Rate: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages & ages 65+

8-Hour Average Model

In a model with the daily 8-hour average the relative risk is 1.006 associated with a 10 ug/m^3 change in ozone, with a 5^{th} and 95^{th} estimate of 1.003 and 1.009 (WHO 2003, p. 44). The paper is not completely clear on how the 8-hour average should be calculated, so we have assumed the average between the hours of 9:00 am and 4:59 pm. In calculating the coefficient and standard error for the C-R function, we assume a conversion of 1.963 ug/m^3 per ppb. This is the standard conversion at 25° C and one atmosphere. We have used this function with a population of all ages, as well as just the population ages 65 and up.

Functional Form: Log-linear
Coefficient: 0.001174
Standard Error: 0.000299
Incidence Rate: county-level daily non-accidental mortality rate (ICD codes <800) per person
Population: population of all ages & ages 65+

Appendix G. Ozone C-R Functions

Exhibit G-2. Concentration-Response (C-R) Functions for Ozone and Chronic Illness

Endpoint Name	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Functional Form	Beta	Std Error
Chronic Asthma	McDonnell et al.	1999	SF, SD, South Coast Air Basin	27+	All	Male	None	annual avg 8-hr avg	Logistic	0.0277	0.0135

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.

G .2 Chronic Illness

Exhibit G-2 summarizes the C-R function ((McDonnell et al., 1999)) used to estimate the relationship between ozone and chronic asthma. A more detailed summary of McDonnell et al. (1999), and the parameters used in the function, is described below.

G .2.1 Chronic Asthma (McDonnell et al., 1999)

McDonnell et al. (1999) used the same cohort of Seventh-Day Adventists as Abbey et al. (1995b), and examined the association between air pollution and the onset of asthma in adults between 1977 and 1992. Males who did not report doctor-diagnosed asthma in 1977, but reported it in 1987 or 1992, had significantly higher ozone exposures, controlling for other covariates; no significant effect was found between ozone exposure and asthma in females. No significant effect was reported for females or males due to exposure to PM, NO_2, SO_2, or SO_4. The C-R function for ozone is based on the single pollutant model for males reported in Table 5 (McDonnell et al., 1999, 1999, p. 117).

Single Pollutant Model

The coefficient and standard error for males is reported in Table 5 for a unit increase in annual average eight-hour ozone concentrations.[142]

Functional Form: Logistic
Coefficient: 0.0277
Standard Error: 0.0135
Incidence Rate: annual asthma incidence rate per person = 0.00219 (McDonnell et al., 1999, 1999, Table 4)
Population: non-asthmatic males age 27 and over = 97.9%[143] of males 27+

[142] The eight-hour ozone concentration is defined as 9:00 A.M. to 4:59 P.M. The study used the 1973-1992 mean 8-hour average ambient ozone concentration (McDonnell et al., 1999, p. 113).

[143] The prevalence of asthma among males 27 and older (2.10 percent) was estimated from the 2000 National Health Interview Survey (NHIS) public use data, available at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000.

Appendix G. Ozone C-R Functions

Exhibit G-3. Concentration-Response (C-R) Functions for Ozone and Hospital Admissions

Endpoint Name	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Functional Form	Beta	Std Error
All Respiratory	Burnett et al.	1997	Toronto, CAN	All	All	All	None	8-hr avg	Log-linear	0.005394	0.001052
All Respiratory	Burnett et al.	1997	Toronto, CAN	All	All	All	$PM_{2.5}$	8-hr avg	Log-linear	0.004985	0.001093
All Respiratory	Burnett et al.	1997	Toronto, CAN	All	All	All	$PM_{10-2.5}$	8-hr avg	Log-linear	0.005231	0.001070
All Respiratory	Burnett et al.	1997	Toronto, CAN	All	All	All	NO_2, $PM_{2.5}$, SO_2	8-hr avg	Log-linear	0.004985	0.001070
All Respiratory	Burnett et al.	2001	Toronto, CAN	<2	All	All	None	1-hr max	Log-linear	0.006607	0.001378
All Respiratory	Burnett et al.	2001	Toronto, CAN	<2	All	All	$PM_{2.5}$	1-hr max	Log-linear	0.006309	0.001834
All Respiratory	Burnett et al.	2001	Toronto, CAN	<2	All	All	$PM_{10-2.5}$	1-hr max	Log-linear	0.005702	0.001901
All Respiratory	Schwartz	1995	New Haven, CT	65+	All	All	None	24-hr avg	Log-linear	0.002284	0.001323
All Respiratory	Schwartz	1995	New Haven, CT	65+	All	All	PM_{10}	24-hr avg	Log-linear	0.002652	0.001398
All Respiratory	Schwartz	1995	Tacoma, WA	65+	All	All	None	24-hr avg	Log-linear	0.007472	0.002638
All Respiratory	Schwartz	1995	Tacoma, WA	65+	All	All	PM_{10}	24-hr avg	Log-linear	0.007147	0.002565
All Respiratory	Thurston et al.	1994	Toronto, CAN	All	All	All	None	1-hr max	Linear	0.0528	0.0197
All Respiratory	Thurston et al.	1994	Toronto, CAN	All	All	All	$PM_{2.5}$	1-hr max	Linear	0.0404	0.0233
All Respiratory	Thurston et al.	1994	Toronto, CAN	All	All	All	PM_{10}	1-hr max	Linear	0.0388	0.0241
Asthma	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	Log-linear	0.003143	0.000679
Asthma	Burnett et al.	1999	Toronto, CAN	All	All	All	CO, $PM_{10-2.5}$	24-hr avg	Log-linear	0.002497	0.000718
Asthma	Sheppard et al.	1999	Seattle, WA	<65	All	All	None	8-hr avg	Log-linear	0.002913	0.001079
Asthma	Thurston et al.	1994	Toronto, CAN	All	All	All	None	1-hr max	Linear	0.0346	0.0124
Asthma	Thurston et al.	1994	Toronto, CAN	All	All	All	$PM_{2.5}$	1-hr max	Linear	0.0265	0.0142
Asthma	Thurston et al.	1994	Toronto, CAN	All	All	All	PM_{10}	1-hr max	Linear	0.029	0.0146
Chronic Lung Disease	Moolgavkar et al.	1997	Minneapolis, MN	65+	All	All	CO, PM_{10}	24-hr avg	Log-linear	0.002743	0.001699
Chronic Lung Disease (less Asthma)	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	Log-linear	0.003608	0.000853
Chronic Lung Disease (less Asthma)	Burnett et al.	1999	Toronto, CAN	All	All	All	CO, $PM_{10-2.5}$	24-hr avg	Log-linear	0.003027	0.001105

Appendix G. Ozone C-R Functions

Endpoint Name	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Functional Form	Beta	Std Error
Chronic Lung Disease (less Asthma)	Schwartz	1994	Detroit, MI	65+	All	All	PM_{10}	24-hr avg	Log-linear	0.00549	0.00205
Pneumonia	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	Log-linear	0.002218	0.000517
Pneumonia	Burnett et al.	1999	Toronto, CAN	All	All	All	NO_2, $PM_{2.5}$	24-hr avg	Log-linear	0.001977	0.000520
Pneumonia	Moolgavkar et al.	1997	Minneapolis, MN	65+	All	All	NO_2, PM_{10}, SO_2	24-hr avg	Log-linear	0.003696	0.001030
Pneumonia	Schwartz	1994	Detroit, MI	65+	All	All	PM_{10}	24-hr avg	Log-linear	0.00521	0.0013
Pneumonia	Schwartz	1994	Minneapolis, MN	65+	All	All	None	24-hr avg	Log-linear	0.003479	0.001616
Pneumonia	Schwartz	1994	Minneapolis, MN	65+	All	All	PM_{10}	24-hr avg	Log-linear	0.003977	0.001865
All Cardiovascular	Burnett et al.	1997	Toronto, CAN	All	All	All	None	8-hr avg	Log-linear	0.006208	0.001612
All Cardiovascular	Burnett et al.	1997	Toronto, CAN	All	All	All	$PM_{2.5}$	8-hr avg	Log-linear	0.005231	0.001503
All Cardiovascular	Burnett et al.	1997	Toronto, CAN	All	All	All	$PM_{10\text{-}2.5}$	8-hr avg	Log-linear	0.005313	0.001420
All Cardiovascular	Burnett et al.	1997	Toronto, CAN	All	All	All	NO_2, $PM_{2.5}$, SO_2	8-hr avg	Log-linear	0.005639	0.001512
Dysrhythmia	Burnett et al.	1999	Toronto, CAN	All	All	All	None	24-hr avg	Log-linear	0.001769	0.001035
Dysrhythmia	Burnett et al.	1999	Toronto, CAN	All	All	All	CO, $PM_{2.5}$	24-hr avg	Log-linear	0.001685	0.001034

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.

G .3 Hospital Admissions

Exhibit G-3 summarizes the C-R functions used to estimate the relationship between ozone and hospital admissions. Detailed summaries of each of the studies used to generate the functions are described below, along with the parameters used in each of the functions.

G .3.1 Hospital Admissions for All Respiratory (Burnett et al., 1997, Toronto)

Burnett et al. (1997) examined the relationship between air pollution and hospital admissions (ICD codes 464-466, 480-486, 490-494, 496) for individuals of all ages in Toronto, Canada during the summers of 1992-1994. In a Poisson regression model, all respiratory admissions were linked to coefficient of haze (COH) and ozone; other PM measures were less strongly linked. In two pollutant models with COH, they found that CO, NO_2, and SO_2 were not significant, while ozone remained significant. In multipollutant models with COH, ozone, NO_2, and SO_2, both ozone and COH remained significant. None of the other PM measures (PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$) were significant in four-pollutant models. The ozone C-R functions are based on the results from the single pollutant model and multipollutant models with PM co-pollutants.

Single Pollutant Model

In a single pollutant model with adjustment for temperature and dew point, the coefficient and standard error are based on the relative risk (1.064) and t-statistic (5.13) reported for an 11.5 ppb increase in 12-hour average ozone (1997, Table 2, p. 617).

Functional Form: Log-linear
Coefficient: 0.005394
Standard Error: 0.001052
Incidence Rate: region-specific daily hospital admission rate for all respiratory admissions per person (ICD 464, 466, 480-487, 490-496)
Population: population of all ages

Multipollutant Model (ozone and $PM_{2.5}$)

In a model with $PM_{2.5}$, the coefficient and standard error are based on the relative risk (1.059) and t-statistic (4.56) reported for an 11.5 ppb increase in 12-hour average ozone (1997, Table 4, p. 618).

Functional Form: Log-linear
Coefficient: 0.004985
Standard Error: 0.001093
Incidence Rate: region-specific daily hospital admission rate for all respiratory admissions per person (ICD 464, 466, 480-487, 490-496)
Population: population of all ages

Multipollutant Model (ozone and $PM_{10-2.5}$)

In a model with $PM_{10-2.5}$, the coefficient and standard error are based on the relative risk (1.062) and t-statistic (4.89) reported for an 11.5 ppb increase in 12-hour average ozone (1997, Table 4, p. 618).

Functional Form: Log-linear
Coefficient: 0.005231
Standard Error: 0.001070
Incidence Rate: region-specific daily hospital admission rate for all respiratory admissions per person (ICD 464, 466, 480-487, 490-496)
Population: population of all ages

Multipollutant Model (ozone, NO_2, $PM_{2.5}$, and SO_2)

In a four-pollutant model with NO_2, $PM_{2.5}$, and SO_2, the coefficient and standard error are based on the relative risk (1.059) and t-statistic (4.66) reported for an 11.5 ppb increase in 12-hour average ozone (1997, Table 6, p. 618).

Functional Form: Log-linear
Coefficient: 0.004985
Standard Error: 0.001070
Incidence Rate: region-specific daily hospital admission rate for all respiratory admissions per person (ICD 464, 466, 480-487, 490-496)
Population: population of all ages

G .3.2 Hospital Admissions for All Respiratory (Burnett et al., 2001, Toronto)

Burnett et al. (2001) studied the association between air pollution and acute respiratory hospital admissions (ICD codes 493, 466, 464.4, 480-486) in Toronto from 1980-1994, among children less than 2 years of age. They collected hourly concentrations of the gaseous pollutants, CO, NO_2, SO_2, and ozone. Daily measures of particulate matter were estimated for the May to August period of 1992-1994 using TSP, sulfates, and coefficient of haze data. The authors report a positive association between ozone in the May through August months and respiratory hospital admissions, for several single days after elevated ozone levels.

The strongest association was found using a five-day moving average of ozone. No association was found in the September through April months. In co-pollutant models with a particulate matter or another gaseous pollutant, the ozone effect was only slightly diminished. The effects for PM and gaseous pollutants were generally significant in single pollutant models but diminished in co-pollutant models with ozone, with the exception of CO. The C-R functions for ozone are based on a single pollutant and two co-pollutant models, using the five-day moving average of one-hour max ozone.

Single Pollutant Model[144]

The single pollutant coefficient and standard error are based on a percent increase (34.8) and 95% confidence interval of the percent increase (19.3 percent, 52.3 percent) for a 45.2 ppb change in the five-day moving average of one-hour max ozone (Burnett et al., 2001, Table 2 and p. 448).

[144] The authors present seven single-pollutant models: the first six of these use single lags of 0 days, 1 day, ..., up to 5 days. The seventh model uses a 5-day moving average of 0-day, 1-day, 2-day, 3-day and 4-day lagged 1-hour maximum ozone concentrations. The authors describe the 5-day moving average model as an attempt to "more fully characterize this pattern of temporally distributed effects" (p. 448). It shows a percentage increase of 34.8%, substantially larger than the percentage increase from any of the single lag models. This suggests that the 5-day moving average is indeed capturing some of the effect of each of the days that were shown to have an effect, individually, in the single lag models.

Functional Form: Log-linear
Coefficient: 0.006607
Standard Error: 0.0001378
Incidence Rate: region-specific daily hospital admission rate for all respiratory admissions per person less than 2 years of age (ICD codes 464, 466, 480-487, 493)
Population: population less than 2 years of age

Multipollutant Model (ozone and $PM_{2.5}$)

In a model with $PM_{2.5}$, the coefficient and standard error are based on the percent increase (33.0) and t-statistic (3.44) associated with a 45.2 ppb increase in the five-day moving average of one-hour max ozone (Burnett et al., 2001, Table 3).

Functional Form: Log-linear
Coefficient: 0.006309
Standard Error: 0.001834
Incidence Rate: region-specific daily hospital admission rate for all respiratory admissions per person less than 2 years of age (ICD codes 464, 466, 480-487, 493)
Population: population less than 2 years of age

Multipollutant Model (ozone and $PM_{10-2.5}$)

In a model with $PM_{10-2.5}$, the coefficient and standard error are based on the percent increase (29.4) and t-statistic (3.00) associated with a 45.2 ppb increase in the five-day moving average of one-hour max ozone (Burnett et al., 2001, Table 3).

Functional Form: Log-linear
Coefficient: 0.005702
Standard Error: 0.001901
Incidence Rate: region-specific daily hospital admission rate for all respiratory admissions per person less than 2 years of age (ICD codes 464, 466, 480-487, 493)
Population: population less than 2 years of age

G .3.3 Hospital Admissions for All Respiratory (Schwartz, 1995, New Haven)

Schwartz (1995) examined the relationship between air pollution and respiratory hospital admissions (ICD codes 460-519) for individuals 65 and older in New Haven, Connecticut, from January 1988 to December 1990. In single-pollutant models, PM_{10} and SO_2 were significant, while ozone was marginally significant. In a co-pollutant model with ozone and PM_{10}, both pollutants were significant. PM_{10} remained significant in a model with SO_2, while ozone was marginally significant when adjusted for SO_2. SO_2 was significant in a co-pollutant model with PM_{10} but not with ozone. The ozone C-R functions are based on results from the single pollutant model and co-pollutant model with PM_{10}.

Single Pollutant Model

In a single-pollutant model, the coefficient and standard error are calculated from the relative risk (1.03) and 95% confidence interval (1.02-1.05) for a 50 μg/m^3 increase in average daily ozone levels (Schwartz, 1995, Table 3, p. 534).[145]

Functional Form: Log-linear
Coefficient: 0.002284
Standard Error: 0.001323
Incidence Rate: region-specific daily hospital admission rate for respiratory admissions per person 65+ (ICD codes 460-519)
Population: population of ages 65 and older

Multipollutant Model (ozone and PM_{10})

In a model with PM_{10}, the coefficient and standard error are estimated from the relative risk (1.07) and 95% confidence interval (1.00-1.15) for a 50 μg/m^3 increase in average daily ozone levels (Schwartz, 1995, Table 3, p. 534).[146]

Functional Form: Log-linear
Coefficient: 0.002652
Standard Error: 0.001398
Incidence Rate: region-specific daily hospital admission rate for respiratory admissions per person 65+ (ICD codes 460-519)
Population: population of ages 65 and older

G .3.4 Hospital Admissions for All Respiratory (Schwartz, 1995, Tacoma)

Schwartz (1995) examined the relationship between air pollution and hospital admissions for individuals 65 and older in Tacoma, Washington, from January 1988 to December 1990. In single-pollutant models, PM_{10}, ozone, and SO_2 were all significant. Ozone remained significant in separate co-pollutant models with PM_{10} and SO_2. PM_{10} remained significant in a co-pollutant model with SO_2, but not in a co-pollutant model with ozone. SO_2 was not significant in either of the co-pollutant models. The ozone C-R functions are based on results from the single pollutant model and co-pollutant model with PM_{10}.

[145] To calculate the coefficient, a conversion of 1.96 μg/m^3 per ppb is used, based on a density of ozone of 1.96 grams per liter (at 25 degrees Celsius). Since there are 1000 liters in a cubic meter and a million μg in a gram, this density means that there are 1.96 billion μg of ozone in a cubic meter of ozone. If a cubic meter has just one ppb of ozone, then this means that this particular cubic meter has 1.96 μg of ozone (i.e., one ppb = 1.96 μg/m^3).

[146] To calculate the coefficient, a conversion of 1.96 μg/m^3 per ppb is used, based on a density of ozone of 1.96 grams per liter (at 25 degrees Celsius). Since there are 1000 liters in a cubic meter and a million μg in a gram, this density means that there are 1.96 billion μg of ozone in a cubic meter of ozone. If a cubic meter has just one ppb of ozone, then this means that this particular cubic meter has 1.96 μg of ozone (i.e., one ppb = 1.96 μg/m^3).

Single Pollutant Model

In a single-pollutant model, the coefficient and standard error are calculated from the relative risk (1.21) and 95% confidence interval (1.06-1.38) for a 50 μg/m^3 increase in average daily ozone levels (Schwartz, 1995, Table 6, p. 535)[147]

Functional Form: Log-linear
Coefficient: 0.007472
Standard Error: 0.002638
Incidence Rate: region-specific daily hospital admission rate for respiratory admissions per person 65+ (ICD codes 460-519)
Population: population of ages 65 and older

Multipollutant Model (ozone and PM_{10})

In a model with PM_{10}, the coefficient and standard error are estimated from the relative risk (1.20) and 95% confidence interval (1.06-1.37) for a 50 μg/m^3 increase in average daily ozone levels (Schwartz, 1995, Table 6, p. 535).[148]

Functional Form: Log-linear
Coefficient: 0.007147
Standard Error: 0.002565
Incidence Rate: region-specific daily hospital admission rate for respiratory admissions per person 65+ (ICD codes 460-519)
Population: population of ages 65 and older

G .3.5 Hospital Admissions for All Respiratory (Thurston et al., 1994, Toronto)

Thurston et al. (1994) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada, for six weeks in July and August 1986-1988. In single-pollutant linear regression models, ozone and various measures of PM were linked to all respiratory admissions (ICD codes 466, 480-482, 485, 490-493). In two-pollutant models, ozone was still significant, but measures of PM were often not significant; only H^+ was significant. The C-R functions for ozone are based on results from single and multipollutant models.

Single Pollutant Model

In a single pollutant model, the ozone coefficient (0.0528) and standard error (0.0197) are reported in Table 3 (Thurston et al., 1994, p. 281) for a unit ppb increase in one-hour maximum ozone levels.

[147] To calculate the coefficient, a conversion of 1.96 μg/m^3 per ppb is used, based on a density of ozone of 1.96 grams per liter (at 25 degrees Celsius). Since there are 1000 liters in a cubic meter and a million μg in a gram, this density means that there are 1.96 billion μg of ozone in a cubic meter of ozone. If a cubic meter has just one ppb of ozone, then this means that this particular cubic meter has 1.96 μg of ozone (i.e., one ppb = 1.96 μg/m^3).

[148] To calculate the coefficient, a conversion of 1.96 μg/m^3 per ppb is used, based on a density of ozone of 1.96 grams per liter (at 25 degrees Celsius). Since there are 1000 liters in a cubic meter and a million μg in a gram, this density means that there are 1.96 billion μg of ozone in a cubic meter of ozone. If a cubic meter has just one ppb of ozone, then this means that this particular cubic meter has 1.96 μg of ozone (i.e., one ppb = 1.96 μg/m^3).

Functional Form: Linear
Coefficient: 0.0528
Standard Error: 0.0197
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

Multipollutant Model (ozone and $PM_{2.5}$)

In a model with $PM_{2.5}$, the ozone coefficient (0.0404) and standard error (0.0233) are reported in Table 3 (Thurston et al., 1994, p. 281) for a unit ppb increase in one-hour maximum ozone levels.

Functional Form: Linear
Coefficient: 0.0404
Standard Error: 0.0233
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

Multipollutant Model (ozone and PM_{10})

In a model with PM_{10}, the ozone coefficient (0.0388) and standard error (0.0241) are reported in Table 3 (Thurston et al., 1994, p. 281) for a unit ppb increase in one-hour maximum ozone levels.

Functional Form: Linear
Coefficient: 0.0388
Standard Error: 0.0241
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

G .3.6 Hospital Admissions for Asthma (Burnett et al., 1999, Toronto)

Burnett et al. (1999) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada from 1980 to 1994. The authors examined single pollutant log-linear models for PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$, CO, NO_2, SO_2, and ozone and found all significantly associated with asthma except SO_2. They estimated multi-pollutant models, where pollutants for best fitting model were chosen using stepwise regression based on AIC criterion. Asthma admissions were linked to ozone, CO, and $PM_{10-2.5}$. The C-R functions for ozone are based on the results of a single pollutant model and three pollutant model (ozone, CO, $PM_{10-2.5}$).[149]

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (6.32) and t-statistic (4.63) reported in Table 3 (Burnett et al., 1999, p. 133) for a 19.5 ppb increase in three-day average ozone concentration.

[149] Burnett et al. (1999) reports results for co-pollutant models with ozone and various PM metrics as well, however, standard errors were not provided so these estimates were not used to derive C-R functions.

Functional Form: Log-linear
Coefficient: 0.003143
Standard Error: 0.000679
Incidence Rate: region-specific daily hospital admission rate for asthma per person (ICD code 493)
Population: population of all ages

Multipollutant Model (ozone, CO, and $PM_{10-2.5}$)

In a model with $PM_{10-2.5}$ and CO, the ozone coefficient and standard error are based on the percent increase (4.99) reported in Table 5 (Burnett et al., 1999, p. 135) and the t-statistic obtained from the authors (3.48)[150] for a 19.5 ppb increase in three-day average ozone concentration.

Functional Form: Log-linear
Coefficient: 0.002497
Standard Error: 0.000718
Incidence Rate: region-specific daily hospital admission rate for asthma per person (ICD code 493)
Population: population of all ages

G .3.7 Hospital Admissions for Asthma (Sheppard et al., 1999, Seattle)

Sheppard et al. (1999) studied the relationship between air pollution in Seattle and nonelderly (<65) hospital admissions for asthma from 1987 to 1994. They used air quality data for PM_{10}, $PM_{2.5}$, $PM_{10-2.5}$, SO_2, ozone, and CO in a Poisson regression model with control for time trends, seasonal variations, and temperature-related weather effects.[151] They found asthma hospital admissions associated with PM_{10}, $PM_{2.5}$, $PM_{10-2.5}$, CO, and ozone. They did not observe an association for SO_2. They found PM and CO to be jointly associated with asthma admissions. The best fitting co-pollutant models were found using ozone. However, ozone data was only available April through October, so they did not consider ozone further. For the remaining pollutants, the best fitting models included $PM_{2.5}$ and CO. Results for other co-pollutant models were not reported. The ozone C-R function is based on the results of a single pollutant model.

Single Pollutant Model

The single pollutant coefficient and standard error are calculated from the relative risk (1.06) and 95% confidence interval (1.02-1.11) associated with a 20 ppb increase in eight-hour average ozone (Sheppard et al., 1999, p. 27).

Functional Form: Log-linear
Coefficient: 0.002913
Standard Error: 0.001079
Incidence Rate: region-specific daily hospital admission rate for asthma per person <65 (ICD code 493)
Population: population of ages 65 and under

[150] Rick Burnett (co-author), personal communication.

[151] $PM_{2.5}$ levels were estimated from light scattering data.

G .3.8 Hospital Admissions for Asthma (Thurston et al., 1994, Toronto)

Thurston et al. (1994) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada, for six weeks in July and August 1986-1988. In single-pollutant linear regression models, ozone was strongly associated with asthma admissions (ICD code 493) and various measures of PM were marginally significant. In two-pollutant models, ozone remained significant, but measures of PM were often not significant. The C-R functions for ozone are based on results from single and multipollutant models.

Single Pollutant Model

In a single pollutant model, the ozone coefficient (0.0346) and standard error (0.0124) are reported in Table 4 (Thurston et al., 1994, p. 282) for a unit ppb increase in one-hour maximum ozone levels.

Functional Form: Linear
Coefficient: 0.0346
Standard Error: 0.0124
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

Multipollutant Model (ozone and $PM_{2.5}$)

In a model with $PM_{2.5}$, the ozone coefficient (0.0265) and standard error (0.0142) are reported in Table 4 (Thurston et al., 1994, p. 282) for a unit ppb increase in one-hour maximum ozone levels.

Functional Form: Linear
Coefficient: 0.0265
Standard Error: 0.0142
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

Multipollutant Model (ozone and PM_{10})

In a model with PM_{10}, the ozone coefficient (0.0290) and standard error (0.0146) are reported in Table 4 (Thurston et al., 1994, p. 282) for a unit ppb increase in one-hour maximum ozone levels.

Functional Form: Linear
Coefficient: 0.0290
Standard Error: 0.0146
Baseline Pop: baseline population in Toronto = 2,400,000 (U.S. EPA, 1997, Table D-7)
Population: population of all ages

G .3.9 Hospital Admissions for Chronic Lung Disease (Moolgavkar et al., 1997, Minneapolis)

Moolgavkar et al. (1997) examined the relationship between air pollution and hospital admissions (ICD codes 490-496) for individuals 65 and older in Minneapolis-St. Paul, Minnesota, from January 1986 to December 1991. In a Poisson regression, they found no significant effect for any of the pollutants (PM_{10}, ozone, or CO). The effect for ozone was marginally significant. The model with a 100 df smoother

was reported to be optimal (p. 368). The C-R function is based on the results from a three-pollutant model (ozone, CO, PM_{10}) using the 100 df smoother.

Multipollutant Model (ozone, CO, PM_{10})

In a model with CO and PM_{10}, the estimated coefficient and standard error are based on the percent increase (4.2) and 95% confidence interval of the percent increase (-1.0-9.4) associated with a change in daily average ozone levels of 15 ppb (Moolgavkar et al., 1997, Table 4 and p. 366).

Functional Form: Log-linear
Coefficient: 0.002743
Standard Error: 0.001699
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease per person 65+ (ICD codes 490-496)
Population: population of ages 65 and older

G .3.10 Hospital Admissions for Chronic Lung Disease (less Asthma) (Burnett et al., 1999, Toronto)

Burnett et al. (1999) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada from 1980 to 1994. The authors examined single pollutant log-linear models for PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$, CO, NO_2, SO_2, and ozone and found $PM_{10-2.5}$, PM_{10}, and ozone significantly associated with chronic lung disease (ICD codes 490-492, 496). They estimated multi-pollutant models, where pollutants for the best fitting model were chosen using stepwise regression based on AIC criterion. In a three pollutant model, admissions for chronic obstructive pulmonary disease (COPD) were linked to ozone and $PM_{10-2.5}$. A non-significant association was found with CO. The C-R functions for ozone are based on the results of a single pollutant model and three-pollutant model (ozone, CO, $PM_{10-2.5}$).[152]

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (7.29) and t-statistic (4.23) reported in Table 3 (Burnett et al., 1999, p. 133) for a 19.5 ppb increase in three-day average ozone concentration.

Functional Form: Log-linear
Coefficient: 0.003608
Standard Error: 0.000853
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease per person (ICD codes 490-492, 494-496)
Population: population of all ages

[152] Burnett et al. (1999) reports results for co-pollutant models with ozone and various PM metrics as well, however, standard errors were not provided so these estimates were not used to derive C-R functions.

Multipollutant Model (ozone, CO, and $PM_{10-2.5}$)

In a model with $PM_{10-2.5}$ and CO, the ozone coefficient and standard error are based on the percent increase (6.08) reported in Table 5 (Burnett et al., 1999, p. 135) and the t-statistic obtained from the authors (2.74)[153] for a 19.5 ppb increase in three-day average ozone concentration.

Functional Form: Log-linear
Coefficient: 0.003027
Standard Error: 0.001105
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease per person (ICD codes 490-492, 494-496)
Population: population of all ages

G .3.11 Hospital Admissions for Chronic Lung Disease (less Asthma) (Schwartz, 1994b, Detroit)

Schwartz (1994b) examined the relationship between air pollution and hospital admissions (ICD codes 491-492, 494-496) for individuals 65 and older in Detroit, Michigan, from January 1986 to December 1989. In a two-pollutant Poisson regression model, Schwartz found both PM_{10} and ozone significantly linked to pneumonia and COPD. The authors state that effect estimates were relatively unchanged compared to the unreported single pollutant models. No significant associations were found between either pollutant and asthma admissions. The C-R function for chronic lung disease incidence is based on the results of the "basic" co-pollutant model (ozone and PM_{10}) presented in Table 4 (p. 651).[154]

Multipollutant Model (ozone and PM_{10})

The coefficient and standard error for the "basic" model are reported in Table 4 (Schwartz, 1994b, p.651) for a one ppb change in daily average ozone.

Functional Form: Log-linear
Coefficient: 0.00549
Standard Error: 0.00205
Incidence Rate: region-specific daily hospital admission rate for chronic lung disease per person 65+ (ICD codes 490-492, 494-496)
Population: population of ages 65 and older

G .3.12 Hospital Admissions for Pneumonia (Burnett et al., 1999, Toronto)

Burnett et al. (1999) examined the relationship between air pollution and hospital admissions for individuals of all ages in Toronto, Canada from 1980 to 1994. The authors examined single pollutant log-linear models for PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$, CO, NO_2, SO_2, and ozone and found all significantly associated with pneumonia and other respiratory infections (ICD codes 464, 466, 480-487, 494). They estimated multipollutant models, where pollutants for the best fitting model were chosen using stepwise regression

[153] Rick Burnett (co-author), personal communication.

[154] Schwartz (1994b) also reports results using generalized additive models to fit time and temperature variables, however no standard error or confidence intervals were reported.

based on AIC criterion. Pneumonia and respiratory infection admissions were linked to ozone, NO_2, and $PM_{2.5}$. The C-R functions for ozone are based on the results of a single pollutant model and three-pollutant model (ozone, NO_2, $PM_{2.5}$).[155]

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (4.42) and t-statistic (4.29) reported in Table 3 (Burnett et al., 1999, p. 133) for a 19.5 ppb increase in two-day average ozone concentration.

Functional Form: Log-linear
Coefficient: 0.002218
Standard Error: 0.000517
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person (ICD codes 464, 466, 480-487)
Population: population of all ages

Multipollutant Model (ozone, NO_2, $PM_{2.5}$)

In a model with $PM_{2.5}$ and NO_2, the ozone coefficient and standard error are based on the percent increase (3.93) reported in Table 5 (Burnett et al., 1999, p. 135) and the t-statistic obtained from the authors (3.80)[156] for a 19.5 ppb increase in two-day average ozone concentration.

Functional Form: Log-linear
Coefficient: 0.001977
Standard Error: 0.000520
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person (ICD codes 464, 466, 480-487)
Population: population of all ages

G .3.13 Hospital Admissions for Pneumonia (Moolgavkar et al., 1997, Minneapolis)

Moolgavkar et al. (1997) examined the relationship between air pollution and pneumonia hospital admissions (ICD 480-487) for individuals 65 and older in Minneapolis-St. Paul, Minnesota, from January 1986 to December 1991. In a four pollutant Poisson model examining pneumonia admissions in Minneapolis, ozone was significant, while NO_2, SO_2, and PM_{10} were not significant. The model with a 130 df smoother was reported to be optimal (p. 368). The ozone C-R function is based on the results from the four-pollutant model with a 130 df smoother.

Multipollutant Model (ozone, NO_2, PM_{10},and SO_2)

In a model with NO_2, PM_{10},and SO_2, the estimated coefficient and standard error are based on the percent increase (5.7) and 95% confidence interval of the percent increase (2.5-8.9) associated with an increase in daily average ozone levels of 15 ppb (Moolgavkar et al., 1997, Table 4 and p. 366).

[155] Burnett et al. (1999) reports results for co-pollutant models with ozone and various PM metrics as well, however, standard errors were not provided so these estimates were not used to derive C-R functions.

[156] Rick Burnett (co-author), personal communication.

Functional Form: Log-linear
Coefficient: 0.003696
Standard Error: 0.00103
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older

G .3.14 Hospital Admissions for Pneumonia (Schwartz, 1994b, Detroit)

Schwartz (1994b) examined the relationship between air pollution and hospital admissions for individuals 65 and older in Detroit, Michigan, from January 1986 to December 1989. In a two-pollutant Poisson regression model, Schwartz found both PM_{10} and ozone significantly linked to pneumonia and COPD. The authors state that effect estimates were relatively unchanged compared to the unreported single pollutant models. No significant associations were found between either pollutant and asthma admissions. The PM_{10} C-R function for pneumonia incidence is based on results of the "basic" co-pollutant model (ozone and PM_{10}).[157]

Multipollutant Model (ozone and PM_{10})

The ozone C-R function for pneumonia incidence is based on the coefficient and standard error for the "basic" co-pollutant model presented in Table 4 (Schwartz, 1994b, p. 651).

Functional Form: Log-linear
Coefficient: 0.00521
Standard Error: 0.0013
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older

G .3.15 Hospital Admissions for Pneumonia (Schwartz, 1994a, Minneapolis)

Schwartz (1994a) examined the relationship between air pollution and hospital admissions for individuals 65 and older in Minneapolis-St. Paul, Minnesota, from January 1986 to December 1989. In single-pollutant Poisson regression models, both ozone and PM_{10} were significantly associated with pneumonia admissions. In a two-pollutant model, Schwartz found PM_{10} significantly related to pneumonia; ozone was weakly linked to pneumonia. The results were not sensitive to the methods used to control for seasonal patterns and weather. The ozone C-R functions are based on the results of the single pollutant model and the two-pollutant model (PM_{10} and ozone) with spline smoothing for temporal patterns and weather.

Single Pollutant Model

The single pollutant coefficient and standard error are based on the relative risk (1.19) and 95% confidence interval (1.02-1.40) for a 50 ppb increase in daily average ozone levels (Schwartz, 1994a, p. 369).

[157] Schwartz (1994b) also reports results using generalized additive models to fit time and temperature variables, however no standard error or confidence intervals were reported.

Functional Form: Log-linear
Coefficient: 0.003479
Standard Error: 0.001616
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older

Multipollutant Model (ozone and PM_{10})

In a model with PM_{10} and spline functions to adjust for time and weather, the coefficient and standard error are based on the relative risk (1.22) and 95% confidence interval (1.02, 1.47) for a 50 ppb increase in daily average ozone levels (Schwartz, 1994a, Table 4).

Functional Form: Log-linear
Coefficient: 0.003977
Standard Error: 0.001865
Incidence Rate: region-specific daily hospital admission rate for pneumonia per person 65+ (ICD codes 480-487)
Population: population of ages 65 and older

G .3.16 Hospital Admissions for All Cardiovascular (Burnett et al., 1997, Toronto)

Burnett et al. (1997) examined the relationship between air pollution and cardiac hospital admissions (ICD codes 410-414, 427, 428) for individuals of all ages in Toronto, Canada during the summers of 1992-1994. In a Poisson regression model, cardiac admissions were linked to coefficient of haze (COH) and ozone; other PM measures were less strongly linked. In two pollutant models, they found that CO, NO_2, and SO_2 were not significant, controlling for COH. They found that ozone was still significant, controlling for COH. In multi-pollutant models with COH, ozone, NO_2, and SO_2, both ozone and COH remained significant. None of the other PM measures (PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$) were significant in four-pollutant models. The ozone C-R functions are based on the results from the single pollutant model and multipollutant models with PM co-pollutants.

Single Pollutant Model

In a single pollutant model with adjustment for temperature and dew point, the ozone coefficient and standard error are based on the relative risk (1.074) and t-statistic (3.85) reported for an 11.5 ppb increase in the three-day average of 12-hour average ozone (Burnett et al., 1997, Table 2, p. 617).

Functional Form: Log-linear
Coefficient: 0.006208
Standard Error: 0.001612
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

Multipollutant Model (ozone and $PM_{2.5}$)

In a model with $PM_{2.5}$, the ozone coefficient and standard error are based on the relative risk (1.062) and t-statistic (3.48) reported for an 11.5 ppb increase in the three-day average of 12-hour average ozone (Burnett et al., 1997, Table 5, p. 618).

Functional Form: Log-linear
Coefficient: 0.005231
Standard Error: 0.001503
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

Multipollutant Model (ozone and $PM_{10-2.5}$)

In a model with $PM_{10-2.5}$, the ozone coefficient and standard error are based on the relative risk (1.063) and t-statistic (3.74) reported for an 11.5 ppb increase in the three-day average of 12-hour average ozone (Burnett et al., 1997, Table 5, p. 618).

Functional Form: Log-linear
Coefficient: 0.005313
Standard Error: 0.001421
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

Multipollutant Model (ozone, NO_2, $PM_{2.5}$, SO_2)

In a four-pollutant model with $PM_{2.5}$, NO_2, and SO_2, the ozone coefficient and standard error are based on the relative risk (1.067) and t-statistic (3.73) reported for an 11.5 ppb increase in the three-day average of 12-hour average ozone (Burnett et al., 1997, Table 6, p. 618).

Functional Form: Log-linear
Coefficient: 0.005639
Standard Error: 0.001512
Incidence Rate: region-specific daily hospital admission rate for all cardiovascular disease per person (ICD codes 410-414, 427, 428)
Population: population of all ages

G .3.17 Hospital Admissions for Dysrhythmia (Burnett et al., 1999, Toronto)

Burnett et al. (1999) examined the relationship between air pollution and hospital admissions (ICD 427) for individuals of all ages in Toronto, Canada from 1980 to 1994. The authors examined single pollutant log-linear models for PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$, CO, NO_2, SO_2, and ozone and found $PM_{2.5}$, PM_{10}, and CO significantly associated with admissions. They estimated multiple pollutant models, where pollutants for best fitting model were chosen using stepwise regression based on AIC criterion. The final model for dysrhythmia admissions included ozone, CO, and $PM_{2.5}$. CO was significantly associated with admissions, while ozone and $PM_{2.5}$ were marginally significant. The C-R functions for ozone are based on the results of a single pollutant model and three-pollutant model (ozone, CO, and $PM_{2.5}$).[158]

[158] Burnett et al. (1999) reports results for co-pollutant models with ozone and various PM metrics as well, however, standard errors were not provided so these estimates were not used to derive C-R functions.

Single Pollutant Model

In a single pollutant model, the coefficient and standard error are based on the percent increase (3.51) and t-statistic (1.71) reported in Table 3 (Burnett et al., 1999, p. 133) for a 19.5 ppb increase in three-day average ozone concentration.

Functional Form: Log-linear
Coefficient: 0.001769
Standard Error: 0.001035
Incidence Rate: region-specific daily hospital admission rate for dysrhythmia per person (ICD code 427)
Population: population of all ages

Multipollutant Model (ozone, CO, $PM_{2.5}$)

In a model with $PM_{2.5}$ and CO, the ozone coefficient and standard error are based on the percent increase (3.34) reported in Table 5 (Burnett et al., 1999, p. 135) and the t-statistic obtained from the authors (1.63)[159] for a 19.5 ppb increase in three-day average ozone concentration.

Functional Form: Log-linear
Coefficient: 0.001685
Standard Error: 0.001034
Incidence Rate: region-specific daily hospital admission rate for dysrhythmia per person (ICD code 427)
Population: population of all ages

[159] Rick Burnett (co-author), personal communication.

Appendix G. Ozone C-R Functions

Exhibit G-4. Concentration-Response (C-R) Functions for Ozone and Emergency Room Visits

Endpoint Name	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Functional Form	Beta	Std Error
Asthma	Cody et al.	1992	New Jersey (Northern)	All	All	All	SO_2	5-hr avg	Linear	0.0203	0.00717
Asthma	Jaffe et al.	2003	Ohio cities	5-34	All	All	None	8-hr max	Log-linear	0.002956	0.001486
Asthma	Norris et al.	1999	Seattle, WA	<18	All	All	None	8-hr avg	Log-linear	0.004305	0.003826
Asthma	Schwartz et al.	1993	Seattle, WA	<65	All	All	None	24-hr avg	Log-linear	-0.002031	0.002812
Asthma	Stieb et al.	1996	New Brunswick, CAN	All	All	All	None	1-hr max	Quadratic	0.00004	0.00002
Asthma	Stieb et al.	1996	New Brunswick, CAN	All	All	All	None	24-hr avg	Quadratic	0.0001	0.00004
Asthma	Weisel et al.	1995	New Jersey (Northern and Central)	All	All	All	None	5-hr avg	Linear	0.0443	0.00723

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.

G .4 Emergency Room Visits

Exhibit G-4 summarizes the C-R functions used to estimate the relationship between ozone and emergency room visits. Detailed summaries of each of the studies used to generate the functions are described below, along with the parameters used in each of the functions.

G .4.1 Emergency Room Visits for Asthma (Cody et al., 1992, Northern NJ)

Cody et al. (1992) examined the relationship between ER visits and air pollution for persons of all ages in central and northern New Jersey, from May to August in 1988-1989. In a two pollutant multiple linear regression model, ozone was linked to asthma visits, and no effect was seen for SO_2. They modeled PM_{10} in separate analysis because of limited (every sixth day) sampling. No significant effect was seen for PM_{10}. The C-R function for ozone is based on results of a co-pollutant model with SO_2 (Cody et al., 1992, Table 6, p. 191).

Multipollutant Model (ozone and SO_2)

The ozone coefficient and standard error are reported per 1 ppm increment of five-hour ozone levels, which are converted to a 1 ppb increment by dividing by 1,000 (Cody et al., 1992, Table 6, p. 191).

Functional Form: Linear
Coefficient: 0.0203
Standard Error: 0.00717
Baseline Population: baseline population of Northern New Jersey[160] = 4,436,976
Population: population of all ages

G .4.2 Emergency Room Visits for Asthma (Jaffe et al., 2003)

Jaffe et al. (2003) examined the relationship between ER visits and air pollution for persons ages 5-34 in Cleveland, Columbus, and Cincinnati, Ohio, from 1991 through 1996. In single-pollutant Poisson regression models, ozone and SO_2 were linked to asthma visits, and no significant effect was seen for NO_2 and PM_{10}.

Single Pollutant Model

The ozone coefficient and standard error are reported per 10 ppb increment of the maximum daily 8-hour average ozone level (Jaffe et al., 2003, Table 3). We used the results from the three cities combined. The relative risk is 1.03, with a 95 percent confidence interval of 1.00 to 1.06.

Functional Form: Log-linear
Coefficient: 0.002956
Standard Error: 0.001486
Incidence: asthma ER rate for ages 0-17, 18-24, and 25-34
Population: population of ages from 5 to 34

[160] The population estimate is based on the 1990 population for the eight counties containing hospitals or in the central core of the study. Cody et al. (1992, Figure 1) presented a map of the study area; the counties are: Bergen, Essex, Hudson, Middlesex, Morris, Passaic, Somerset, and Union.

G .4.3 Emergency Room Visits for Asthma (Norris et al., 1999)

Norris et al. (1999) examined the relation between air pollution in Seattle and childhood (<18) hospital admissions for asthma from 1995 to 1996. The authors used air quality data for PM_{10}, light scattering (used to estimate fine PM), CO, SO_2, NO_2, and ozone in a Poisson regression model with adjustments for day of the week, time trends, temperature, and dew point. They found significant associations between asthma ER visits and light scattering (converted to $PM_{2.5}$), PM_{10}, and CO. No association was found between ozone, NO_2, or SO_2 and asthma ER visits, although ozone had a significant amount of missing data. In multi-pollutant models with either PM metric (light scattering or PM_{10}) and NO_2 and SO_2, the PM coefficients remained significant while the gaseous pollutants were not associated with increased asthma ER visits. The C-R function for ozone is based on the result of a single pollutant model.

Single Pollutant Model

The coefficient and standard error are calculated from a relative risk of 1.02 (95% CI 0.98-1.05) for a 4.6 ppb increase in maximum eight-hour ozone levels (Norris et al., 1999, p. 491).

Functional Form: Log-linear
Coefficient: 0.004305
Standard Error: 0.003826
Incidence Rate: region-specific daily emergency room rate for asthma per person <18 (ICD code 493)
Population: population of ages under 18

G .4.4 Emergency Room Visits for Asthma (Schwartz et al., 1993, Seattle)

Schwartz et al. (1993) examined the relationship between air quality and emergency room visits for asthma (ICD codes 493, 493.01, 493.10, 493.90, 493.91) in persons under 65 and 65 and over, living in Seattle from September 1989 to September 1990. Using single-pollutant models they found daily levels of PM_{10} linked to ER visits in individuals ages under 65, and they found no effect in individuals ages 65 and over. They did not find a significant effect for SO_2 and ozone in either age group. The C-R function is based on the results of the single pollutant model for ozone.

Single Pollutant Model

The ozone coefficient and standard error are based on the relative risk (0.97) and 95% confidence interval (0.89-1.05) for a 15 ppb increase in daily ozone levels (Schwartz et al., 1993, p. 829).

Functional Form: Log-linear
Coefficient: -0.002031
Standard Error: 0.002812
Incidence Rate: region-specific daily emergency room rate for asthma per person <65 (ICD code 493)
Population: population of ages under 65

G .4.5 Emergency Room Visits for Asthma (Stieb et al., 1996, New Brunswick)

Stieb et al. (1996) examined the relationship between ER visits and air pollution for persons of all ages in St. John, New Brunswick, Canada, from May through September in 1984-1992. Ozone was

significantly linked to ER visits, especially when ozone levels exceeded 75 ppb. The authors reported results from a linear model, quadratic model, and linear-quadratic model using daily average and 1-hour maximum ozone. In the linear model, ozone was borderline significant. In the quadratic and linear-quadratic models, ozone was highly significant. This is consistent with the author's conclusion that "only ozone appeared to have a nonlinear relationship with visit rates" (p. 1356) and that "quadratic, linear-quadratic, and indicator models consistently fit the data better than the linear model ..." (p. 1358). The linear term in the linear-quadratic model is negative, implying that at low ozone levels, increases in ozone are associated with decreases in risk. Since this does not seem biologically plausible, the ozone C-R functions described here are based on the results of the quadratic regression models presented in Table 2 (Stieb et al., 1996, p. 1356).

Single Pollutant Model (one-hour max ozone)

The coefficient and standard error of the quadratic model are reported in Table 2 (Stieb et al., 1996, p. 1356) for a 1 ppb increase in 1-hour daily maximum ozone levels. The C-R function to estimate avoided emergency visits derived from a quadratic regression model is shown below:

$$\Delta\ Asthma\ ERVisits = \frac{\beta}{BasePop} \cdot [(O_{3,baseline})^2 - (O_{3,control})^2] \cdot pop,$$

Functional Form: Quadratic
Coefficient: 0.00004
Standard Error: 0.00002
Baseline Population: baseline population of St. John, New Brunswick (Stieb et al., 1996, p. 1354) = 125,000
Population: population of all ages

Single Pollutant Model (daily average ozone)

The coefficient and standard error of the quadratic model are reported in Table 2 (p. 1356) for a 1 ppb increase in daily average ozone levels. The C-R function to estimate avoided emergency visits derived from a quadratic regression model is shown below:

$$\Delta\ Asthma\ ERVisits = \frac{\beta}{BasePop} \cdot [(O_{3,baseline})^2 - (O_{3,control})^2] \cdot pop,$$

Functional Form: Quadratic
Coefficient: 0.0001
Standard Error: 0.00004
Baseline Population: baseline population of St. John, New Brunswick (Stieb et al., 1996, p. 1354) = 125,000
Population: population of all ages

G .4.6 Emergency Room Visits for Asthma (Weisel et al., 1995, Northern NJ)

Weisel et al. (1995) examined the relationship between ER visits and air pollution for persons of all ages in central and northern New Jersey, from May to August in 1986-1990. A significant relationship was reported for ozone. The C-R function is based on the results of the single pollutant models reported by Weisel et al. (1995, Table 2).

Single Pollutant Model

The coefficient (β) used in the C-R function is a weighted average of the coefficients in Weisel et al. (1995, Table 2) using the inverse of the variance as the weight:

$$\beta = \left(\frac{\sum_{i=1986}^{1990} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\sum_{i=1986}^{1990} \frac{1}{\sigma_{\beta_i}^2}} \right) = 0.0443 .$$

The standard error of the coefficient (σ_β) is calculated as follows, assuming that the estimated year-specific coefficients are independent:

$$\sigma_\beta^2 = \text{var}\left(\frac{\sum_{i=1986}^{1990} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\sum_{i=1986}^{1990} \frac{1}{\sigma_{\beta_i}^2}} \right) = \left(\frac{\sum_{i=1986}^{1990} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\gamma} \right) = \sum_{i=1986}^{1990} \text{var}\left(\frac{\beta_i}{\sigma_{\beta_i}^2 \cdot \gamma} \right) .$$

This eventually reduces down to:

$$\sigma_\beta^2 = \frac{1}{\gamma} \Rightarrow \sigma_\beta = \sqrt{\frac{1}{\gamma}} = 0.00723 .$$

Functional Form: Linear
Coefficient: 0.0443
Standard Error: 0.00723
Baseline Population: baseline population of Northern New Jersey[161] = 4,436,976
Population: population of all ages

[161] The population estimate is based on the 1990 population for the eight counties containing hospitals or in the central core of the study. Cody et al. (1992, Figure 1) presented a map of the study area; the counties are: Bergen, Essex, Hudson, Middlesex, Morris, Passaic, Somerset, and Union.

Appendix G. Ozone C-R Functions

Exhibit G-5. Concentration-Response (C-R) Functions for Ozone and Acute Effects

Endpoint Name	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Functional Form	Beta	Std Error
Any of 19 Respiratory Symptoms	Krupnick	1990	Los Angeles, CA	18-64	All	All	COH	1-hr max	Linear	0.000137	0.000070
Minor Restricted Activity Days	Ostro and Rothschild	1989	nationwide	18-64	All	All	$PM_{2.5}$	24-hr avg	Log-linear	0.0022	0.000658
School Loss Days, All Cause	Chen et al.	2000	Washoe Co, NV	6-11	All	All	CO, PM_{10}	1-hr max	Linear	0.013247	0.004985
School Loss Days, All Cause	Gilliland et al.	2001	Southern California	9-10	All	All	None	8-hr avg	Log-linear	0.00755	0.004527
School Loss Days, Illness-Related	Gilliland et al.	2001	Southern California	9-10	All	All	None	8-hr avg	Log-linear	0.024398	0.008138
School Loss Days, Respiratory-Related	Gilliland et al.	2001	Southern California	9-10	All	All	None	8-hr avg	Log-linear	0.030188	0.014436
Worker Productivity	Crocker and Horst	1981	nationwide	18-64	All	All	None	24-hr avg	Linear	0.14	–

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.

G .5 Acute Morbidity

Exhibit G-5 summarizes the C-R functions used to estimate the relationship between ozone and acute morbidity. Detailed summaries of each of the studies used to generate the functions are described below, along with the parameters used in each of the functions.

G .5.1 Any of 19 Respiratory Symptoms: Krupnick (1990)

Krupnick et al. (1990) estimated the impact of air pollution on the incidence of any of 19 respiratory symptoms or conditions in 570 adults and 756 children living in three communities in Los Angeles, California from September 1978 to March 1979. Krupnick et al. (1990) listed 13 specific "symptoms or conditions": head cold, chest cold, sinus trouble, croup, cough with phlegm, sore throat, asthma, hay fever, doctor-diagnosed ear infection, flu, pneumonia, bronchitis, and bronchiolitis. The other six symptoms or conditions are not specified.

In their analysis, they included coefficient of haze (COH, a measure of particulate matter concentrations), ozone, NO_2, and SO_2, and they used a logistic regression model that takes into account whether a respondent was well or not the previous day. A key difference between this and the usual logistic model, is that the model they used includes a lagged value of the dependent variable. In single-pollutant models, daily ozone, COH, and SO_2 were significantly related to respiratory symptoms in adults. Controlling for other pollutants, they found that ozone was still significant. The results were more variable for COH and SO_2, perhaps due to collinearity. NO_2 had no significant effect. No effect was seen in children for any pollutant. The results from the two-pollutant model with COH and ozone are used to develop a C-R function.

Multipollutant Model (ozone and coefficient of haze)

The C-R function used to estimate the change in ARD2 associated with a change in daily one-hour maximum ozone[162] is based on Krupnick et al. (1990, p. 12):[163]

$$\Delta ARD2 \cong \beta^* \cdot \Delta O_3 \cdot pop \,,$$

Functional Form: Linear
Coefficient: first derivative of the stationary probability = 0.000137
Standard Error: 0.0000697
Population: population of ages 18-64 years[164]

The logistic regression model used by Krupnick et al. (1990) takes into account whether a respondent was well or not the previous day. Following Krupnick et al. (p. 12), the probability that one is sick is on a given day is:

[162]Krupnick et al. (1990) used parts per hundred million (pphm) to measure ozone; the coefficient used here is based on ppb.

[163]Krupnick and Kopp (1988, p. 2-24) and ESEERCO (1994, p. V-32) used the same C-R functional form as that used here.

[164]The coefficient estimates are based on the sample of "adults," and assumes that individuals 18 and older were considered adult. According to Krupnick et al. (1990, Table 1), about 0.6 percent of the study sample was over the age of 60. This is a relatively small fraction, so it is further assumed that the results do not apply to individuals 65 years of age and older.

$$probability(ARD2) = \frac{p_0}{1 - p_1 + p_0}$$

$$p_i = probability(ARD2 \mid sickness\ or\ not_{t-1}) = \frac{1}{1 - e^{\beta_0 + \beta_1 \cdot ARD2_{t-1} + X \cdot \beta}}, for\ i = 0,1.$$

where:

X	=	the matrix of explanatory variables
p_0	=	the probability of sickness on day t, given wellness on day t-1, and
p_1	=	the probability of sickness on day t, given sickness on day t-1.

In other words, the transition probabilities are estimated using a logistic function; the key difference between this and the usual logistic model, is that the model includes a lagged value of the dependent variable.

To calculate the impact of ozone (or other pollutants) on the probability of ARD2, it is possible, in principle, to estimate ARD2 before the change in ozone and after the change:

$$\Delta ARD2 = ARD2_{after} - ARD2_{before}.$$

However the full suite of coefficient estimates are not available.[165] Rather than use the full suite of coefficient values, the impact of ozone on the probability of ARD2 may be approximated by the derivative of ARD2 with respect to ozone:[166]

$$\frac{\partial probability(ARD2)}{\partial O_3} = \frac{p_0 \cdot (1 - p_1) \cdot \beta \left[p_1 + (1 - p_0) \right]}{(1 - p_1 + p_0)^2} = \beta^*,$$

where β is the reported logistic regression coefficient for ozone. The change in the incidence of ARD2 associated with a given change in ozone is then estimated by:

$$\frac{\partial ARD2}{\partial O_3} \cong \frac{\Delta ARD2}{\Delta O_3}$$

$$\Rightarrow \frac{\Delta ARD2}{\Delta O_3} \cong \beta^*$$

[165]The model without NO_2 (Krupnick et al., 1990, Table V equation 3) was used in this analysis, but the full suite of coefficient estimates for this model were not reported. Krupnick et al. (Table IV) reported all of the estimated coefficients for a model of children and for a model of adults when four pollutants were included (ozone, COH, SO_2, and NO_2). However, because of high collinearity between NO_2 and COH, NO_2 was dropped from some of the reported analyses (Krupnick et al., p. 10), and the resulting coefficient estimates changed substantially (see Krupnick et al., Table V). Both the ozone and COH coefficients dropped by about a factor of two or more.

[166]The derivative result is reported by Krupnick et al. (1990, p. 12).

$$\Rightarrow \Delta ARD2 \cong \beta^* \cdot \Delta O_3 \ .$$

This analysis uses transition probabilities obtained from Krupnick et al. as reported by ESEERCO (1994, p. V-32) for the adult population: $p_1 = 0.7775$ and $p_0 = 0.0468$. This implies:

$$\beta^* = \frac{0.0468 \cdot (1 - 0.7775) \cdot 0.00055 \cdot [0.7775 + (1 - 0.0468)]}{(1 - 0.7775 + 0.0468)^2} = 0.000137 \ .$$

The *standard error* for the coefficient is derived using the reported standard error of the logistic regression coefficient in Krupnick et al. (1990, Table V):

$$\beta_{high} = 0.00055 + (1.96 \cdot 0.00027) = 0.00108$$

$$\Rightarrow \beta^*_{high} = \frac{0.0468 \cdot (1 - 0.7775) \cdot 0.00108 \cdot [0.7775 + (1 - 0.0468)]}{(1 - 0.7775 + 0.0468)^2} = 0.000268$$

$$\sigma_{\beta, high} = \frac{\beta_{high} - \beta}{1.96} = \frac{(0.000268 - 0.000137)}{1.96} = 0.0000668$$

$$\beta_{low} = 0.00055 - (1.96 \cdot 0.00027) = 0.0000208$$

$$\Rightarrow \beta^*_{low} = \frac{0.0468 \cdot (1 - 0.7775) \cdot 0.0000208 \cdot [0.7775 + (1 - 0.0468)]}{(1 - 0.7775 + 0.0468)^2} = 5.17 \cdot 10^{-6}$$

$$\Rightarrow \ \sigma_{\beta, low} = \frac{\beta - \beta_{low}}{1.96} = \frac{\left(0.000137 + 5.17 \cdot 10^{-6}\right)}{1.96} = 0.0000725$$

$$\sigma_\beta = \frac{\sigma_{\beta, high} + \sigma_{\beta, low}}{2} = 0.0000697 \ .$$

G .5.2 Minor Restricted Activity Days: Ostro and Rothschild (1989)

Ostro and Rothschild (1989) estimated the impact of $PM_{2.5}$ and ozone on the incidence of minor restricted activity days (MRADs) and respiratory-related restricted activity days (RRADs) in a national sample of the adult working population, ages 18 to 65, living in metropolitan areas.[167] The annual national

[167] The study population is based on the Health Interview Survey (HIS), conducted by the National Center for Health Statistics. In publications from this ongoing survey, non-elderly adult populations are generally reported as ages 18-64. From the study, it is not clear if the age range stops at 65 or includes 65 year olds. We apply the C-R function to individuals ages 18-64 for

survey results used in this analysis were conducted in 1976-1981. Controlling for $PM_{2.5}$, two-week average ozone had a highly variable association with RRADs and MRADs. Controlling for ozone, two-week average $PM_{2.5}$ was significantly linked to both health endpoints in most years. The C-R function for ozone is based on the co-pollutant model with $PM_{2.5}$.

The study is based on a "convenience" sample of non-elderly individuals. Applying the C-R function to this age group is likely a slight underestimate, as it seems likely that elderly are at least as susceptible to ozone as individuals under 65. A number of studies have found that hospital admissions for the elderly are related to ozone exposures (e.g., Schwartz, 1994b; Schwartz, 1995).

Multipollutant Model (ozone and $PM_{2.5}$)

The coefficient and standard error used in the C-R function are based on a weighted average of the coefficients in Ostro and Rothschild (1989, Table 4). The derivation of these estimates is described below.

Functional Form: Log-linear
Coefficient: 0.00220
Standard Error: 0.000658
Incidence Rate: daily incidence rate for minor restricted activity days (MRAD) = 0.02137 (Ostro and Rothschild, 1989, p. 243)
Population: adult population ages 18 to 64

The coefficient used in the C-R function is a weighted average of the coefficients in Ostro and Rothschild (1989, Table 4) using the inverse of the variance as the weight:[168]

$$\beta = \left(\frac{\sum_{i=1976}^{1981} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\sum_{i=1976}^{1981} \frac{1}{\sigma_{\beta_i}^2}} \right) = 0.00220 .$$

The standard error of the coefficient is calculated as follows, assuming that the estimated year-specific coefficients are independent:

$$\sigma_\beta^2 = \text{var}\left(\frac{\sum_{i=1976}^{1981} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\sum_{i=1976}^{1981} \frac{1}{\sigma_{\beta_i}^2}} \right) = \left(\frac{\sum_{i=1976}^{1981} \frac{\beta_i}{\sigma_{\beta_i}^2}}{\gamma} \right) = \sum_{i=1976}^{1981} \text{var}\left(\frac{\beta_i}{\sigma_{\beta_i}^2 \cdot \gamma} \right) .$$

This reduces down to:

consistency with other studies estimating impacts to non-elderly adult populations.

[168] The calculation of the MRAD coefficient and its standard error is exactly analogous to the calculation done for the work-loss days coefficient based on Ostro (1987).

$$\sigma_\beta^2 = \frac{1}{\gamma} \Rightarrow \sigma_\beta = \sqrt{\frac{1}{\gamma}} = 0.000658 .$$

G .5.3 School Loss Days, All Cause (Chen et al., 2000)

Chen et al. (2000) studied the association between air pollution and elementary school absenteeism (grades 1-6)[169] in Washoe County, Nevada. Daily absence data were available for all elementary schools in the Washoe Country School District. The authors regressed daily total absence rate on the three air pollutants, meteorological variables, and indicators for day of the week, month, and holidays. They reported statistically significant associations between both ozone and CO and daily total absence rate for grades one through six. PM_{10} was negatively associated with absence rate, after adjustment for ozone, CO, and meteorological and temporal variables. The C-R function for ozone is based on the results from a multiple linear regression model with CO, ozone, and PM_{10}.

Multipollutant Model (ozone, CO, and PM_{10})

The coefficient and standard error are presented in Table 3 (Chen et al., 2000, p. 1008) for a unit ppm increase in the two-week average of daily one-hour maximum ozone concentration. This is converted to unit ppb increase by dividing by 1,000.

The reported coefficient represents an *absolute* increase in absenteeism rate for a unit increase in ozone. If we apply this study to other locations, we assume that the same absolute increase will occur for a unit increase in ozone, regardless of the baseline rate. If the study location has a particularly high baseline rate, we may be overestimating decreases in absenteeism nationally, and vice-versa. As an example, consider if the baseline absenteeism rate were 10% in the study and 5% nationally. An absolute increase in absence rate of 2% associated with a given increase in ozone reflects a relative increase in absence rate of 20% for the study population. However, in the national estimate, we would assume the same absolute increase of 2%, but this would reflect a relative increase in the absenteeism rate of 40%.

An alternative approach is to estimate apply the *relative* increase in absenteeism rate in the C-R function by adjusting the results by the ratio of the national absenteeism rate to the study-specific rate. As a result, the percent increase in absenteeism rate associated with an increase in ozone is extrapolated nationally rather than the absolute increase in absenteeism rate. The incidence derivation section above describes the data used to estimate national and study-specific absence rates.

In addition to this scaling factor, there are two other scaling factors which are applied to the function. A scaling factor of 0.01 is used to convert the beta from a percentage (x 100) per unit increase of ozone to a proportion per unit increase of ozone. As a result it can be applied directly to the national population of school children ages 6 through 11 to estimate the number of absences avoided.

The final scaling factor adjusts for the number of school days in the ozone season. In the modeling program, the function is applied to every day in the ozone season (May 1 - September 30), however, in reality, school absences will be avoided only on school days. We assume that children are in school during weekdays for all of May, two weeks in June, one week in August, and all of September.

[169] Assuming that most children start kindergarten at age 5, the corresponding ages for grades 1 through 6 would be 6 through 11.

This corresponds to approximately 2.75 months out of the 5 month season, resulting in an estimate of 39.3% of days (2.75/5*5/7). The C-R function parameters are shown below.

Functional Form: Linear
Coefficient: 0.013247
Standard Error: 0.004985
Population: population of children ages 6-11
Scaling Factor 1: Ratio of national school absence rate to study-specific school absence rate[170] = 1.081
Scaling Factor 2: Convert beta in percentage terms to a proportion = 0.01
Scaling Factor 3: Proportion of days that are school days in the ozone season[171] = 0.393

G .5.4 School Loss Days, All Cause (Gilliland et al., 2001)

Gilliland et al. (2001) examined the association between air pollution and school absenteeism among 4th grade school children (ages 9-10) in 12 southern Californian communities. The study was conducted from January through June 1996. The authors used school records to collect daily absence data and parental telephone interviews to identify causes. They defined illness-related absences as respiratory or non-respiratory. A respiratory illness was defined as an illness that included at least one of the following: runny nose/sneezing, sore throat, cough, earache, wheezing, or asthma attack. The authors used 15 and 30 day distributed lag models to quantify the association between ozone, PM_{10}, and NO_2 and incident school absences. Ozone levels were positively associated with all school absence measures and significantly associated with all illness-related school absences (non-respiratory illness, respiratory illness, URI and LRI). Neither PM_{10} nor NO_2 was significantly associated with illness-related school absences, but PM_{10} was associated with non-illness related absences. The C-R function for ozone is based on the results of the single pollutant model.

Gilliland et al. (2001) defines an incident absence as an absence that followed attendance on the previous day and the incidence rate as the number of incident absences on a given day over the population at risk for an absence on a given day (i.e. those children who were not absent on the previous day). Since school absences due to air pollution may last longer than one day, an estimate of the average duration of school absences could be used to calculated the total avoided school loss days from an estimate of avoided new absences. A simple ratio of the total absence rate divided by the new absence rate would provide an estimate of the average duration of school absences, which could be applied to the estimate of avoided new absences as follows:

$$Duration = \frac{totalAbsences}{newAbsences}$$

$$\Delta\, TotalAbsences = -\left[incidence \cdot \left(e^{-\beta \cdot \Delta O_3} - 1\right)\right] \cdot duration \cdot pop$$

[170] National school absence rate of 5.5% obtained from the U.S. Department of Education (1996, Table 42-1). Study-specific school absence rate of 5.09% obtained from Chen et al. (2000, Table 1).

[171] Ozone is modeled for the 5 months from May 1 through September 30. We assume that children are in school during weekdays for all of May, 2 weeks in June, 1 week in August, and all of September. This corresponds to approximately 2.75 months out of the 5 month season, resulting in an estimate of 39.3% of days (2.75/5*5/7).

Since the function is log-linear, the baseline incidence rate (in this case, the rate of new absences) is multiplied by duration, which reduces to the total school absence rate. Therefore, the same result would be obtained by using a single estimate of the total school absence rate in the C-R function. Using this approach, we assume that the same relationship observed between pollutant and new school absences in the study would be observed for total absences on a given day. As a result, the total school absence rate is used in the function below. The derivation of this rate is described in the section on baseline incidence rate estimation.

Single Pollutant Model

For all absences, the coefficient and standard error are based on a percent increase of 16.3 percent (95% CI -2.6 percent, 38.9 percent) associated with a 20 ppb increase in 8-hour average ozone concentration (2001, Table 6, p. 52).

A scaling factor is used to adjust for the number of school days in the ozone season. In the modeling program, the function is applied to every day in the ozone season (May 1 - September 30), however, in reality, school absences will be avoided only on school days. We assume that children are in school during weekdays for all of May, two weeks in June, one week in August, and all of September. This corresponds to approximately 2.75 months out of the 5 month season, resulting in an estimate of 39.3% of days (2.75/5*5/7).

In addition, not all children are at-risk for a new school absence, as defined by the study. On average, 5.5% of school children are absent from school on a given day (U.S. Department of Education, 1996, Table 42-1). Only those who are in school on the previous day are at risk for a new absence (1-0.055 = 94.5%). As a result, a factor of 94.5% is used in the function to estimate the population of school children at-risk for a new absence.

Functional Form: Log-linear
Coefficient: 0.007550
Standard Error: 0.004527
Incidence Rate: daily school absence rate = 0.055 (U.S. Department of Education, 1996, Table 42-1)
Population: population of children ages 9-10 not absent from school on a given day[172] = 94.5% of children ages 9-10
Scaling Factor: Proportion of days that are school days in the ozone season[173] = 0.393

G .5.5 School Loss Days, Illness-Related (Gilliland et al., 2001)

Gilliland et al. (2001) examined the association between air pollution and school absenteeism among 4th grade school children (ages 9-10) in 12 southern Californian communities. The study was conducted from January through June 1996. The authors used school records to collect daily absence data and parental telephone interviews to identify causes. They defined illness-related absences as respiratory or non-respiratory. A respiratory illness was defined as an illness that included at least one of the following: runny nose/sneezing, sore throat, cough, earache, wheezing, or asthma attack. The authors used

[172] The proportion of children not absent from school on a given day (5.5%) is based on 1996 data from the U.S. Department of Education (1996, Table 42-1).

[173] Ozone is modeled for the 5 months from May 1 through September 30. We assume that children are in school during weekdays for all of May, 2 weeks in June, 1 week in August, and all of September. This corresponds to approximately 2.75 months out of the 5 month season, resulting in an estimate of 39.3% of days (2.75/5*5/7).

15 and 30 day distributed lag models to quantify the association between ozone, PM_{10}, and NO_2 and incident school absences. Ozone levels were positively associated with all school absence measures and significantly associated with all illness-related school absences (non-respiratory illness, respiratory illness, URI and LRI). Neither PM_{10} nor NO_2 was significantly associated with illness-related school absences, but PM_{10} was associated with non-illness related absences. The C-R function for ozone is based on the results of the single pollutant model.

Gilliland et al. (2001) defines an incident absence as an absence that followed attendance on the previous day and the incidence rate as the number of incident absences on a given day over the population at risk for an absence on a given day (i.e. those children who were not absent on the previous day). Since school absences due to air pollution may last longer than one day, an estimate of the average duration of school absences could be used to calculated the total avoided school loss days from an estimate of avoided new absences. A simple ratio of the total absence rate divided by the new absence rate would provide an estimate of the average duration of school absences, which could be applied to the estimate of avoided new absences as follows:

$$Duration = \frac{totalAbsences}{newAbsences}$$

$$\Delta\, TotalAbsences = -\left[incidence \cdot (e^{-\beta \cdot \Delta O_3} - 1)\right] \cdot duration \cdot pop$$

Since the function is log-linear, the baseline incidence rate (in this case, the rate of new absences) is multiplied by duration, which reduces to the total school absence rate. Therefore, the same result would be obtained by using a single estimate of the total school absence rate in the C-R function. Using this approach, we assume that the same relationship observed between pollutant and new school absences in the study would be observed for total absences on a given day. As a result, the total school absence rate is used in the function below. The derivation of this rate is described in the section on baseline incidence rate estimation.

Single Pollutant Model

For total illness-related absences, the coefficient and standard error are based on a percent increase of 62.9 percent (95% CI 18.4 percent, 124.1 percent) associated with a 20 ppb increase in 8-hour average ozone concentration (2001, Table 6, p. 52).

A scaling factor is used to adjust for the number of school days in the ozone season. In the modeling program, the function is applied to every day in the ozone season (May 1 - September 30), however, in reality, school absences will be avoided only on school days. We assume that children are in school during weekdays for all of May, two weeks in June, one week in August, and all of September. This corresponds to approximately 2.75 months out of the 5 month season, resulting in an estimate of 39.3% of days (2.75/5*5/7).

In addition, not all children are at-risk for a new school absence, as defined by the study. On average, 5.5% of school children are absent from school on a given day (U.S. Department of Education, 1996, Table 42-1). Only those who are in school on the previous day are at risk for a new absence (1-0.055 = 94.5%). As a result, a factor of 94.5% is used in the function to estimate the population of school children at-risk for a new absence.

Functional Form: Log-linear
Coefficient: 0.024398
Standard Error: 0.008138
Incidence Rate: region-specific daily illness-related school absence rate (Adams et al., 1999, Table 47), assuming 180 school days per year.
Population: population of children ages 9-10 not absent from school on a given day[174] = 94.5% of children ages 9-10
Scaling Factor: Proportion of days that are school days in the ozone season[175] = 0.393

G .5.6 School Loss Days, Respiratory Illness-Related (Gilliland et al., 2001)

Gilliland et al. (2001) examined the association between air pollution and school absenteeism among 4th grade school children (ages 9-10) in 12 southern Californian communities. The study was conducted from January through June 1996. The authors used school records to collect daily absence data and parental telephone interviews to identify causes. They defined illness-related absences as respiratory or non-respiratory. A respiratory illness was defined as an illness that included at least one of the following: runny nose/sneezing, sore throat, cough, earache, wheezing, or asthma attack. The authors used 15 and 30 day distributed lag models to quantify the association between ozone, PM_{10}, and NO_2 and incident school absences. Ozone levels were positively associated with all school absence measures and significantly associated with all illness-related school absences (non-respiratory illness, respiratory illness, URI and LRI). Neither PM_{10} nor NO_2 was significantly associated with illness-related school absences, but PM_{10} was associated with non-illness related absences. The C-R function for ozone is based on the results of the single pollutant model.

Gilliland et al. (2001) defines an incident absence as an absence that followed attendance on the previous day and the incidence rate as the number of incident absences on a given day over the population at risk for an absence on a given day (i.e. those children who were not absent on the previous day). Since school absences due to air pollution may last longer than one day, an estimate of the average duration of school absences could be used to calculated the total avoided school loss days from an estimate of avoided new absences. A simple ratio of the total absence rate divided by the new absence rate would provide an estimate of the average duration of school absences, which could be applied to the estimate of avoided new absences as follows:

$$Duration = \frac{totalAbsences}{newAbsences}$$

$$\Delta\, TotalAbsences = -\left[incidence \cdot (e^{-\beta \cdot \Delta O_3} - 1)\right] \cdot duration \cdot pop$$

Since the function is log-linear, the baseline incidence rate (in this case, the rate of new absences) is multiplied by duration, which reduces to the total school absence rate. Therefore, the same result would be obtained by using a single estimate of the total school absence rate in the C-R function. Using this

[174] The proportion of children not absent from school on a given day (5.5%) is based on 1996 data from the U.S. Department of Education (1996, Table 42-1).

[175] Ozone is modeled for the 5 months from May 1 through September 30. We assume that children are in school during weekdays for all of May, 2 weeks in June, 1 week in August, and all of September. This corresponds to approximately 2.75 months out of the 5 month season, resulting in an estimate of 39.3% of days (2.75/5*5/7).

approach, we assume that the same relationship observed between pollutant and new school absences in the study would be observed for total absences on a given day. As a result, the total school absence rate is used in the function below. The derivation of this rate is described in the section on baseline incidence rate estimation.

Single Pollutant Model

For respiratory illness-related absences, the coefficient and standard error are based on a percent increase of 82.9 percent (95% CI 3.9 percent, 222.0 percent) associated with a 20 ppb increase in 8-hour average ozone concentration (2001, Table 6, p. 52).

A scaling factor is used to adjust for the number of school days in the ozone season. In the modeling program, the function is applied to every day in the ozone season (May 1 - September 30), however, in reality, school absences will be avoided only on school days. We assume that children are in school during weekdays for all of May, two weeks in June, one week in August, and all of September. This corresponds to approximately 2.75 months out of the 5 month season, resulting in an estimate of 39.3% of days (2.75/5*5/7).

In addition, not all children are at-risk for a new school absence, as defined by the study. On average, 5.5% of school children are absent from school on a given day (U.S. Department of Education, 1996, Table 42-1). Only those who are in school on the previous day are at risk for a new absence (1-0.055 = 94.5%). As a result, a factor of 94.5% is used in the function to estimate the population of school children at-risk for a new absence.

Functional Form: Log-linear
Coefficient: 0.030188
Standard Error: 0.014436
Incidence Rate: region-specific daily respiratory illness-related school absence rate (Adams et al., 1999, Table 47), assuming 180 school days per year.
Population: population of children ages 9-10 not absent from school on a given day[176] = 94.5% of children ages 9-10
Scaling Factor: Proportion of days that are school days in the ozone season[177] = 0.393

G .5.7 Worker Productivity: Crocker and Horst (1981)

To monetize benefits associated with increased worker productivity resulting from improved ozone air quality, we used information reported in Crocker and Horst (1981) and summarized in EPA (1994). Crocker and Horst examined the impacts of ozone exposure on the productivity of outdoor citrus workers. The study measured productivity impacts as the change in income associated with a change in ozone exposure, given as the elasticity of income with respect to ozone concentration (-0.1427).[178] The

[176] The proportion of children not absent from school on a given day (5.5%) is based on 1996 data from the U.S. Department of Education (1996, Table 42-1).

[177] Ozone is modeled for the 5 months from May 1 through September 30. We assume that children are in school during weekdays for all of May, 2 weeks in June, 1 week in August, and all of September. This corresponds to approximately 2.75 months out of the 5 month season, resulting in an estimate of 39.3% of days (2.75/5*5/7).

[178] The relationship estimated by Crocker and Horst between wages and ozone is a log-log relationship. Therefore the elasticity of wages with respect to ozone is a constant, equal to the coefficient of the log of ozone in the model.

reported elasticity translates a ten percent reduction in ozone to a 1.4 percent increase in income. Given the national median daily income for outdoor workers engaged in strenuous activity reported by the U.S. Census Bureau (2002), $68 per day (2000$),[179] a ten percent reduction in ozone yields about $0.97 in increased daily wages. We adjust the national median daily income estimate to reflect regional variations in income using a factor based on the ratio of county median household income to national median household income. No information was available for quantifying the uncertainty associated with the central valuation estimate. Therefore, no uncertainty analysis was conducted for this endpoint.

Single Pollutant Model

The C-R function for estimating changes in worker productivity is shown below:

$$\Delta productivity = \beta \cdot \frac{Q_1 - Q_0}{Q_1} \cdot dailyincome \cdot pop \cdot ,$$

Functional Form: Linear
Coefficient: 0.1427
Daily Income: median daily income for outdoor workers[180]
Population: population of adults 18 to 64 employed as farm workers.

[179] The national median daily income for workers engaged in "farming, forestry, and fishing" from the U.S. Census Bureau (2002, Table 621, p. 403) is used as a surrogate for outdoor workers engaged in strenuous activity.

[180] The national median daily income for workers engaged in "farming, forestry, and fishing" was obtained from the U.S. Census Bureau (2002, Table 621, p. 403) and is used as a surrogate for outdoor workers engaged in strenuous activity. This national median daily income ($68) is then scaled by the ratio of national median income to county median income to estimate county median daily income for outdoor workers.

Exhibit G-6. Concentration-Response (C-R) Functions for Ozone and Asthma-Related Effects

Endpoint Name	Author	Year	Location	Age	Race	Gender	Other Pollutants	Averaging Time[1]	Functional Form	Beta	Std Error	Notes
Asthma Exacerbation, Asthma Attacks	Whittemore and Korn	1980	Los Angeles, CA	All	All	All	TSP	1-hr max	Logistic	0.001843	0.000715	
Asthma Exacerbation, Cough	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	1-hr max	Logistic	-0.001814	0.000824	Probability of symptoms
Asthma Exacerbation, Cough	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	1-hr max	Logistic	-0.003196	0.001456	Probability of a new onset of symptoms
Asthma Exacerbation, Shortness of Breath	Ostro et al.	1995	Los Angeles, CA	7-12	Black	All	None	1-hr max	Logistic	0.003834	0.001859	
Asthma Exacerbation, Shortness of Breath	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	1-hr max	Logistic	0.000249	0.001140	Probability of symptoms
Asthma Exacerbation, Shortness of Breath	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	1-hr max	Logistic	0	0.001835	Probability of a new onset of symptoms
Asthma Exacerbation, Wheeze	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	1-hr max	Logistic	-0.001547	0.000815	Probability of symptoms
Asthma Exacerbation, Wheeze	Ostro et al.	2001	Los Angeles, CA	8-13	Black	All	None	1-hr max	Logistic	-0.001282	0.001212	Probability of a new onset of symptoms

1. The averaging time refers to the metric used in the benefits model. This may differ slightly from the averaging time used in the study. Refer to the study summaries below for more detail on the specific averaging time used in the study.

G .6 Asthma-Related Effects

Exhibit G-6 summarizes the C-R functions used to estimate the relationship between ozone and asthma-related effects. Detailed summaries of each of the studies used to generate the functions are described below, along with the parameters used in each of the functions.

G .6.1 Asthma Attacks (Whittemore and Korn, 1980)

Whittemore and Korn (1980) examined the relationship between air pollution and asthma attacks in a survey of 443 children and adults, living in six communities in southern California during three 34-week periods in 1972-1975. The analysis focused on TSP and oxidants (O_x). Respirable PM, NO_2, SO_2 were highly correlated with TSP and excluded from the analysis. In a two pollutant model, daily levels of both TSP and oxidants were significantly related to reported asthma attacks. The results from this model were used, and the oxidant result was adjusted so it may be used with ozone data.

Multipollutant Model (ozone and PM_{10})

The daily one-hour ozone coefficient is based on an oxidant coefficient (1.66) estimated from data expressed in ppm. The coefficient is converted to ppb by dividing by 1,000 and to ozone by multiplying by 1.11.[181] The standard error is calculated from the two-tailed p-value (<0.01) reported by Whittemore and Korn (1980, Table 5), which implies a t-value of at least 2.576 (assuming a large number of degrees of freedom).

Functional Form: Logistic
Coefficient: 0.001843
Standard Error: 0.000715
Incidence Rate: daily incidence of asthma attacks = 0.0550[182]
Population: population of asthmatics of all ages = 3.86% of the population of all ages (American Lung Association, 2002c, Table 7)

G .6.2 Asthma Exacerbation, Cough (Ostro et al., 2001)

Ostro et al. (2001) studied the relation between air pollution in Los Angeles and asthma exacerbation in African-American children (8 to 13 years old) from August to November 1993. They used air quality data for PM_{10}, $PM_{2.5}$, NO_2, and O_3 in a logistic regression model with control for age, income,

[181] The study used oxidant measurements in ppm (Whittemore and Korn, 1980, p. 688); these have been converted to ozone measurements in ppb, assuming ozone comprises 90% of oxidants (i.e., 1.11*ozone=oxidant). It is assumed that the harm of oxidants is caused by ozone. The view expressed in the Ozone Staff Paper (U.S. EPA, 1996, p.164) is consistent with assuming that ozone is the oxidant of concern at normal ambient concentrations: "Further, among the photochemical oxidants, the acute-exposure chamber, field, and epidemiological human health data base raises concern only for ozone at levels of photochemical oxidants commonly reported in ambient air. Thus, the staff recommends that ozone remain as the pollutant indicator for protection of public health from exposure to all photochemical oxidants found in the ambient air."

[182] Based on an analysis of the 1999 National Health Interview Survey, the daily incidence of wheezing attacks for adult asthmatics is estimated to be 0.0550. In the same survey, wheezing attacks for children were examined, however, the number of wheezing attacks per year were censored at 12 (compared to censoring at 95 for adults). Due to the potential for underestimation of the number of children's wheezing attacks, we used the adult rate for all individuals.

time trends, and temperature-related weather effects.[183] Asthma symptom endpoints were defined in two ways: "probability of a day with symptoms" and "onset of symptom episodes". New onset of a symptom episode was defined as a day with symptoms followed by a symptom-free day. The authors found cough prevalence associated with PM_{10} and $PM_{2.5}$ and cough incidence associated with $PM_{2.5}$ PM_{10}, and NO_2. Ozone was not significantly associated with cough among asthmatics. The ozone C-R functions are based on the results of single pollutant models looking at both the probability of symptoms and the onset of new symptoms.

Single Pollutant Model (probability of symptoms)

The coefficient and standard error are based on the odds ratio (0.93) and 95% confidence interval (0.87-0.99) reported for a 40 ppb increase in one-hour maximum ozone levels (Ostro et al., 2001, Table 4, p.204).

Functional Form: Logistic
Coefficient: -0.001814
Standard Error: 0.000824
Incidence Rate: daily cough rate per person (Ostro et al., 2001, p.202) = 0.145
Population: asthmatic African-American population ages 8 to 13 = 7.26%[184] of African-American population ages 8 to 13

Single Pollutant Model (probability of a new onset of symptoms)

The coefficient and standard error are based on the odds ratio (0.88) and 95% confidence interval (0.78-0.98) reported for a 40 ppb increase in one-hour maximum ozone levels (Ostro et al., 2001, Table 5, p.204).

The C-R function based on this model will estimate the number of new onset episodes of cough avoided. In order to convert this estimate to the total number of episodes avoided, the results are adjusted by an estimate of the duration of symptom episodes. The average duration can be estimated from Ostro et al. (2001) using the ratio of the probability of a symptom episode to the probability of a new onset episode. For cough, this ratio is 2.2 (14.5% divided by 6.7%) (Ostro et al., 2001, p.202).

In addition, not all children are at-risk for a new onset of cough, as defined by the study. On average, 14.5% of African-American asthmatics have cough on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day are at-risk for a new onset episode (1-0.145 = 85.5%). As a result, a factor of 85.5% is used in the function to estimate the population of African-American 8 to 13 year old children at-risk for a new cough episode.

[183] The authors note that there were 26 days in which $PM_{2.5}$ concentrations were reported higher than PM_{10} concentrations. The majority of results the authors reported were based on the full dataset. These results were used for the basis for the C-R functions.

[184] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

Functional Form: Logistic
Coefficient: -0.003196
Standard Error: 0.001456
Incidence Rate: daily new onset cough rate per person (Ostro et al., 2001, p.202) = 0.067
Population: asthmatic African-American population ages 8 to 13 at-risk for a new episode of cough = 6.21% of African-American population ages 8 to 13 (85.5% at-risk[185] times 7.26% asthmatic[186])
Scaling Factor: average number of consecutive days with a cough episode (days) = 2.2

G .6.3 Asthma Exacerbation, Shortness of Breath (Ostro et al., 1995)

Using a logistic regression estimation, Ostro et al. (1995) estimated the impact of PM_{10}, ozone, NO_2, and SO_2 on the incidence of coughing, shortness of breath, and wheezing in 83 African-American asthmatic children ages 7-12 living in Los Angeles from August through September 1992. Regression results show both PM_{10} and ozone significantly linked to shortness of breath; the beginning of an asthma episode was also significantly linked to ozone. No effect was seen for NO_2 and SO_2. Results for single-pollutant models only were presented in the published paper. The C-R function is based on the model with adjustment for respiratory infection, temperature, and outdoor mold levels (Ostro et al., 1995, Table 3).

Single Pollutant Model

The ozone coefficient and standard error are based on the odds ratio (1.36) and 95% confidence interval (1.02-1.83) (Ostro et al., 1995, Table 3) associated with a change in one-hour daily maximum ozone of 8.02 pphm (80.2 ppb) (Ostro et al., 1995, Table 2).

Functional Form: Logistic
Coefficient: 0.003834
Standard Error: 0.001859
Incidence Rate: daily shortness of breath incidence rate per person (Ostro et al., 1995, p. 715) = 0.056
Population: asthmatic African-American population ages 7 to 12 = 7.26%[187] of African-American population ages 7 to 12

G .6.4 Asthma Exacerbation, Shortness of Breath (Ostro et al., 2001)

Ostro et al. (2001) studied the relationship between air pollution in Los Angeles and asthma exacerbation in African-American children (8 to 13 years old) from August to November 1993. They used air quality data for PM_{10}, $PM_{2.5}$, NO_2, and ozone in a logistic regression model with control for age, income, time trends, and temperature-related weather effects. Asthma symptom endpoints were defined in two ways: "probability of a day with symptoms" and "new onset of a symptom episode". New onset of a symptom episode was defined as a day with symptoms followed by a symptom-free day. The authors

[185] On average, 14.5% of African-American asthmatics have cough episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day (1-0.145 = 85.5%) are at-risk for a new onset episode.

[186] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

[187] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

found that both the prevalent and incident episodes of shortness of breath were associated with $PM_{2.5}$ and PM_{10}. Neither ozone nor NO_2 were significantly associated with shortness of breath among asthmatics. The ozone C-R functions are based on the results of single pollutant models looking at both the probability of symptoms and the onset of new symptoms.

Single Pollutant Model (probability of symptoms)

The coefficient and standard error are based on the odds ratio (1.01) and 95% confidence interval (0.92-1.10) reported for a 40 ppb increase in one-hour maximum ozone levels (Ostro et al., 2001, Table 4, p.204).

Functional Form: Logistic
Coefficient: 0.000249
Standard Error: 0.001140
Incidence Rate: daily shortness of breath rate per person (Ostro et al., 2001, p.202) = 0.074
Population: asthmatic African-American population ages 8 to 13 = 7.26%[188] of African-American population ages 8 to 13

Single Pollutant Model (probability of a new onset of symptoms)

The coefficient and standard error are based on the odds ratio (1.00) and 95% confidence interval (0.87-1.16) reported for a 40 ppb increase in one-hour maximum ozone levels (Ostro et al., 2001, Table 5, p.204).

The C-R function based on this model will estimate the number of new onset episodes of shortness of breath avoided. In order to convert this estimate to the total number of episodes avoided, the results are adjusted by an estimate of the duration of symptom episodes. The average duration can be estimated from Ostro et al. (2001) using the ratio of the probability of a symptom episode to the probability of a new onset episode. For shortness of breath, this ratio is 2.0 (7.4% divided by 3.7%) (Ostro et al., 2001, p.202).

In addition, not all children are at-risk for a new onset of shortness of breath, as defined by the study. On average, 7.4% of African-American asthmatics have shortness of breath episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day are at-risk for a new onset episode (1-0.074 = 92.6%). As a result, a factor of 92.6% is used in the function to estimate the population of African-American 8 to 13 year old children at-risk for a new shortness of breath episode.

Functional Form: Logistic
Coefficient: 0
Standard Error: 0.001835
Incidence Rate: daily new onset shortness of breath rate per person (Ostro et al., 2001, p.202) = 0.037
Population: asthmatic African-American population ages 8 to 13 at-risk for a new episode of shortness of breath = 6.72% of African-American population ages 8 to 13 (92.6% at-risk[189] times 7.26% asthmatic[190])
Scaling Factor: average number of consecutive days with a shortness of breath episode (days) = 2.0

---

[188] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

[189] On average, 7.4% of African-American asthmatics have shortness of breath episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day (1-0.074 = 92.6%) are at-risk for a new onset episode.

[190] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

G .6.5 Asthma Exacerbation, Wheeze (Ostro et al., 2001)

Ostro et al. (2001) studied the relation between air pollution in Los Angeles and asthma exacerbation in African-American children (8 to 13 years old) from August to November 1993. They used air quality data for PM_{10}, $PM_{2.5}$, NO_2, and O_3 in a logistic regression model with control for age, income, time trends, and temperature-related weather effects. Asthma symptom endpoints were defined in two ways: "probability of a day with symptoms" and "onset of symptom episodes". New onset of a symptom episode was defined as a day with symptoms followed by a symptom-free day. The authors found both the prevalence and incidence of wheeze associated with $PM_{2.5}$ PM_{10}, and NO_2. Ozone was not significantly associated with wheeze among asthmatics. The ozone C-R functions are based on the results of single pollutant models looking at both the probability of symptoms and the onset of new symptoms.

Single Pollutant Model (probability of symptoms)

The coefficient and standard error are based on the odds ratio (0.94) and 95% confidence interval (0.88-1.00) reported for a 40 ppb increase in one-hour maximum ozone levels (Ostro et al., 2001, Table 4, p.204).

Functional Form: Logistic
Coefficient: -0.001547
Standard Error: 0.000815
Incidence Rate: daily wheeze rate per person (Ostro et al., 2001, p.202) = 0.173
Population: asthmatic African-American population ages 8 to 13 = 7.26%[191] of African-American population ages 8 to 13

Single Pollutant Model (probability of a new onset of symptoms)

The coefficient and standard error are based on the odds ratio (0.95) and 95% confidence interval (0.86-1.04) reported for a 40 ppb increase in one-hour maximum ozone levels (Ostro et al., 2001, Table 5, p.204).

The C-R function based on this model will estimate the number of new onset episodes of wheeze avoided. In order to convert this estimate to the total number of episodes avoided, the results are adjusted by an estimate of the duration of symptom episodes. The average duration can be estimated from Ostro et al. (2001) using the ratio of the probability of a symptom episode to the probability of a new onset episode. For wheeze, this ratio is 2.3 (17.3% divided by 7.6%) (Ostro et al., 2001, p.202).

In addition, not all children are at-risk for a new onset of wheeze, as defined by the study. On average, 17.3% of African-American asthmatics have wheeze on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day are at-risk for a new onset episode (1-0.173 = 82.7%). As a result, a factor of 82.7% is used in the function to estimate the population of African-American 8 to 13 year old children at-risk for a new wheeze episode.

[191] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

Functional Form: Logistic
Coefficient: -0.001282
Standard Error: 0.001212
Incidence Rate: daily new onset wheeze rate per person (Ostro et al., 2001, p.202) = 0.076
Population: asthmatic African-American population ages 8 to 13 at-risk for a new episode of wheeze = 6.00% of African-American population ages 8 to 13 (82.7% at-risk[192] times 7.26% asthmatic[193])
Scaling Factor: average number of consecutive days with a wheeze episode (days) = 2.3

[192] On average, 17.3% of African-American asthmatics have wheeze episodes on a given day (Ostro et al., 2001, p.202). Only those who are symptom-free on the previous day (1-0.173 = 82.7%) are at-risk for a new onset episode.

[193] The American Lung Association (2002c, Table 9) estimates asthma prevalence for African-American children ages 5 to 17 at 7.26% (based on data from the 1999 National Health Interview Survey).

Appendix H: Economic Value of Health Effects

This appendix first presents an overview of valuation, and then presents the unit values that are available in BenMAP for each of the health endpoints included in the current suite of C-R functions. Wherever possible, we present a distribution of the unit value, characterizing the uncertainty surrounding any point estimate. The mean of the distribution is taken as the point estimate of the unit value, and the distribution itself is used to characterize the uncertainty surrounding the unit value, which feeds into the uncertainty surrounding the monetary benefits associated with reducing the incidence of the health endpoint. Below we give detailed descriptions of the derivations of unit values and their distributions, as well as tables listing the unit values and their distributions, available for each health endpoint. The definitions of the distributions and their parameters is given in Exhibit H-11, at the end of this Appendix.

H.1 Overview of Valuation

Reductions in ambient concentrations of air pollution generally lower the risk of future adverse health affects by a fairly small amount for a large population. A lower risk for everyone means that fewer cases of the adverse health effect are expected, although we don't know *ex ante* which cases will be avoided. For example, the analysis may predict 100 hospital admissions for respiratory illnesses avoided, but the analysis does not estimate which individuals will be spared those cases of respiratory illness that would have required hospitalization. The health benefits conferred on individuals by a reduction in pollution concentrations are, then, actually *reductions in the risk* of having to endure certain health problems. These benefits (reductions in risk) may not be the same for all individuals (and could be zero for some individuals). Likewise, the WTP for a given benefit is likely to vary from one individual to another. In theory, the total social value associated with the decrease in risk of a given health problem resulting from a given reduction in pollution concentrations is generally taken to be the sum of everyone's WTP for the benefits they receive:

$$\sum_{i=1}^{N} WTP_i\left(B_i\right) ,$$

where B_i is the benefit (i.e., the reduction in risk of having to endure the health problem conferred on the i^{th} individual by the reduction in pollution concentrations, $WTP_i(B_i)$ is the i^{th} individual's WTP for that benefit, and N is the number of people exposed to the pollution.[194] If a reduction in pollution concentrations affects the risks of several health endpoints, the total health-related social value of the reduction in pollution concentrations is:

$$\sum_{i=1}^{N}\sum_{j=1}^{J} WTP_i\left(B_{i,j}\right) ,$$

where B_{ij} is the benefit related to the j^{th} health endpoint (i.e., the reduction in risk of having to endure the j^{th} health problem) conferred on the i^{th} individual by the reduction in pollution concentrations, and $WTP_i(B_{ij})$ is the i^{th} individual's WTP for that benefit.

The reduction in risk of each health problem for each individual is not known, however (nor is each individual's WTP for each possible risk reduction he or she might receive). Instead, epidemiological

[194] WTP may also include altruism – that is, a person may be WTP not only for his own benefits, but for the benefits that would be enjoyed by others.

studies allow us to estimate the number of cases of an adverse health effect that would be avoided by a given reduction in pollutant concentrations. Therefore, in practice, benefit analyses take an *ex post* approach and estimate the value of a *statistical* health problem avoided. If we have an estimate of the average individual's WTP for the risk reduction conferred upon him, we can derive from that an estimate of the value of a statistical case avoided. Suppose, for example, that a given reduction in pollutant concentrations results in a decrease in mortality risk of 1/10,000. Then for every 10,000 individuals, one individual would be expected to die in the absence of the reduction in pollutant concentrations (who would not be expected to die in the presence of the reduction in pollutant concentrations). If the average individual's WTP for this 1/10,000 decrease in mortality risk is $100, then the value of a statistical life is 10,000 x $100, or $1 million. In general, the *ex ante* WTP for a risk reduction of x can be converted into an *ex post* value of a statistical case avoided by dividing the average individual's WTP for the risk reduction of x by x (e.g. $100/0.0001 = $1,000,000). The same type of calculation can produce values for statistical incidences of other health endpoints.

The value of a statistical case avoided is referred to here as a "unit value." The total dollar value for a specific health effect is the number of statistical cases of the health effect avoided times the unit value for that health effect. Whereas ideally the unit value would reflect the underlying WTP for the *ex ante* risk reduction (as in the above example), in practice we usually have estimates of the value of the *ex post* statistical case avoided. Sometimes those values come from contingent valuation studies, in which study participants are queried about their WTP to avoid a specific adverse health effect. Sometimes, when WTP estimates are not available, WTP is approximated by other measures, most notably cost of illness measures.

An individual's WTP to avoid an adverse health effect will include, at a minimum, the amount of money he would have to pay for medical expenses associated with the illness. Because medical expenditures are to a significant extent shared by society, via medical insurance, Medicare, etc., however, the medical expenditures actually incurred by the individual are likely to be less than the total medical cost to society. The total value to society of an individual's avoidance of an adverse health effect, then, might be thought of as having two components: (1) the cost of the illness (COI) to society, including the total value of the medical resources used (some portion of which will be paid by the individual), plus the value of the lost productivity, as well as (2) the WTP of the individual, as well as that of others, to avoid the pain and suffering resulting from the illness.

These two components might be rephrased as (1) the market component and (2) the non-market component. When an individual becomes ill, there is some amount of resources (medical goods and services) that are used to address the illness. The value of those resources, whoever pays the bill, is the market component of the value of avoiding the illness – i.e., the value of the resources that would *not* have to be used up if the individual had not incurred the illness. This may be a small value – e.g., the cost of aspirin used for a headache, or a very large value – e.g., the value of medical goods and services used to treat someone who goes to the hospital with a life-threatening illness. The COI approach attempts to estimate the total value of the medical resources used up as well as the value of the individual's time lost as a result of the illness. Because this method does not include the value of avoiding the pain and suffering resulting from the illness (a potentially large component), it is generally believed to underestimate the total value of avoiding the illness, perhaps substantially.

The contingent valuation method attempts to elicit from people what they would be willing to pay to avoid the illness. Because of the distortion in the market for medical goods and services, whereby individuals generally do not pay the full value of the medical resources used to address their illnesses, however, this method too is likely to understate the total value of avoiding the illness.

Although the COI and contingent valuation approaches to valuing health effects avoided are the two most common methods, other methods have been used in certain circumstances. The method the benefit analyst chooses to value a particular health endpoint will depend in part on what is available. Benefit analysts typically do not do primary research to generate data for valuation to be used in a benefit analysis – it is too expensive and time consuming. Instead, the benefit analyst uses data or estimates that have been collected or generated by researchers and can be readily obtained in publicly available databases or in the open literature. The unit values currently available for use in BenMap are all of this type.

Sometimes more than one estimate of a unit value for a health effect is available. For chronic bronchitis, for example, we have both a WTP estimate and COI estimates. In that case, you may select one or pool two or more estimates. As research continues and new unit values become available, the database of unit values available for use in BenMAP will be updated. The discussion below refers to the set of unit values that are currently available for use in BenMAP. Unless otherwise stated, all unit values are in 2000$.

H.2 Mortality

The economics literature concerning the appropriate method for valuing reductions in premature mortality risk is still developing. The adoption of a value for the projected reduction in the risk of premature mortality is the subject of continuing discussion within the economics and public policy analysis communities. Issues such as the appropriate discount rate and whether there are factors, such as age or the quality of life, that should be taken into consideration when estimating the value of avoided premature mortality are still under discussion. BenMAP currently offers a variety of options reflecting the uncertainty surrounding the unit value for premature mortality.

H.2.1 Value of a Statistical Life Based on 26 Studies

One unit value available in BenMAP is $6.3 million. This estimate is the mean of a distribution fitted to 26 "value of statistical life" (VSL) estimates that appear in the economics literature and that have been identified in the Section 812 Reports to Congress as "applicable to policy analysis." This represents an intermediate value from a variety of estimates, and it is a value EPA has frequently used in Regulatory Impact Analyses (RIAs) as well as in the Section 812 Retrospective and Prospective Analyses of the Clean Air Act.

The VSL approach and the set of selected studies mirrors that of Viscusi (1992) (with the addition of two studies), and uses the same criteria as Viscusi in his review of value-of-life studies. The $6.3 million estimate is consistent with Viscusi's conclusion (updated to 2000$) that "most of the reasonable estimates of the value of life are clustered in the $3.8 to $8.9 million range." Five of the 26 studies are contingent valuation (CV) studies, which directly solicit WTP information from subjects; the rest are wage-risk studies, which base WTP estimates on estimates of the additional compensation demanded in the labor market for riskier jobs. Because this VSL-based unit value does not distinguish among people based on the age at their death or the quality of their lives, it can be applied to all premature deaths.

H.2.2 Value of a Statistical Life Based on Selected Studies

In addition to the value of a statistical based on the results of 26 studies, we have included four alternatives based loosely on the results of recent work by Mrozek and Taylor (2002) and Viscusi and

Aldy (2003). Each of the four alternatives has a mean value of $5.5 million (2000$), but with a different distributions: normal, uniform, triangular, and beta. Exhibit H-1 presents the distribution parameters for the suite of mortality valuations currently available in BenMAP.

Exhibit H-1. Unit Values Available for Mortality

Basis for Estimate [a]	Age Range at Death		Unit Value (VSL) (2000$)	Distribution of Unit Value**	Parameters of Distribution	
	min.	*max.*			*P1*	*P2*
VSL, based on 26 value-of-life studies.	0	99	$6,324,101	Weibull	5.32E-6	1.509588
VSL based on range from $1 million to $10 million – 95% CI of assumed normal distribution.	0	99	$5,500,000	Normal	2,295,960.54	--
VSL based on range from $1 million to $10 million – assumed uniform distribution.	0	99	$5,500,000	Uniform	1,000,000	10,000,000
VSL based on range from $1 million to $10 million – assumed triangular distribution.	0	99	$5,500,000	Triangular	1,000,000	10,000,000
VSL based on range from $1 million to $10 million – 95% CI of assumed beta distribution. [B]	0	99	$5,500,000	Beta	1.95	1.95

[a] The original value of a statistical life was calculated in 1990 $. We have used a factor of 1.3175, based on the All-Items CPI-U.

[b] The Beta distribution in this instance also has a scale parameter equal to 10993993.6.

H.3 Chronic Illness

This sub-section presents the unit values developed for chronic bronchitis, chronic asthma, and non-fatal myocardial infarctions.

H.3.1 Chronic Bronchitis

PM-related chronic bronchitis is expected to last from the initial onset of the illness throughout the rest of the individual's life. WTP to avoid chronic bronchitis would therefore be expected to incorporate the present discounted value of a potentially long stream of costs (e.g., medical expenditures and lost earnings) as well as WTP to avoid the pain and suffering associated with the illness. Both WTP and COI estimates are currently available in BenMAP.

Unit Value Based on Two Studies of WTP

Two contingent valuation studies, Viscusi et al. (1991) and Krupnick and Cropper (1992), provide estimates of WTP to avoid a case of chronic bronchitis. Viscusi et al. (1991) and Krupnick and Cropper (1992) were experimental studies intended to examine new methodologies for eliciting values for morbidity endpoints. Although these studies were not specifically designed for policy analysis, they can be used to provide reasonable estimates of WTP to avoid a case of chronic bronchitis. As with other contingent valuation studies, the reliability of the WTP estimates depends on the methods used to obtain the WTP values. The Viscusi et al. and the Krupnick and Cropper studies are broadly consistent with current contingent valuation practices, although specific attributes of the studies may not be.

The study by Viscusi et al. (1991) uses a sample that is larger and more representative of the general population than the study by Krupnick and Cropper (1992), which selects people who have a relative with the disease. However, the chronic bronchitis described to study subjects in the Viscusi study is severe, whereas a pollution-related case may be less severe.

The relationship between the severity of a case of chronic bronchitis and WTP to avoid it was estimated by Krupnick and Cropper (1992). We used that estimated relationship to derive a relationship between WTP to avoid a severe case of chronic bronchitis, as described in the Viscusis study, and WTP to avoid a less severe case. The estimated relationship (see Table 4 in Krupnick and Cropper) can be written as:

$$\ln(WTP) = \alpha + \beta * sev$$

where α denotes all the other variables in the regression model and their coefficients, β is the coefficient of sev, estimated to be 0.18, and sev denotes the severity level (a number from 1 to 13). Let x (< 13) denote the severity level of a pollution-related case of chronic bronchitis, and 13 denote the highest severity level (as described in Viscusi et al., 1991). Then

$$\ln(WTP_{13}) = \alpha + \beta * 13$$

and

$$\ln(WTP_x) = \alpha + \beta * x.$$

Subtracting one equation from the other,

$$\ln(WTP_{13}) - \ln(WTP_x) = \beta * (13 - x)$$

or

$$\ln\left(\frac{WTP_{13}}{WTP_x}\right) = \beta * (13 - x) \quad .$$

Exponentiating and rearranging terms,

$$WTP_x = WTP_{13} * e^{-\beta * (13 - x)} .$$

There is uncertainty surrounding the exact values of WTP_{13}; x, and β, and this uncertainty can be incorporated in the equation, if you request that the analysis be carried out in "uncertainty mode." The distribution of WTP to avoid a severe case of chronic bronchitis, WTP_{13} ,is based on the distribution of WTP responses in the Viscusi et al. (1991) study. The distribution of x, the severity level of an average case of pollution-related chronic bronchitis, is modeled as a triangular distribution centered at 6.5, with endpoints at 1.0 and 12.0. And the distribution of β is normal with mean = 0.18 and std. dev.= 0.0669 (the estimate of b and standard error reported in Krupnick and Cropper, 1992).

In uncertainty mode, BenMAP uses a Monte Carlo approach. On each Monte Carlo iteration, random draws for these three variables are made, and the resulting WTP_x is calculated from the equation above. Because this function is non-linear, the expected value of WTP for a pollution-related case of CB

cannot be obtained by using the expected values of the three uncertain inputs in the function (doing that will substantially understate mean WTP). A Monte Carlo analysis suggests, however, that the mean WTP to avoid a case of pollution-related chronic bronchitis is about $340,000. Therefore, if you request that the analysis be carried out in "point estimate" mode, that is the unit value that is used.

Alternative Cost of Illness Estimates

Cost of illness estimates for chronic bronchitis were derived from estimates of annual medical costs and annual lost earnings by Cropper and Krupnick (1990). This study estimated annual lost earnings resulting from chronic bronchitis as a function of age at onset of the illness, for the following age categories: 25-43, 35-44, 45-54, and 55-65 (see Cropper and Krupnick, Table 8). Annual medical expenses were estimated for 10-years age groups (0-9, 10-19, 20-29, ..., 80-89). We derived estimates of the present discounted value of the stream of medical and opportunity costs for people whose age of onset is 30, 40, 50, 60, 70, and 80. Medical costs (which are in 1977$ in the Cropper and Krupnick study) were inflated to 2000$ using the CPI-U for medical care; lost earnings (opportunity costs) were inflated to 2000$ using the Employment Cost Index for Wages and Salaries. Life expectancies were assumed to be unaffected by the illness.[195] For example, an individual at age 70 has a life expectancy of 14.3 more years, and we assumed that someone whose age of onset of chronic bronchitis is 70 will also live for 14.3 more years. We also assumed that opportunity costs at ages 66 and over were zero. Present discounted values were calculated using three and seven percent discount rates.

For each of the two discount rates, there are three cost of illness unit values for chronic bronchitis available in BenMAP, for the following age categories: 27-44, 45-64, and 65+. These are the age categories that were used in the epidemiological study that estimated a concentration-response function for chronic bronchitis (Abbey et al., 1995b). The estimate for the 27-44 age group is an average of the present discounted values calculated for ages 30 and 40; the estimate for the 45-64 age category is an average of the present discounted values calculated for ages 50 and 60; and the estimate for the 65+ age category is an average of the present discounted values calculated for ages 70 and 80. The suite of unit values available for use in BenMAP are shown in Exhibit H-2 below.

Exhibit H-2. Unit Values Available for Chronic Bronchitis

Basis for Estimate	Age of Onset		Present Discounted Value of Medical Costs	Present Discounted Value of Opportunity Costs	Unit Value	Distribution
	min.	max.				
WTP: average severity	30	99	N/A	N/A	$340,482	custom
COI: med costs + wage loss, 3% DR	27	44	$18,960	$135,463	$154,422	none
	45	64	$23,759	$76,029	$99,788	none
	65	99	$11,088	$0	$11,088	none
COI: med costs + wage loss, 7% DR	27	44	$7,886	$80,444	$88,331	none
	45	64	$14,390	$59,577	$73,967	none
	65	99	$9,030	$0	$9,030	none

[195] Source of life expectancies: National Vital Statistics Reports, Volume 47, No. 19, June 30, 1999. Table 5: "Life expectancy at selected ages by race and sex: United States, 1997.

H.3.2 Chronic Bronchitis Reversals

The unit value for chronic bronchitis reversals assumes that this is chronic bronchitis with a severity level of 1. The method for generating a distribution of unit values in BenMAP is therefore the same as the WTP-based unit value method for chronic bronchitis (see above), with x=1. The mean of this distribution is $150,221.

H.3.3 Chronic Asthma

Two studies have estimated WTP to avoid chronic asthma in adults. Blumenschein and Johannesson (1998) used two different contingent valuation (CV) methods, the dichotomous choice method and a bidding game, to estimate mean willingness to pay for a cure for asthma. The mean WTP elicited from the bidding game was $189 per month, or $2,268 per year (in 1996$). The mean WTP elicited from the dichotomous choice approach was $343 per month, or $4,116 per year (in 1996$). Using $2,268 per year, a three percent discount rate, and 1997 life expectancies for males in the United States (National Center for Health Statistics, 1999, Table 5), the present discounted value of the stream of annual WTPs is $47,637 (in 2000$).

O'Conor and Blomquist (1997) estimated WTP to avoid chronic asthma from estimates of risk-risk tradeoffs. Combining the risk-risk tradeoffs with a statistical value of life, the annual value of avoiding asthma can be derived. Assuming a value of a statistical life of $6 million, they derived an annual WTP to avoid asthma of $1500 (O'Connor and Blomquist, 1997, p. 677). For a value of a statistical life of $5,894,400 (in 1997 $), the corresponding implied annual value of avoiding chronic asthma, based on O'Conor and Blomquist would be $1,474. Assuming a three percent discount rate and 1997 life expectancies for males in the United States, the present discounted value of the stream of annual WTPs would be $30,257 (in 2000$). A unit value, based on a three percent discount rate, is the average of the two estimates, or $38,947. Following the method used for the §812 Prospective analysis, the uncertainty surrounding the WTP to avoid a case of chronic asthma among adult males was characterized by a triangular distribution on the range determined by the two study-specific WTP estimates.

A second unit value, using a seven percent discount rate, is also available for use in BenMAP. The method used to derive this unit value is the same as that described above for the three percent discount rate unit value. The unit values available for use in BenMAP are summarized in Exhibit H-3 below.

Exhibit H-3. Unit Values Available for Chronic Asthma

Basis for Estimate	Age Range		Unit Value	Distribution of Unit Value	Parameters of Distribution	
	min.	max.			P1	P2
WTP: 3% DR (Discount Rate)	27	99	$38,947	triangular	$30,257	$47,637
WTP: 7% DR	27	99	$25,357	triangular	$19,699	$31,015

H.3.4 Non-Fatal Myocardial Infarctions (Heart Attacks)

In the absence of a suitable WTP value for reductions in the risk of non-fatal heart attacks, there are a variety of cost-of-illness unit values available for use in BenMAP. These cost-of-illness unit values incorporate two components: the direct medical costs and the opportunity cost (lost earnings) associated

with the illness event. Because the costs associated with a heart attack extend beyond the initial event itself, the unit values include costs incurred over five years. Using age-specific annual lost earnings estimated by Cropper and Krupnick (1990), and a three percent discount rate, we estimated the following present discounted values in lost earnings over 5 years due to a heart attack: $8,774 for someone between the ages of 25 and 44, $12,932 for someone between the ages of 45 and 54, and $74,746 for someone between the ages of 55 and 65. The corresponding age-specific estimates of lost earnings using a seven percent discount rate are $7,855, $11,578, and $66,920, respectively. Cropper and Krupnick do not provide lost earnings estimates for populations under 25 or over 65. As such we do not include lost earnings in the cost estimates for these age groups.

We have found three possible sources of estimates of the direct medical costs of a myocardial infarction (MI) in the literature:

- Wittels et al. (1990) estimated expected total medical costs of MI over 5 years to be $51,211 (in 1986$) for people who were admitted to the hospital and survived hospitalization. (There does not appear to be any discounting used.) Wittels et al. was used to value coronary heart disease in the 812 Retrospective Analysis of the Clean Air Act. Using the CPI-U for medical care, the Wittels estimate is $109,474 in year 2000$. This estimated cost is based on a medical cost model, which incorporated therapeutic options, projected outcomes and prices (using "knowledgeable cardiologists" as consultants). The model used medical data and medical decision algorithms to estimate the probabilities of certain events and/or medical procedures being used. The authors note that the average length of hospitalization for acute MI has decreased over time (from an average of 12.9 days in 1980 to an average of 11 days in 1983). Wittels et al. used 10 days as the average in their study. It is unclear how much further the length of stay (LOS) for MI may have decreased from 1983 to the present. The average LOS for ICD code 410 (MI) in the year-2000 AHQR HCUP database is 5.5 days. However, this may include patients who died in the hospital (not included among our non-fatal MI cases), whose LOS was therefore substantially shorter than it would be if they hadn't died.

- Eisenstein et al. (2001) estimated 10-year costs of $44,663, in 1997$ (using a three percent discount rate), or $49,651 in 2000$ for MI patients, using statistical prediction (regression) models to estimate inpatient costs. Only inpatient costs (physician fees and hospital costs) were included.

- Russell et al. (1998) estimated first-year direct medical costs of treating nonfatal MI of $15,540 (in 1995$), and $1,051 annually thereafter. Converting to year 2000$, that would be $18,880 for a 5-year period, using a three percent discount rate, or $17,850, using a seven percent discount rate.

The age group-specific estimates of opportunity cost over a five-year period are combined with the medical cost estimates from each of the three studies listed above. Because opportunity costs are derived for each of five age groups, there are 3 x 5 = 15 unit values for each of 2 discount rates, or 30 unit values available for use in BenMAP.[196] These are given in Exhibit H-4 below.

[196] We were unable to achieve complete consistency, unfortunately, because of limitations in the input studies. For example, although we calculated opportunity costs over a five-year period using a 3 percent and a 7 percent discount rate, we were not able to do the same for medical costs, except for the medical costs estimated by Russell et al. (in which they estimate an annual cost). Wittels et al. appear to have used no discounting in their estimate; Eisenstein et al. used a 3 percent discount rate. Similarly, although almost all cost estimates (opportunity costs and medical costs) are for a 5-year period, the medical cost estimate reported by Eisenstein et al. is for a 10-year period. There was no reasonable method for inferring from that study what costs over a 5-year period would be.

Exhibit H-4 Unit Values Available for Myocardial Infarction

Basis of Estimate	Age Range		Medical Cost [a]	Opportunity Cost [b]	Total Cost
	Min	*Max*			
COI: 5 yrs med, 5 yrs wages, 3% DR, Wittels (1990)	0	24	$109,474	$0	$109,474
	25	44	$109,474	$9,033	$118,507
	45	54	$109,474	$13,313	$122,787
	55	65	$109,474	$76,951	$186,425
	66	99	$109,474	$0	$109,474
COI: 10 yrs med, 5 yrs wages, 3% DR, Eisenstein (2001)	0	24	$49,651	$0	$49,651
	25	44	$49,651	$9,033	$58,683
	45	54	$49,651	$13,313	$62,964
	55	65	$49,651	$76,951	$126,602
	66	99	$49,651	$0	$49,651
COI: 5 yrs med, 5 yrs wages, 3% DR, Russell (1998)	0	24	$22,331	$0	$22,331
	25	44	$22,331	$9,033	$31,363
	45	54	$22,331	$13,313	$35,644
	55	65	$22,331	$76,951	$99,281
	66	99	$22,331	$0	$22,331
COI: 5 yrs med, 5 yrs wages, 7% DR, Wittels (1990)	0	24	$109,474	$0	$109,474
	25	44	$109,474	$8,087	$117,561
	45	54	$109,474	$11,919	$121,393
	55	65	$109,474	$68,894	$178,368
	66	99	$109,474	$0	$109,474
COI: 10 yrs med, 5 yrs wages, 7% DR, Eisenstein (2001)	0	24	$49,651	$0	$49,651
	25	44	$49,651	$8,087	$57,738
	45	54	$49,651	$11,919	$61,570
	55	65	$49,651	$68,894	$118,545
	66	99	$49,651	$0	$49,651
COI: 5 yrs med, 5 yrs wages, 7% DR, Russell (1998)	0	24	$21,113	$0	$21,113
	25	44	$21,113	$8,087	$29,200
	45	54	$21,113	$11,919	$33,032
	55	65	$21,113	$68,894	$90,007
	66	99	$21,113	$0	$21,113

[a] From Cropper and Krupnick (1990). Present discounted value of 5 yrs of lost earnings, at 3% and 7% discount rate, adjusted from 1977$ to 2000$ using CPI-U "all items".

[b] An average of the 5-year costs estimated by Wittels et al. (1990) and Russell et al. (1998). Note that Wittels et al. appears not to have used discounting in deriving a 5-year cost of $109,474; Russell et al. estimated first-year direct medical costs and annual costs thereafter. The resulting 5-year cost is $22,331, using a 3% discount rate, and $21,113, using a 7% discount rate. Medical costs were inflated to 2000$ using CPI-U for medical care.

H.4 Hospital Admissions & Emergency Room Visits

This section presents the values for avoided hospital admissions, as well as avoided emergency room visits. We assume that hospital admissions due to acute exposure to air pollution pass through the emergency room. However, the value of hospital admissions that we have calculated here does not account for the cost incurred in the emergency room visit.

H.4.1 Hospital Admissions

As suggested above, the total value to society of an individual's avoidance of a hospital admission can be thought of as having two components: (1) the cost of illness (COI) to society, including the total medical costs plus the value of the lost productivity, as well as (2) the WTP of the individual, as well as that of others, to avoid the pain and suffering resulting from the illness.

In the absence of estimates of social WTP to avoid hospital admissions for specific illnesses (components 1 plus 2 above), estimates of total COI (component 1) are available for use in BenMAP as conservative (lower bound) estimates. Because these estimates do not include the value of avoiding the pain and suffering resulting from the illness (component 2), they are biased downward. Some analyses adjust COI estimates upward by multiplying by an estimate of the ratio of WTP to COI, to better approximate total WTP. Other analyses have avoided making this adjustment because of the possibility of over-adjusting -- that is, possibly replacing a known downward bias with an upward bias. Based on Science Advisory Board (SAB) advice, the COI values currently available for use in BenMAP are not adjusted.

Unit values are based on ICD-code-specific estimated hospital charges and opportunity cost of time spent in the hospital (based on the average length of a hospital stay for the illness). The opportunity cost of a day spent in the hospital is estimated as the value of the lost daily wage, regardless of whether or not the individual is in the workforce.

For all hospital admissions endpoints available in BenMAP, estimates of hospital charges and lengths of hospital stays were based on discharge statistics provided by the Agency for Healthcare Research and Quality's Healthcare Utilization Project (2000). The total COI for an ICD-code-specific hospital stay lasting n days is estimated as the mean hospital charge plus n times the daily lost wage. Year 2000 county-specific median annual wages[197] divided by (52*5) were used to estimate county-specific median daily wages. Because wage data used in BenMAP are county-specific, the unit value for a hospital admission varies from one county to another.

Most hospital admissions categories considered in epidemiological studies consisted of sets of ICD codes. The unit value for the set of ICD codes was estimated as the weighted average of the ICD-code-specific COI estimates. The weights were the relative frequencies of the ICD codes among hospital discharges in the United States, as estimated by the National Hospital Discharge Survey (Owings and Lawrence, 1999, Table 1). The hospital admissions for which unit values are available in BenMAP are given in Exhibit H-5. Although unit values available for use in BenMAP are county-specific, the national median daily wage was used to calculate opportunity costs and total costs for the table below, to give a general idea of the cost of illness estimates for the different hospital admissions endpoints.

[197] Source: U.S. Year 2000 Census, compiled by Geolytics.

The mean hospital charges and mean lengths of stay provided by (AHRQ 2000) are based on a very large nationally representative sample of about seven million hospital discharges, and are therefore the best estimates of mean hospital charges and mean lengths of stay available, with negligible standard errors.

Exhibit H-5. Unit Values Available for Hospital Admissions

EndPoint	ICD Codes	Age Range		Mean Hospital Charge [a]	Mean Length of Stay (days) [a]	Total Cost of Illness (Unit Value) [b]
		min.	*max.*			
HA, All Cardiovascular	390-429	65	99	$20,607	5.07	$21,191
HA, All Cardiovascular	390-429	0	99	$20,873	4.71	$21,415
HA, All Cardiovascular	390-429	20	64	$22,300	4.15	$22,778
HA, Congestive Heart Failure	428	65	99	$14,573	5.60	$15,218
HA, Dysrhythmia	427	0	99	$14,811	3.70	$15,237
HA, Ischemic Heart Disease	410-414	65	99	$25,322	4.81	$25,876
HA, All Respiratory	460-519	65	99	$17,600	6.88	$18,393
HA, All Respiratory	460-519	0	99	$14,999	5.63	$15,647
HA, All Respiratory	460-519	0	2	$7,416	2.97	$7,759
HA, Asthma	493	0	64	$7,448	2.95	$7,788
HA, Asthma	493	65	99	$11,417	4.99	$11,991
HA, Asthma	493	0	99	$8,098	3.30	$8,478
HA, Chronic Lung Disease	490-496	65	99	$12,781	5.59	$13,425
HA, Chronic Lung Disease	490-496	0	99	$10,882	4.59	$11,412
HA, Chronic Lung Disease	490-496	20	64	$10,194	4.04	$10,660
HA, Chronic Lung Disease (less Asthma)	490-492, 494-496	65	99	$12,993	5.69	$13,648
HA, Chronic Lung Disease (less Asthma)	490-492, 494-496	0	99	$12,742	5.45	$13,370
HA, Chronic Lung Disease (less Asthma)	490-492, 494-496	20	64	$11,820	4.48	$11,820
HA, Pneumonia	480-487	65	99	$17,030	7.07	$17,844
HA, Pneumonia	480-487	0	99	$14,693	5.92	$15,375

[a] Source of hospital charges and lengths of stay: Agency for Healthcare Research and Quality. 2000. HCUPnet, Healthcare Cost and Utilization Project. http://www.agrq.gov/data/hcup/hcupnet.htm .

[b] The opportunity cost of a day spent in the hospital was estimated, for this exhibit, at the median daily wage of all workers, $115.20, regardless of age. The median daily wage was calculated by dividing the median weekly wage ($576 in 2000$) by 5. The median weekly wage was obtained from U.S. Census Bureau, Statistical Abstract of the United States: 2001, Section 12, Table 621: "Full-Time Wage and Salary Workers – Numbers and Earnings: 1985 to 2000." Actual unit values used in BenMAP are based on county-specific wages, and are therefore county-specific.

H.4.2 Emergency Room Visits for Asthma

Two unit values are currently available for use in BenMAP for asthma emergency room (ER) visits. One is $311.55, from Smith et al., 1997, who reported that there were approximately 1.2 million asthma-related ER visits made in 1987, at a total cost of $186.5 million, in 1987$. The average cost per visit was therefore $155 in 1987$, or $311.55 in 2000 $ (using the CPI-U for medical care to adjust to 2000$). The uncertainty surrounding this estimate, based on the uncertainty surrounding the number of

ER visits and the total cost of all visits reported by Smith et al. is characterized by a triangular distribution centered at $311.55, on the interval [$230.67, $430.93].

A second unit value is $260.67 from Stanford et al. (1999). This study considered asthmatics in 1996-1997, in comparison to the Smith et al. (1997) study, which used 1987 National Medical Expenditure Survey (NMES) data). In comparing their study, the authors note that the 1987 NMES, used by Smith et al., "may not reflect changes in treatment patterns during the 1990s." In addition, its costs are the costs to the hospital (or ER) for treating asthma rather than charges or payments by the patient and/or third party payer. Costs to the ER are probably a better measure of the value of the medical resources used up on an asthma ER visit (see above for a discussion of costs versus charges).

The unit values and the corresponding distributions available in BenMAP for asthma-related ER visits are summarized in Exhibit H-6.

Exhibit H-6. Unit Values Available for Asthma-Related ER Visits

Basis for Estimate	Age Range		Unit Value	Distribution of Unit Value	Parameters of Distribution	
	min.	max.			P1	P2
COI: Smith et al. (1997)	0	99	$312	triangular	$231	$431
COI: Standford et al. (1999)	0	99	$261	normal	5.22	–

H.5 Acute Symptoms and Illness Not Requiring Hospitalization

Several acute symptoms and illnesses have been associated with air pollution, including acute bronchitis in children, upper and lower respiratory symptoms, and exacerbation of asthma (as indicated by one of several symptoms whose occurrence in an asthmatic generally suggests the onset of an asthma episode). In addition, several more general health endpoints which are associated with one or more of these acute symptoms and illnesses, such as minor restricted activity days, school loss days, and work loss days, have also been associated with air pollution. We briefly discuss the derivation of the unit values for each of these acute symptoms and illnesses, and then present all of these unit values in exhibits H-9 and H-10 at the end of this section.

For several of the acute symptoms and illnesses for which more than one unit value is available in BenMAP, one of these is the value that EPA used in several recent benefits analyses. These "original" unit values were all based on a set of three CV studies, in which respondents were asked their WTP to avoid a day of specific symptoms. These study- and symptom-specific WTP estimates, along with the recommended midrange estimates derived by IEc (1993) on which the original unit values were based, are presented in Exhibit H-7 below.

Exhibit H-7. Median WTP Estimates and Derived Midrange Estimates (in 1999 $)

Symptom [a]	Dickie et al. (1987)	Tolley et al. (1986)	Loehman et al. (1979)	Mid-Range Estimate
Throat congestion	4.81	20.84	-	12.75
Head/sinus congestion	5.61	22.45	10.45	12.75
Coughing	1.61	17.65	6.35	8.93
Eye irritation	-	20.03	-	20.03
Headache	1.61	32.07	-	12.75
Shortness of breath	0.00	-	13.47	6.37
Pain upon deep inhalation (PDI)	5.63	-	-	5.63
Wheeze	3.21	-	-	3.21
Coughing up phlegm	3.51 [b]	-	-	3.51
Chest tightness	8.03	-	-	8.03

[a] All estimates are WTP to avoid one day of symptom. Midrange estimates were derived by IEc (1993).

[b] 10% trimmed mean.

H.5.1 Acute Bronchitis in Children

Estimating WTP to avoid a case of acute bronchitis is difficult for several reasons. First, WTP to avoid acute bronchitis itself has not been estimated. Estimation of WTP to avoid this health endpoint therefore must be based on estimates of WTP to avoid symptoms that occur with this illness. Second, a case of acute bronchitis may last more than one day, whereas it is a day of avoided symptoms that is typically valued. Finally, the C-R function used in the benefit analysis for acute bronchitis was estimated for children, whereas WTP estimates for those symptoms associated with acute bronchitis were obtained from adults.

Three unit values are available in BenMAP for acute bronchitis in children. In previous benefits analyses, EPA used a unit value of $59.31. This is the midpoint between a low estimate and a high estimate. The low estimate is the sum of the midrange values recommended by IEc (1994) for two symptoms believed to be associated with acute bronchitis: coughing and chest tightness. The high estimate was taken to be twice the value of a minor respiratory restricted activity day. For a more complete description of the derivation of this estimate, see Abt Associates (2000, p. 4-30).

The above unit value assumes that an episode of acute bronchitis lasts only one day. However, this is generally not the case. More typically, it can last for 6 or 7 days. A simple adjustment, then, would be to multiply the original unit value of $59.31 by 6 or 7. A second unit value of $356 (=$59.31 x 6) was therefore derived.

Finally, as noted above, the epidemiological study relating air pollution to the incidence of acute bronchitis referred to children specifically. The value of an avoided case should therefore be WTP to avoid a case in a child, which may be different from WTP to avoid a case in an adult. Recent work by Dickie and Ulery (2002) suggests, in fact, that parents are generally willing to pay about twice as much to

avoid sickness in their children as in themselves.[198] In one of several models they estimated, the natural logarithm of parents' WTP was related both to the number of symptom-days avoided and to whether it was their child or themselves at issue. Dickie and Ulery noted that "experiencing all of the symptoms [considered in their study – cough and phlegm, shortness of breath/wheezing, chest pain, and fever] for 7 days, or 28 symptom-days altogether, is roughly equivalent to a case of acute bronchitis ..." Using this model, and assuming that a case of acute bronchitis can be reasonably modeled as consisting of 28 symptom-days, we estimated parents' WTP to avoid a case of acute bronchitis in a child to be $374.[199] This is the third unit value available in BenMAP.

H.5.2 Upper Respiratory Symptoms (URS) in Children

In past benefits analyses, EPA based willingness to pay to avoid a day of URS on symptom-specific WTPs to avoid those symptoms identified as part of the URS complex of symptoms. Pope et al. (1991) defined a day of URS as consisting of one or more of the following symptoms: runny or stuffy nose; wet cough; and burning, aching, or red eyes. The three contingent valuation (CV) studies shown in Exhibit H-7 above have estimated WTP to avoid various morbidity symptoms that are either within the URS symptom complex defined by Pope et al., or are similar to those symptoms. The three individual symptoms that were identified as most closely matching those listed by Pope et al. for URS are cough, head/sinus congestion, and eye irritation, corresponding to "wet cough," "runny or stuffy nose," and "burning, aching or red eyes," respectively. A day of URS could consist of any one of the seven possible "symptom complexes" consisting of at least one of these three symptoms. The original unit value for URS was based on the assumption that each of these seven URS complexes is equally likely. This unit value for URS, $24.64, is just an average of the seven estimates of mean WTP for the different URS complexes.

The WTP estimates on which the first unit value is based were elicited from adults, whereas the health endpoint associated with air pollution in the epidemiological study is in children. As noted above, recent research by Dickie and Ulery (2002) suggests that parental WTP to avoid symptoms and illnesses in their children is about twice what it is to avoid those symptoms and illnesses in themselves. We therefore derived a second unit value of $49.28 (=2 x $24.64) from the first unit value.

A third unit value was derived by using Model 1, Table III in Dickie and Ulery (2002) (the same model used for acute bronchitis), assuming that a day of URS consists of 2 symptoms. As noted above, this model relates parental WTP to the number of symptom-days avoided and to whether it is the parent or the child at issue. The unit value derived from this model is $187.[200]

---

[198] This is, to our knowledge, the only estimate, based on empirical data, of parental WTP for their children versus themselves.

[199] The mean household income among participants in the Dickie and Ulery CV survey was slightly higher than the national average. We therefore adjusted all WTP estimates that resulted from their models downward slightly, using an income elasticity of WTP of 0.147, the average of the income elasticities estimated in the four models in the study. The adjustment factor thus derived was 0.9738.

[200] A WTP estimate elicited from parents concerning their WTP to avoid symptoms in their children may well include some calculation of lost earnings resulting from having to lose a day of work. Estimates from the Dickie and Ulery model therefore (appropriately) probably include not only their WTP to have their children avoid the pain and suffering associated with their illness, but also the opportunity cost of a parent having to stay home with a sick child.

H.5.3 Lower Respiratory Symptoms (LRS) in Children

The three unit values for LRS in children currently available in BenMAP follow the same pattern as those for URS in children. In past benefits analyses, EPA based willingness to pay to avoid a day of LRS on symptom-specific WTPs to avoid those symptoms identified as part of the LRS complex of symptoms. Schwartz et al. (1994) defined a day of LRS as consisting of at least two of the following symptoms: cough, chest tightness, coughing up phlegm, and wheeze. Of the symptoms for which WTP estimates are available (listed in Exhibit H-7), those that most closely match the symptoms listed by Schwartz et al. are coughing, chest tightness, coughing up phlegm, and wheeze. A day of LRS, as defined by Schwartz et al., could consist of any one of 11 possible combinations of at least two of these four symptoms. In the absence of any further information, each of the 11 possible "symptom clusters" was considered equally likely. The original unit value for LRS, $15.57, is just an average of the eleven estimates of mean WTP for the different LRS symptom clusters.

A second unit value is twice the original unit value, or $31.15, based on the evidence from Dickie and Ulery (2002) that parents are willing to pay about twice as much to avoid symptoms and illness in their children as in themselves. The third unit value is based on Model 1, Table III in Dickie and Ulery, assuming that, as for URS, a day of LRS consists of 2 symptoms. As noted above, this model relates parental WTP to the number of symptom-days avoided and to whether it is the parent or the child at issue. The unit value derived from this model is $187.

H.5.4 "Any of 19 Respiratory Symptoms"

The presence of "any of 19 acute respiratory symptoms" is a somewhat subjective health effect used by Krupnick et al. (1990). Moreover, not all 19 symptoms are listed in the Krupnick et al. study. It is therefore not clear exactly what symptoms were included in the study. Even if all 19 symptoms were known, it is unlikely that WTP estimates could be obtained for all of the symptoms. Finally, even if all 19 symptoms were known and WTP estimates could be obtained for all 19 symptoms, the assumption of additivity of WTPs becomes tenuous with such a large number of symptoms. The likelihood that all 19 symptoms would occur simultaneously, moreover, is very small.

Acute respiratory symptoms must be either upper respiratory symptoms or lower respiratory symptoms. In the absence of further knowledge about which of the two types of symptoms is more likely to occur among the "any of 19 acute respiratory symptoms," we assumed that they occur with equal probability. Because this health endpoint may also consist of combinations of symptoms, it was also assumed that there is some (smaller) probability that upper and lower respiratory symptoms occur together. To value avoidance of a day of "the presence of any of 19 acute respiratory symptoms" we therefore assumed that this health endpoint consists either of URS, or LRS, or both. We also assumed that it is as likely to be URS as LRS and that it is half as likely to be both together. That is, it was assumed that "the presence of any of 19 acute respiratory symptoms" is a day of URS with 40 percent probability, a day of LRS with 40 percent probability, and a day of both URS and LRS with 20 percent probability. Using the point estimates of WTP to avoid a day of URS and LRS derived above, the point estimate of WTP to avoid a day of "the presence of any of 19 acute respiratory symptoms" is:

$$(0.40)(\$24.64) + (0.40)(\$15.57) + (0.20)(\$24.64 + \$15.57) = \$24.12.$$

Because this health endpoint is only vaguely defined, and because of the lack of information on the relative frequencies of the different combinations of acute respiratory symptoms that might qualify as "any of 19 acute respiratory symptoms," the unit dollar value derived for this health endpoint must be considered only a rough approximation.

H.5.5 Work Loss Days (WLDs)

Work loss days are valued at a day's wage. BenMAP calculates county-specific median daily wages from county-specific annual wages by dividing by (52*5), on the theory that a worker's vacation days are valued at the same daily rate as work days.

H.5.6 Minor Restricted Activity Days (MRADs)

Two unit values are currently available in BenMAP for MRADs. No studies are reported to have estimated WTP to avoid a minor restricted activity day (MRAD). However, IEc (1993) derived an estimate of WTP to avoid a minor respiratory restricted activity day (MRRAD), using WTP estimates from Tolley et al. (1986) for avoiding a three-symptom combination of coughing, throat congestion, and sinusitis. This estimate of WTP to avoid a MRRAD, so defined, is $38.37 (1990 $). Although Ostro and Rothschild (1989) estimated the relationship between $PM_{2.5}$ and MRADs, rather than MRRADs (a component of MRADs), it is likely that most of the MRADs associated with exposure to $PM_{2.5}$ are in fact MRRADs. The original unit value, then, assumes that MRADs associated with PM exposure may be more specifically defined as MRRADs, and uses the estimate of mean WTP to avoid a MRRAD.

Any estimate of mean WTP to avoid a MRRAD (or any other type of restricted activity day other than WLD) will be somewhat arbitrary because the endpoint itself is not precisely defined. Many different combinations of symptoms could presumably result in some minor or less minor restriction in activity. Krupnick and Kopp (1988) argued that mild symptoms will not be sufficient to result in a MRRAD, so that WTP to avoid a MRRAD should exceed WTP to avoid any single mild symptom. A single severe symptom or a combination of symptoms could, however, be sufficient to restrict activity. Therefore WTP to avoid a MRRAD should, these authors argue, not necessarily exceed WTP to avoid a single severe symptom or a combination of symptoms. The "severity" of a symptom, however, is similarly not precisely defined; moreover, one level of severity of a symptom could induce restriction of activity for one individual while not doing so for another. The same is true for any particular combination of symptoms.

Given that there is inherently a substantial degree of arbitrariness in any point estimate of WTP to avoid a MRRAD (or other kinds of restricted activity days), the reasonable bounds on such an estimate must be considered. By definition, a MRRAD does not result in loss of work. WTP to avoid a MRRAD should therefore be less than WTP to avoid a WLD. At the other extreme, WTP to avoid a MRRAD should exceed WTP to avoid a single mild symptom. The highest IEc midrange estimate of WTP to avoid a single symptom is $20.03 (1999 $), for eye irritation. The point estimate of WTP to avoid a WLD in the benefit analysis is $83 (1990 $). If all the single symptoms evaluated by the studies are not severe, then the estimate of WTP to avoid a MRRAD should be somewhere between $16 and $83. Because the IEc estimate of $38 falls within this range (and acknowledging the degree of arbitrariness associated with any estimate within this range), the IEc estimate is used as the mean of a triangular distribution centered at $38, ranging from $16 to $61. Adjusting to 2000 $, this is a triangular distribution centered at $50.55, ranging from $21 to $80.

A second unit value is based on Model 1, Table III in Dickie and Ulery (2002). This model estimates the natural logarithm of parents' WTP to avoid symptoms as a linear function of the natural logarithm of the number of symptom-days avoided and whether or not the person avoiding the symptoms is the parent or the child. The unit value derived from this model, assuming that an MRAD consists of one day of 3 symptoms in an adult, is $98.

H.5.7 Asthma Exacerbation

Several respiratory symptoms in asthmatics or characterizations of an asthma episode have been associated with exposure to air pollutants. All of these can generally be taken as indications of an asthma exacerbation ("asthma attack") when they occur in an asthmatic. BenMAP therefore uses the same set of unit values for all of the variations of "asthma exacerbation" that appear in the epidemiological literature.

Two unit values are currently available in BenMAP for asthma exacerbation in adults, and three are currently available for asthma exacerbation in children. In past benefits analyses, EPA based willingness to pay to avoid an asthma exacerbation on four WTP estimates from Rowe and Chestnut (1986) for avoiding a "bad asthma day." The mean of the four average WTPs is $32 (1990 $), or $43 in 2000$. The uncertainty surrounding this estimate was characterized by a continuous uniform distribution on the range defined by the lowest and highest of the four average WTP estimates from Rowe and Chestnut, [$12, $54] in 1990$, or [$16, $71] in 2000 $. This unit value is available for both adults and children.

A second unit value for adults was derived by using Model 1, Table III in Dickie and Ulery (2002) (the same model used for acute bronchitis, LRS, and URS), assuming that an asthma exacerbation consists of 1 symptom-day. As noted above, this model relates parental WTP to the number of symptom-days avoided and to whether it is the parent or the child at issue. The unit value derived from this model for adults is $74.

Two additional unit values are available for children. One of these is twice the original unit value, or $86, based on the evidence from Dickie and Ulery (2002) that parents are willing to pay about twice as much to avoid symptoms and illness in their children as in themselves. The third unit value is based on Model 1, Table III in Dickie and Ulery (the same model used for asthma exacerbation in adults, only now with the "adult or child" variable set to 1 rather than 0). The unit value derived from this model is $156.

H.5.8 School Loss Days

There is currently one unit value available in BenMAP for school loss days, based on (1) the probability that, if a school child stays home from school, a parent will have to stay home from work to care for the child, and (2) the value of the parent's lost productivity. We first estimated the proportion of families with school-age children in which both parents work, and then valued a school loss day as the probability of a work loss day resulting from a school loss day (i.e., the proportion of households with school-age children in which both parents work) times a measure of lost wages.

From the U.S. Bureau of the Census (2002) we obtained (1) the numbers of single, married, and "other" (i.e., widowed, divorced, or separated) women with children in the workforce, and (2) the rates of participation in the workforce of single, married, and "other" women with children. From these two sets of statistics, we calculated a weighted average participation rate of 72.85 percent, as shown in Exhibit H-8.

Our estimated daily lost wage (if a mother must stay at home with a sick child) is based on the median weekly wage among women age 25 and older in 2000 (U.S. Bureau of the Census, 2002, Table 621). This median weekly wage is $551. Dividing by 5 gives an estimated median daily wage of $103. The expected loss in wages due to a day of school absence in which the mother would have to stay home with her child is estimated as the probability that the mother is in the workforce times the daily wage she would lose if she missed a day = 72.85% of $103, or $75. We currently have insufficient information to characterize the uncertainty surrounding this estimate.

Exhibit H-8. Women with Children: Number and Percent in the Labor Force, 2000, and Weighted Average Participation Rate

Category	Women in Labor Force (millions) [a]	Participation Rate (%) [a]	Implied Total Number in Population (in millions)	Implied Percent in Population	Population-Weighted Average Participation Rate
	(1)	*(2)*	*(3) = (1)/(2)*	*(4)*	*[=sum (2)*(4) over rows]*
Single	3.1	73.9%	4.19	11.84%	--
Married	18.2	70.6%	25.78	72.79%	--
Other [b]	4.5	82.7%	5.44	15.36%	--
Total	--	--	35.42	--	**72.85%**

[a] Source: U.S. Bureau of the Census (2002, Table 577).

[b] Widowed, divorced, or separated.

A unit value based on the approach described above is likely to understate the value of a school loss day in two ways. First, it omits WTP to avoid the symptoms/illness which resulted in the school absence. Second, it effectively gives zero value to school absences which do not result in a work loss day. The unit value of $75 is therefore considered an "interim" value until such time as alternative means of estimating this unit value become available.

Exhibit H-9. Unit Values Available for Acute Symptoms and Illnesses

Health Endpoint	Basis for Estimate [a]	Age Range		Unit Value	Distribution of Unit Value	Parameters of Distribution	
		min.	*max.*			*P1*	*P2*
Acute Bronchitis	WTP: 1 day illness, CV studies	0	17	$59	uniform	17.5099	101.107
	WTP: 6 day illness, CV studies	0	17	$356	uniform	105.059	606.639
	WTP: 28 symptom-days, Dickie and Ulery (2002)	0	17	$374	lognormal	5.9470	0.0907
Any of 19 Respiratory Symptoms	WTP: 1 day illness, CV studies	18	65	$24	uniform	0	48.2476
Minor Restricted Activity Days	WTP: 1 day, CV studies	18	99	$51	triangular	20.7114	80.3688
	WTP: 3 symptoms 1 day, Dickie and Ulery (2002).	18	99	$98	lognormal	4.60884	0.06486
Lower Respiratory Symptoms	WTP: 1 day, CV studies	0	17	$16	uniform	6.94334	24.4664
	WTP: 2 symptoms 1 day, Dickie and Ulery (2002).	0	17	$187	lognormal	5.2556	0.07048
	WTP: 2 x 1 day, CV studies	0	17	$31	uniform	13.8867	48.9327
School Loss Days	0	0	17	$75	none	N/A	N/A
Upper Respiratory Symptoms	WTP: 1 day, CV studies	0	17	$25	uniform	9.22265	43.1093
	WTP: 2 symptoms 1 day, Dickie and Ulery (2002)	0	17	$187	lognormal	5.2556	0.07048
	WTP: 2 x 1 day, CV studies	0	17	$49	uniform	18.4453	86.2186
Work Loss Days [b]	Median daily wage, county-specific	18	65	$115	none	N/A	N/A

[a] All unit values pulled from a lognormal distribution from Model 1, Table III in Dickie and Ulery (2002) are multiplied by 0.973811 to adjust for a difference in mean household income between the study participants and the general population. The unit values shown here have already been adjusted.

[b] Unit values for work loss days are county-specific, based on county-specific median wages. The unit value shown here is the national median daily wage, given for illustrative purposes only.

Exhibit H-10. Unit Values Available for Asthma-related Acute Symptoms and Illnesses

Health Endpoint	Basis for Estimate*	Age Range		Unit Value	Unit Value Distribution	Parameters of Distribution	
		min.	*max.*			*P1*	*P2*
Asthma Exacerbation, Asthma Attacks	WTP: bad asthma day, Rowe and Chestnut (1986)	18	99	$43	uniform	15.5599	70.8826
	WTP: 1 symptom-day, Dickie and Ulery (2002)	18	99	$74	lognormal	4.321	0.09569
	WTP: bad asthma day, Rowe and Chestnut (1986)	0	17	$43	uniform	15.5599	70.8826
	WTP: 2 x bad asthma day, Rowe and Chestnut (1986)	0	17	$86	uniform	31.1198	141.765
	WTP: 1 symptom-day, Dickie and Ulery (2002)	0	17	$156	lognormal	5.074	0.09253
Asthma Exacerbation, Cough	WTP: bad asthma day, Rowe and Chestnut (1986)	18	99	$43	uniform	15.5599	70.8826
	WTP: 1 symptom-day, Dickie and Ulery (2002)	18	99	$74	lognormal	4.321	0.09569
	WTP: bad asthma day, Rowe and Chestnut (1986)	0	17	$43	uniform	15.5599	70.8826
	WTP: 2 x bad asthma day, Rowe and Chestnut (1986)	0	17	$86	uniform	31.1198	141.765
	WTP: 1 symptom-day, Dickie and Ulery (2002)	0	17	$156	lognormal	5.074	0.09253
Asthma Exacerbation, Moderate or Worse	WTP: bad asthma day, Rowe and Chestnut (1986)	18	99	$43	uniform	15.5599	70.8826
	WTP: 1 symptom-day, Dickie and Ulery (2002)	18	99	$74	lognormal	4.321	0.09569
	WTP: bad asthma day, Rowe and Chestnut (1986)	0	17	$43	uniform	15.5599	70.8826
	WTP: 2 x bad asthma day, Rowe and Chestnut (1986)	0	17	$86	uniform	31.1198	141.765
	WTP: 1 symptom-day, Dickie and Ulery (2002)	0	17	$156	lognormal	5.074	0.09253
Asthma Exacerbation, One or More Symptoms	WTP: bad asthma day, Rowe and Chestnut (1986)	18	99	$43	uniform	15.5599	70.8826
	WTP: 1 symptom-day, Dickie and Ulery (2002)	18	99	$74	lognormal	4.321	0.09569
	WTP: bad asthma day, Rowe and Chestnut (1986)	0	17	$43	uniform	15.5599	70.8826
	WTP: 2 x bad asthma day, Rowe and Chestnut (1986)	0	17	$86	uniform	31.1198	141.765
	WTP: 1 symptom-day, Dickie and Ulery (2002)	0	17	$156	lognormal	5.074	0.09253
Asthma Exacerbation, Shortness of Breath	WTP: bad asthma day, Rowe and Chestnut (1986)	18	99	$43	uniform	15.5599	70.8826
	WTP: 1 symptom-day, Dickie and Ulery (2002)	18	99	$74	lognormal	4.321	0.09569
	WTP: bad asthma day, Rowe and Chestnut (1986)	0	17	$43	uniform	15.5599	70.8826
	WTP: 2 x bad asthma day, Rowe and Chestnut (1986)	0	17	$86	uniform	31.1198	141.765
	WTP: 1 symptom-day, Dickie and Ulery (2002)	0	17	$156	lognormal	5.074	0.09253
Asthma Exacerbation, Wheeze	WTP: bad asthma day, Rowe and Chestnut (1986)	18	99	$43	uniform	15.5599	70.8826
	WTP: 1 symptom-day, Dickie and Ulery (2002)	18	99	$74	lognormal	4.321	0.09569
	WTP: bad asthma day, Rowe and Chestnut (1986)	0	17	$43	uniform	15.5599	70.8826
	WTP: 2 x bad asthma day, Rowe and Chestnut (1986)	0	17	$86	uniform	31.1198	141.765
	WTP: 1 symptom-day, Dickie and Ulery (2002)	0	17	$156	lognormal	5.074	0.09253

*All unit values pulled from a lognormal distribution from Model 1, Table III in Dickie and Ulery, 2002, are multiplied by 0.973811 to adjust for a difference in mean household income between the study participants and the general population. The unit values shown here have already been adjusted.

Exhibit H-11. Unit Value Uncertainty Distributions and Their Parameters

Distribution [a]	Parameter 1 (P1)	Parameter 2 (P2)
Normal	standard deviation	–
Triangular	minimum value	maximum value
Lognormal [b]	mean of corresponding normal distribution	standard deviation of corresponding normal distribution
Uniform	minimum value	maximum vaue
Weibull [c]	α	β

[a] In all cases, BenMAP calculates the mean of the distribution, which is used as the "point estimate" of the unit value.

[b] If Y is a normal random variable, and $Y = \log_e X$, then X is lognormally distributed. Equivalently, X is lognormally distributed if $X = e^Y$, where Y is normally distributed.

[c] The Weibull distribution has the following probability density function:

$$\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-(x/\alpha)^{\beta}}$$

Appendix I: Uncertainty & Pooling

This Appendix discusses the treatment of uncertainty in BenMAP, both for incidence changes and associated dollar benefits. Some background is then given on pooling methodology. Finally, the mechanics of the various *Pooling Methods* available in BenMAP are discussed in detail, including *Subjective Weight* based pooling, *Fixed Effects* pooling, *Random / Fixed Effects* pooling, and independent and dependent *Sum* and *Subtraction*.

I.1 Uncertainty

Although there are several sources of uncertainty affecting estimates of incidence changes and associated benefits, the sources of uncertainty that are most readily quantifiable in benefits analyses are uncertainty surrounding the C-R relationships and uncertainty surrounding unit dollar values. The total dollar benefit associated with a given endpoint group depends on how much the endpoint group will change in the control scenario (e.g., how many premature deaths will be avoided) and how much each unit of change is worth (e.g., how much a statistical death avoided is worth).

Both the uncertainty about the incidence changes and uncertainty about unit dollar values can be characterized by *distributions*. Each "uncertainty distribution" characterizes our beliefs about what the true value of an unknown (e.g., the true change in incidence of a given health effect) is likely to be, based on the available information from relevant studies.[201] Unlike a sampling distribution (which describes the possible values that an *estimator* of an unknown value might take on), this uncertainty distribution describes our beliefs about what values the unknown value itself might be. Such uncertainty distributions can be constructed for each underlying unknown (such as a particular pollutant coefficient for a particular location) or for a function of several underlying unknowns (such as the total dollar benefit of a regulation). In either case, an uncertainty distribution is a characterization of our beliefs about what the unknown (or the function of unknowns) is likely to be, based on all the available relevant information. Uncertainty statements based on such distributions are typically expressed as 90 percent credible intervals. This is the interval from the fifth percentile point of the uncertainty distribution to the ninety-fifth percentile point. The 90 percent credible interval is a "credible range" within which, according to the available information (embodied in the uncertainty distribution of possible values), we believe the true value to lie with 90 percent probability. The uncertainty surrounding both incidence estimates and dollar benefits estimates can be characterized quantitatively in BenMAP. Each is described separately below.

I.1.1 Characterization of Uncertainty Surrounding Incidence Changes

To calculate point estimates of the changes in incidence of a given adverse health effect associated with a given set of air quality changes, BenMAP performs a series of calculations at each grid-cell. First, it accesses the C-R functions needed for the analysis, and then it accesses any data needed by the C-R functions. Typically, these include the grid-cell population, the change in population exposure at the grid-cell, and the appropriate baseline incidence rate. BenMAP then calculates the change in incidence of adverse health effects for each selected C-R function. This is described more fully in Chapter 5. The

[201] Although such an "uncertainty distribution" is not formally a Bayesian posterior distribution, it is very similar in concept and function (see, for example, the discussion of the Bayesian approach in Kennedy 1990, pp. 168-172).

resulting incidence change is stored, and BenMAP proceeds to the next grid-cell, where the above process is repeated.

In *Latin Hypercube* mode (see Chapter 5), BenMAP reflects the uncertainty surrounding estimated incidence changes (resulting from the sampling uncertainty surrounding the pollutant coefficients in the C-R functions used) by producing a *distribution* of possible incidence changes rather than a single point estimate. To do this, it uses the distribution (*Dist Beta*) associated with the pollutant coefficient (*Beta*, or β), and potentially the point estimate (*Beta*) and two parameters (*P1Beta*, *P2Beta*). Typically, pollutant coefficients are normally distributed, with mean *Beta* and standard deviation *P1Beta*. See Chapter 8 for more information on these C-R Function variables.

BenMAP uses an N-point Latin Hypercube[202] to represent the underlying distribution of β and to create a corresponding distribution of incidence changes in each population grid cell, where N is specified by you (as *Latin Hypercube Points* - see Chapter 5). The Latin Hypercube method represents an underlying distribution by N percentile points of the distribution, where the n^{th} percentile point is equal to:

$$(n-1)\cdot\frac{100}{N}+\frac{100}{2N}$$

Suppose, for example, that you elect to use a 20-point Latin Hypercube. BenMAP would then represent the distribution of β by 20 percentile points, specifically the 2.5^{th}, 7.5^{th}, ..., 97.5^{th}. To do this, the inverse cumulative distribution function specified by the distribution of β is called with the input probability equal to each the 20 percentile points. BenMAP then generates an estimate of the incidence change in a grid-cell for each of these values of β, resulting in a distribution of N incidence changes. This distribution is stored, and BenMAP proceeds to the next population grid-cell, where the process is repeated.

I.1.2 Characterization of Uncertainty Surrounding Dollar Benefits

The uncertainty distribution of the dollar benefits associated with a given health or welfare effect is derived from the two underlying uncertainty distributions – the distribution of the change in incidence of the effect (number of cases avoided) and the distribution of the value of a case avoided (the "unit value"). The derivation of the uncertainty distribution for incidence change is described above. The distributions used to characterize the uncertainty surrounding unit values are described in detail in Appendix H. As noted in that Appendix, a variety of distributions have been used to characterize the uncertainty of unit values, including uniform, triangular, normal, and Weibull.

To represent the underlying distribution of uncertainty surrounding unit values, a 100-point Latin Hypercube is generated in the same way described in the previous section for the distribution of β. That is, the unit value distribution is represented using the 0.5^{th}, 1.5^{th}, ..., and 99.5^{th} percentile values of its distribution.

[202]The Latin Hypercube method is used to enhance computer processing efficiency. It is a sampling method that divides a probability distribution into intervals of equal probability, with an assumption value for each interval assigned according to the interval's probability distribution. Compared with conventional Monte Carlo sampling, the Latin Hypercube approach is more precise over a fewer number of trials because the distribution is sampled in a more even, consistent manner (Decisioneering, 1996, pp. 104-105).

A distribution of the uncertainty surrounding the dollar benefits associated with a given endpoint is then derived from latin hypercube values generated to represent the change in incidence and the latin hypercube values generated to represent the unit value distribution. To derive this new distribution, each of the 100 unit values is multiplied by each of the N incidence change values, yielding a set of 100 * N dollar benefits. These values are sorted low to high and binned down to a final distribution of N dollar benefit values.

I.2 Pooling

There is often more than one study that has estimated a C-R function for a given pollutant-health endpoint combination. Each study provides an estimate of the pollutant coefficient, β, in the C-R function, along with a measure of the uncertainty of the estimate. Because uncertainty decreases as sample size increases, combining data sets is expected to yield more reliable estimates of β, and therefore more reliable estimates of the incidence change predicted using β. Combining data from several comparable studies in order to analyze them together is often referred to as meta-analysis.

For a number of reasons, including data confidentiality, it is often impractical or impossible to combine the original data sets. Combining the *results* of studies in order to produce better estimates of β provides a second-best but still valuable way to synthesize information (DerSimonian and Laird, 1986). This is referred to as pooling. Pooling β's requires that all of the studies contributing estimates of β use the same functional form for the concentration-response function. That is, the β's must be measuring the same thing.

It is also possible to pool the study-specific estimates of incidence change derived from the C-R functions, instead of pooling the underlying β's themselves. For a variety of reasons, this is often possible when it is not feasible to pool the underlying β's. For example, if one study is log-linear and another is linear, we could not pool the β's because they are not different estimates of a coefficient in the same C-R function, but are instead estimates of coefficients in different C-R functions. We can, however, calculate the incidence change predicted by each C-R function (for a given change in pollutant concentration and, for the log-linear function, a given baseline incidence rate), and pool these incidence changes. BenMAP allows the pooling of incidence changes predicted by several studies for the same pollutant-health endpoint group combination. It also allows the pooling of the corresponding study-specific estimates of monetary benefits.

As with estimates based on only a single study, BenMAP allows you to characterize the uncertainty surrounding pooled estimates of incidence change and/or monetary benefit. To do this, BenMAP pools the study-specific distributions of incidence changes (or monetary benefit) to derive a pooled distribution. This pooled distribution incorporates information from all the studies used in the pooling procedure.

I.2.1 Weights Used for Pooling

The relative contribution of any one study in the pooling process depends on the weight assigned to that study. A key component of the pooling process, then, is the determination of the weight given to each study. There are various methods that can be used to assign weights to studies (these are three of the *Pooling Methods* - see Chapter 6 for more information). Below we discuss the possible weighting schemes that are available in BenMAP.

Subjective (User-specified) Weights

BenMAP allows you the option of specifying the weights to be used. Suppose, for example, you want to simply average all study-specific results. You would then assign a weight of 1/N to each of the N study-specific distributions that are to be pooled. Note that subjective weights are limited to two decimal places, and are normalized if they do not sum to one.

Automatically Generated Weights

A simple average has the advantage of simplicity but the disadvantage of not taking into account the uncertainty of each of the estimates. Estimates with great uncertainty surrounding them are given the same weight as estimates with very little uncertainty. A common method for weighting estimates involves using their variances. Variance takes into account both the consistency of data and the sample size used to obtain the estimate, two key factors that influence the reliability of results. BenMAP has two methods of automatically generating pooling weights using the variances of the input distributions - *Fixed Effects Pooling* and *Random / Fixed Effects Pooling*.

The discussion of these two weighting schemes is first presented in terms of pooling the pollutant coefficients (the β's), because that most closely matches the discussion of the method for pooling study results as it was originally presented by DerSimonian and Laird (1986). We then give an overview of the analogous weighting process used within BenMAP to generate weights for incidence changes rather than β's.

Fixed Effects Weights

The fixed effects model assumes that there is a single true concentration-response relationship and therefore a single true value for the parameter β that applies everywhere. Differences among β's reported by different studies are therefore simply the result of sampling error. That is, each reported β is an estimate of the *same underlying parameter*. The certainty of an estimate is reflected in its variance (the larger the variance, the less certain the estimate). Fixed effects pooling therefore weights each estimate under consideration in proportion to the *inverse* of its variance.

Suppose there are n studies, with the ith study providing an estimate β_i with variance v_i (I = 1, ..., n). Let

$$S = \Sigma \frac{1}{v_i},$$

denote the sum of the inverse variances. Then the weight, w_i, given to the ith estimate, β_i, is

$$w_i = \frac{1/v_i}{S} .$$

This means that estimates with small variances (i.e., estimates with relatively little uncertainty surrounding them) receive large weights, and those with large variances receive small weights.

The estimate produced by pooling based on a fixed effects model, then, is just a weighted average of the estimates from the studies being considered, with the weights as defined above. That is,

$$\beta_{fe} = \Sigma\ w_i * \beta_i .$$

The variance associated with this pooled estimate is the inverse of the sum of the inverse variances:

$$v_{fe} = \frac{1}{\Sigma\ 1/v_i} .$$

Exhibit I-2 shows the relevant calculations for this pooling for three sample studies.

Exhibit I-2. Example of Fixed Effects Model Calculations

Study	β_i	v_i	$1/v_i$	w_i	$w_i*\beta_i$
1	0.75	0.1225	8.16	0.016	0.012
2	1.25	0.0025	400	0.787	0.984
3	1.00	0.0100	100	0.197	0.197
Sum			$\sum$ = 508.16	$\sum$ = 1.000	$\sum$ = 1.193

The sum of weighted contributions in the last column is the pooled estimate of β based on the fixed effects model. This estimate (1.193) is considerably closer to the estimate from study 2 (1.25) than is the estimate (1.0) that simply averages the study estimates. This reflects the fact that the estimate from study 2 has a much smaller variance than the estimates from the other two studies and is therefore more heavily weighted in the pooling.

The variance of the pooled estimate, v_{fe}, is the inverse of the sum of the variances, or 0.00197. (The sums of the β_i and v_i are not shown, since they are of no importance. The sum of the $1/v_i$ is S, used to calculate the weights. The sum of the weights, w_i , i=1, ..., n, is 1.0, as expected.)

Random / Fixed Effects Weights

An alternative to the fixed effects model is the random effects model, which allows the possibility that the estimates β_i from the different studies may in fact be estimates of *different* parameters, rather than just different estimates of a single underlying parameter. In studies of the effects of PM_{10} on mortality, for example, if the composition of PM_{10} varies among study locations the underlying relationship between mortality and PM_{10} may be different from one study location to another. For example, fine particles make up a greater fraction of PM_{10} in Philadelphia than in El Paso. If fine particles are disproportionately responsible for mortality relative to coarse particles, then one would expect the true value of β in Philadelphia to be greater than the true value of β in El Paso. This would violate the assumption of the fixed effects model.

The following procedure can test whether it is appropriate to base the pooling on the random effects model (vs. the fixed effects model):

A test statistic, Q_w , the weighted sum of squared differences of the separate study estimates from the pooled estimate based on the fixed effects model, is calculated as:

$$Q_w = \Sigma_i \frac{1}{v_i} (\beta_{fe} - \beta_i)^2.$$

Under the null hypothesis that there is a single underlying parameter, β, of which all the β_i 's are estimates, Q_w has a chi-squared distribution with n-1 degrees of freedom. (Recall that n is the number of studies in the meta-analysis.) If Q_w is greater than the critical value corresponding to the desired confidence level, the null hypothesis is rejected. That is, in this case the evidence does not support the fixed effects model, and the random effects model is assumed, allowing the possibility that each study is estimating a different β. (BenMAP uses a five percent one-tailed test).

The weights used in a pooling based on the random effects model must take into account not only the within-study variances (used in a meta-analysis based on the fixed effects model) but the between-study variance as well. These weights are calculated as follows:

Using Q_w , the between-study variance, η^2, is:

$$\eta^2 = \frac{Q_w - (n-1)}{\Sigma 1/v_i - \frac{\Sigma 1/v_i^2}{\Sigma 1/v_i}} .$$

It can be shown that the denominator is always positive. Therefore, if the numerator is negative (i.e., if $Q_w < n-1$), then η^2 is a negative number, and it is not possible to calculate a random effects estimate. In this case, however, the small value of Q_w would presumably have led to accepting the null hypothesis described above, and the meta-analysis would be based on the fixed effects model. The remaining discussion therefore assumes that η^2 is positive.

Given a value for η^2 , the random effects estimate is calculated in almost the same way as the fixed effects estimate. However, the weights now incorporate both the within-study variance (v_i) and the between-study variance (η^2). Whereas the weights implied by the fixed effects model used only v_i, the within-study variance, the weights implied by the random effects model use $v_i + \eta^2$.

Let $v_i^* = v_i + \eta^2$. Then

$$S^* = \Sigma \frac{1}{v_i^*} ,$$

and

$$w_i^* = \frac{1/v_i^*}{S^*} .$$

The estimate produced by pooling based on the random effects model, then, is just a weighted average of the estimates from the studies being considered, with the weights as defined above. That is,

$$\beta_{rand} = \Sigma\ w_i^* * \beta_i \ .$$

The variance associated with this random effects pooled estimate is, as it was for the fixed effects pooled estimate, the inverse of the sum of the inverse variances:

$$v_{rand} = \frac{1}{\Sigma\ 1/v_i^*} \ .$$

The weighting scheme used in a pooling based on the random effects model is basically the same as that used if a fixed effects model is assumed, but the variances used in the calculations are different. This is because a fixed effects model assumes that the variability among the estimates from different studies is due only to sampling error (i.e., each study is thought of as representing just another sample from the same underlying population), while the random effects model assumes that there is not only sampling error associated with each study, but that there is also *between-study* variability -- each study is estimating a different underlying β. Therefore, the sum of the within-study variance and the between-study variance yields an overall variance estimate.

Fixed Effects and Random / Fixed Effects Weighting to Pool Incidence Change Distributions and Dollar Benefit Distributions

Weights can be derived for pooling incidence changes predicted by different studies, using either the fixed effects or the fixed / random effects model, in a way that is analogous to the derivation of weights for pooling the β's in the C-R functions. As described above, BenMAP generates a latin hypercube representation of the distribution of incidence change corresponding to each C-R Function selected. The means of those study-specific latin hypercube distributions of incidence change are used in exactly the same way as the reported β's are used in the calculation of fixed effects and random effects weights described above. The variances of incidence change are used in the same way as the variances of the β's. The formulas above for calculating fixed effects weights, for testing the fixed effects hypothesis, and for calculating random effects weights can all be used by substituting the mean incidence change for the ith C-R Function for β_i and the variance of incidence change for the ith C-R Function for v_i.[203]

Similarly, weights can be derived for dollar benefit distributions. As described above, BenMAP generates a latin hypercube representation of the distribution of dollar benefits . The means of those latin hypercube distributions are used in exactly the same way as the reported β's are used in the calculation of fixed effects and random effects weights described above. The variances of dollar benefits are used in the same way as the variances of the β's. The formulas above for calculating fixed effects weights, for testing

[203] There may be a problem with transferring the fixed effects hypothesis test to "incidence change space." The test statistic to test the fixed effects model is a chi-squared random variable. In the original paper on this pooling method, DerSimonian and Laird, 1986, were discussing the pooling of estimates of parameters, which are generally normally distributed. The incidence changes predicted from a C-R function will not be normally distributed if the C-R function is not a linear function of the pollutant coefficient, which, in most cases it is not. (Most C-R functions are log-linear.) In that case, the test statistic may not be chi-square distributed. However, most log-linear C-R functions are *nearly* linear because their coefficients are very small. In that case the test statistic is likely to be *nearly* chi-square distributed.

the fixed effects hypothesis, and for calculating random effects weights can all be used by substituting the mean dollar benefit change for the ith valuation for β_i and the variance of dollar benefits for the ith valuation for v_i.

BenMAP always derives *Fixed Effects* and *Random / Fixed Effects* weights using nationally aggregated results, and uses those weights for pooling at each grid cell (or county, etc. if you choose to aggregate results prior to pooling). This is done because BenMAP does not include any regionally based uncertainty – that is, all uncertainty is at the national level in BenMAP, and all regional differences (population, for example) are treated as certain.

I.2.2 The Mechanics of Pooling in BenMAP

Once weights are generated for each input distribution, BenMAP has three options for using these weights to combine the input distributions into a single new distribution. These options are referred to as *Advanced Pooling Methods* (see Chapter 6 for more details).

Round Weights to Two Digits

This is BenMAP's default *Advanced Pooling Method*, and is always the method used when *Subjective Weights* are used. The first step is converting the weights to two digit integers by multiplying them by 100 and rounding to the nearest integer. If all the integral weights thus generated are divisible by the smallest weight, they are each divided by that smallest weight. For example, if the original weights were 0.1, 0.2, 0.3, and 0.4, the resulting integral weights would be 10/10, 20/10, 30/10, and 40/10 (or 1, 2, 3, and 4).

BenMAP then creates a new distribution by sampling each entire input distribution according to its weight. That is, in the above example the first distribution would be sampled once, the second distribution twice, and so forth. The advantage of sampling whole distributions is that it preserves the characteristics (i.e., the moments - the mean, the variance, etc.) of the underlying distributions. Assuming *n* latin hypercube points, the resulting distribution will contain a maximum of 100 * *n* values, which are then sorted low to high and binned down to *n* values, which will represent the new, pooled distribution.

Round Weights to Three Digits

This *Advanced Pooling Method* is essentially the same as rounding weights to two digits, except that the weights are converted to three digit integers, and so forth. That is, the weights are multiplied by 1000 and rounded to the nearest integer. Again, if all the integral weights thus generated are divisible by the smallest weight, they are each divided by that smallest weight. Assuming *n* latin hypercube points, the resulting distribution with this *Advanced Pooling Method* can contain a maximum of 1000 * *n* values, which are sorted low to high and binned down to *n* values, which represent the new, pooled distribution.

Exact Weights for Monte Carlo

This *Advanced Pooling Method* uses a Monte Carlo method to combine the input distributions. Using this method, on each of many iterations, (1) an input distribution is selected (with the probability of selection equal to the weight assigned to the distribution), and (2) a value is randomly drawn from that distribution. Values chosen in this way are placed into a temporary pooled distribution, which will have one point per iteration of the Monte Carlo method. The number of iterations is specified by the user (see

Chapter 6), and defaults to 5,000. After the temporary distribution is fully generated, it is sorted low to high and binned down to *n* values (where *n* is the number of *Latin Hypercube Points* chosen for the analysis - see Chapter 5).

I.2.3 Summing Distributions

Sometimes rather than pooling distributions we want to add them. For example, some studies have estimated a C-R function for hospital admissions for COPD and another C-R function for hospital admissions for pneumonia. From each of these C-R functions, BenMAP can derive the corresponding distributions for incidence change. Hospital admissions for COPD and pneumonia are two of the most important components of respiratory hospital admissions, and we may want to estimate the number of cases of "respiratory hospital admissions," as characterized by being either COPD or pneumonia. To do this we would add the two distributions.

Summing across distributions can be done in one of two ways: We can assume the two distributions are independent of each other or dependent. Which is the more reasonable assumption depends on the particulars of the distributions being summed.

Assuming Independence

This is the *Sum (Independent) Pooling Method* (see Chapter 6 for details). To sum two distributions that are independent, on each of many iterations of a Monte Carlo procedure, BenMAP (1) randomly selects a value from the first input distribution, (2) randomly selects a value from the second input distribution, and (3) adds the two values together. To sum N distributions that are independent, BenMAP follows an analogous procedure in which, on each iteration it makes a random selection from each of the input distributions and then adds the results together. When the Monte Carlo procedure is completed, all such generated results are sorted low to high and binned down to the appropriate number of latin hypercube points. The number of iterations is determined by the *Monte Carlo Iterations* setting (see Chapter 6).

Assuming Dependence

This is the *Sum (Dependent) Pooling Method* (see Chapter 6 for details). Recall that the uncertainty distributions in BenMAP are latin hypercube representations, consisting of N percentile points. To sum two distributions assumed to be dependent, BenMAP simply generates a new N point latin hypercube where each point is the sum of the corresponding points from the input latin hypercubes. That is, the first point in the new latin hypercube is the sum of the first points in the two input latin hypercubes, and so forth. To sum *n* distributions that are assumed to be dependent, BenMAP follows an analogous procedure in which each point in the new latin hypercube is the sum of the corresponding points from each of the input latin hypercubes.

I.2.4 Subtracting Distributions

In some cases, you may want to subtract one or more distribution(s) from another. For example, one study may have estimated a C-R function for minor restricted activity days (MRADs), and another study may have estimated a C-R function for asthma "episodes." You may want to subtract the change in

incidence of asthma episodes from the change in incidence from MRADs before estimating the monetary value of the MRADs, so that the monetary value of asthma episodes avoided will not be included.

Subtracting across distributions can be done in one of two ways: we can assume the two distributions are independent of each other or dependent. Which is the more reasonable assumption depends on the particulars of the distributions being subtracted.

Assuming Independence

This is the *Subtraction (Independent) Pooling Method*. To subtract one distribution from another, assuming independence, on each of many iterations of a Monte Carlo procedure, BenMAP (1) randomly selects a value from the first input distribution, (2) randomly selects a value from the second input distribution, and (3) subtracts the second value from the first. To subtract N distributions from another distribution, assuming independence, BenMAP follows an analogous procedure in which, on each iteration it makes a random selection from each of the input distributions and then subtracts the second through the Nth from the first. When the Monte Carlo procedure is completed, all such generated results are sorted low to high and binned down to the appropriate number of latin hypercube points. The number of iterations is determined by the *Monte Carlo Iterations* setting (see Chapter 6).

Assuming Dependence

This is the *Subtraction (Dependent) Pooling Method* (see Chapter 6 for details). Recall that the uncertainty distributions in BenMAP are latin hypercube representations, consisting of N percentile points. To subtract one distribution from another, assuming them to be dependent, BenMAP simply generates a new N point latin hypercube where each point is the result of subtracting the corresponding point of the second input latin hypercube from the corresponding point of the first input latin hypercube. That is, the first point in the new latin hypercube is the result of subtracting the first point in the second latin hypercube from the first point of the first latin hypercube, and so forth. To subtract *n* distributions from another distribution, assuming dependence, BenMAP follows an analogous procedure in which each point in the new latin hypercube is the result of subtracting the corresponding points of the second through the Nth input latin hypercubes from the corresponding point of the first.

References

Abbey, D.E., F. Petersen, P.K. Mills and W.L. Beeson. 1993. Long-Term Ambient Concentrations of Total Suspended Particulates, Ozone, and Sulfur Dioxide and Respiratory Symptoms in a Nonsmoking Population. *Archives of Environmental Health*. Vol. 48(1): 33-46.

Abbey, D.E., B.L. Hwang, R.J. Burchette, T. Vancuren and P.K. Mills. 1995a. Estimated long-term ambient concentrations of PM10 and development of respiratory symptoms in a nonsmoking population. *Arch Environ Health*. Vol. 50(2): 139-52.

Abbey, D.E., B.E. Ostro, G. Fraser, T. Vancuren and R.J. Burchette. 1995b. Estimating Fine Particulates Less Than 2.5 Microns in Aerodynamic Diameter (Pm2.5) From Airport Visibility Data in California. *J Expo Anal Environ Epidemiol*. Vol. 5(2): 161-180.

Abbey, D.E., B.E. Ostro, F. Petersen and R.J. Burchette. 1995c. Chronic Respiratory Symptoms Associated with Estimated Long-Term Ambient Concentrations of Fine Particulates Less Than 2.5 Microns in Aerodynamic Diameter (PM2.5) and Other Air Pollutants. *J Expo Anal Environ Epidemiol*. Vol. 5(2): 137-159.

Abbey, D.E., R.J. Burchette, S.F. Knutsen, W.F. McDonnell, M.D. Lebowitz and P.L. Enright. 1998. Long-term particulate and other air pollutants and lung function in nonsmokers. *Am J Respir Crit Care Med*. Vol. 158(1): 289-98.

Abt Associates Inc. 1996a. An Analysis of the Monetized Benefits Associated with National Attainment of Alternative Particulate Matter Standards in the Year 2007. Prepared for U.S. EPA, Office of Air Quality Planning and Standards. Research Triangle Park, NC. July 5.

Abt Associates Inc. 1996b. A Particulate Matter Risk Assessment for Philadelphia and Los Angeles. Prepared for U.S. EPA, Office of Air Quality Planning and Standards. Research Triangle Park, NC. July 3.

Abt Associates Inc. 2000. Final Heavy Duty Engine/Diesel Fuel Rule: Air Quality Estimation, Selected Health and Welfare Benefits Methods, and Benefit Analysis Results. Prepared for U.S. EPA, Office of Air Quality Planning and Standards, Research Triangle Park, NC. Bethesda, MD. December.

Ackermann-Liebrich, U., P. Leuenberger, J. Schwartz, C. Schindler, C. Monn, C. Bolognini, J.P. Bongard, O. Brandli, G. Domenighetti, S. Elsasser, L. Grize, W. Karrer, R. Keller, H. KellerWossidlo, N. Kunzli, B.W. Martin, T.C. Medici, A.P. Perruchoud, M.H. Schoni, J.M. Tschopp, B. Villiger, B. Wuthrich, J.P. Zellweger and E. Zemp. 1997. Lung function and long term exposure to air pollutants in Switzerland. Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) Team. *Am J Respir Crit Care Med*. Vol. 155(1): 122-129.

Adams, P.F., G.E. Hendershot and M.A. Marano. 1999. Current Estimates from the National Health Interview Survey, 1996. *Vital Health Stat*. Vol. 10(200): 1-212.

Agency for Healthcare Research and Quality. 2000. HCUPnet, Healthcare Cost and Utilization Project.

References

American Lung Association. 2002a. Trends in Morbidity and Mortality: Pneumonia, Influenza, and Acute Respiratory Conditions. American Lung Association, Best Practices and Program Services, Epidemiology and Statistics Unit.

American Lung Association. 2002b. Trends in Chronic Bronchitis and Emphysema: Morbidity and Mortality. American Lung Association, Best Practices and Program Services, Epidemiology and Statistics Unit.

American Lung Association. 2002c. Trends in Asthma Morbidity and Mortality. American Lung Association, Best Practices and Program Services, Epidemiology and Statistics Unit.

Blumenschein, K. and M. Johannesson. 1998. Relationship between quality of life instruments, health state utilities, and willingness to pay in patients with asthma. *Ann Allergy Asthma Immunol*. Vol. 80(2): 189-94.

Bobak, M. and D.A. Leon. 1992. Air pollution and infant mortality in the Czech Republic, 1986-88. *Lancet*. Vol. 340(8826): 1010-4.

Brunekreef, B., P.L. Kinney, J.H. Ware, D. Dockery, F.E. Speizer, J.D. Spengler and B.G. Ferris. 1991. Sensitive Subgroups and Normal Variation in Pulmonary Function Response to Air Pollution Episodes. *Environmental Health Perspectives*. Vol. 90(JAN): 189-193.

Burnett, R.T., S. Cakmak, J.R. Brook and D. Krewski. 1997. The role of particulate size and chemistry in the association between summertime ambient air pollution and hospitalization for cardiorespiratory diseases. *Environ Health Perspect*. Vol. 105(6): 614-20.

Burnett, R.T., M. Smith-Doiron, D. Stieb, S. Cakmak and J.R. Brook. 1999. Effects of particulate and gaseous air pollution on cardiorespiratory hospitalizations. *Archives Environmental Health*. Vol. 54(2): 130-139.

Burnett, R.T., M. Smith-Doiron, D. Stieb, M.E. Raizenne, J.R. Brook, R.E. Dales, J.A. Leech, S. Cakmak and D. Krewski. 2001. Association between ozone and hospitalization for acute respiratory diseases in children less than 2 years of age. *Am J Epidemiol*. Vol. 153(5): 444-52.

CARB (California Air Resources Board). 1982. California Ambient Air Quality Standard for Particulate Matter. Sacramento, CA. December.

Chen, L., B.L. Jennison, W. Yang and S.T. Omaye. 2000. Elementary school absenteeism and air pollution. *Inhal Toxicol*. Vol. 12(11): 997-1016.

Cody, R.P., C.P. Weisel, G. Birnbaum and P.J. Lioy. 1992. The effect of ozone associated with summertime photochemical smog on the frequency of asthma visits to hospital emergency departments. *Environ Res*. Vol. 58(2): 184-94.

Collet, D. 1994. Modelling Survival Data in Medical Research. Chapman & Hall: New York.

Crocker, T.D. and R.L. Horst, Jr. 1981. Hours of Work, Labor Productivity, and Environmental Conditions: A Case Study. *The Review of Economics and Statistics*. Vol. 63: 361-368.

Cropper, M.L. and A.J. Krupnick. 1990. The Social Costs of Chronic Heart and Lung Disease. Resources for the Future. Washington, DC. Discussion Paper QE 89-16-REV.

Cummings, R., H. Burness and R. Norton. 1985. Methods Development for Environmental Control Benefits Assessment, Volume V. Measuring Household Soiling Damages from Suspended Air Particulates, A Methodological Inquiry. Prepared for U.S. Environmental Protection Agency. Washington, DC.

Cunningham, J., D.W. Dockery, D.R. Gold and F.E. Speizer. 1995. Racial Differences in the Association Between Maternal Smoking During Pregnancy and Lung Function in Children. *American Journal of Respiratory and Critical Care Medicine*. Vol. 152(2): 565-569.

Daniels, M.J., F. Dominici, J.M. Samet and S.L. Zeger. 2000. Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities [see comments]. *Am J Epidemiol*. Vol. 152(5): 397-406.

Decisioneering. 1996. Crystal Ball: Forecasting and Risk Analysis for Spreadsheet Users: User Manual. www.decisioneering .com.

DerSimonian, R. and N. Laird. 1986. Meta-Analysis in Clinical Trials. *Controlled Clinical Trials*. Vol. 7: 177-188.

Detels, R., D.P. Tashkin, J.W. Sayre, S.N. Rokaw, F.J. Massey, A.H. Coulson and D.H. Wegman. 1991. The Ucla Population Studies of Cord .10. a Cohort Study of Changes in Respiratory Function Associated With Chronic Exposure to Sox, Nox, and Hydrocarbons. *American Journal of Public Health*. Vol. 81(3): 350-359.

Dickie, M. and S. Gerking. 1987. Reconciling Averting Behavior and Contingent Valuation Benefit Estimates of Reducing Symptoms of Ozone Exposure (draft), as cited in Neumann, J.E., M. Dickie, and R.E. Unsworth. 1994. Prepared by Industrial Economics. Prepared for Jim DeMocker, U.S. EPA, Office of Air and Radiation. March 31.

Dickie, M. and V.L. Ulery. 2002. Parental Altruism and the Value of Avoiding Acute Illness: Are Kids Worth More Than Parents? (Paper to be submitted for publication. Presented at Association of Environmental and Resource Economists 2001 Workshop, "Assessing and Managing Environmental and Public Health Risks."). December.

Dockery, D.W., F.E. Speizer, D.O. Stram, J.H. Ware, J.D. Spengler and B.G. Ferris, Jr. 1989. Effects of Inhalable Particles on Respiratory Health of Children. *Am Rev Respir Dis*. Vol. 139: 587-594.

Dockery, D.W., C.A. Pope, X.P. Xu, J.D. Spengler, J.H. Ware, M.E. Fay, B.G. Ferris and F.E. Speizer. 1993. An association between air pollution and mortality in six U.S. cities. *N Engl J Med*. Vol. 329(24): 1753-1759.

Dockery, D.W., J. Cunningham, A.I. Damokosh, L.M. Neas, J.D. Spengler, P. Koutrakis, J.H. Ware, M. Raizenne and F.E. Speizer. 1996. Health Effects of Acid Aerosols On North American Children - Respiratory Symptoms. *Environmental Health Perspectives*. Vol. 104(5): 500-505.

Eisenstein, E.L., L.K. Shaw, K.J. Anstrom, C.L. Nelson, Z. Hakim, V. Hasselblad and D.B. Mark. 2001. Assessing the clinical and economic burden of coronary artery disease: 1986-1998. *Med Care*. Vol. 39(8): 824-35.

Empire State Electric Energy Research Corporation (ESEERCO). 1994. New York State Environmental Externalities Cost Study. Report 2: Methodology. Prepared by RCG/Hagler, Bailly, Inc. November.

Fairley, D. 1999. Daily mortality and air pollution in Santa Clara County, California: 1989-1996. *Environ Health Perspect*. Vol. 107(8): 637-41.

Fairley, D. 2003. Mortality and Air Pollution for Santa Clara County, California, 1989-1996. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May. pp. 97-106.

GeoLytics Inc. 2001a. Geolytics CensusCD® 1990 Blocks, Release 1.1. CD-ROM. GeoLytics, Inc. East Brunswick, NJ. September.

GeoLytics Inc. 2001b. CensusCD® 1990 + Maps, Release 4.1. CD-ROM. GeoLytics, Inc. East Brunswick, NJ.

GeoLytics Inc. 2002. Geolytics CensusCD® 2000 Short Form Blocks. CD-ROM. GeoLytics, Inc. Release 1.0.

Gilliland, F.D., K. Berhane, E.B. Rappaport, D.C. Thomas, E. Avol, W.J. Gauderman, S.J. London, H.G. Margolis, R. McConnell, K.T. Islam and J.M. Peters. 2001. The effects of ambient air pollution on school absenteeism due to respiratory illnesses. *Epidemiology*. Vol. 12(1): 43-54.

Greenbaum, D. 2002. Letter to colleagues dated May 30, 2002. [Available at www.healtheffects.org]. Letter from L.D. Grant, Ph.D. to Dr. P. Hopke re: external review of EPA's Air Quality Criteria for Particulate Matter, with copy of 05/30/02 letter from Health Effects Institute re: re-analysis of National Morbidity, Mortality and Air Pollution Study data attached. Docket No. A-2000-01. Document No. IV-A-145.

Greene, W.H. 1997. Econometric Analysis. Prentice Hall: Upper Saddle River, NJ.

Haase, N., American Heart Association. 2002. Phone conversation. October.

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

Horst, R. and M. Duff. 1995. Concentration Data Transformation and the Quadratic Rollback Methodology (Round 2, Revised). Unpublished memorandum to R. Rodriguez, U.S. EPA. June 8.

Industrial Economics Incorporated (IEc). 1993. Memorandum to Jim DeMocker, U.S. Environmental Protection Agency, Office of Air and Radiation, Office of Policy Analysis and Review. September 30.

Industrial Economics Incorporated (IEc). 1994. Linkage Between Health Effects Estimation and Morbidity Valuation in the Section 812 Analysis -- Draft Valuation Document. Memorandum to Jim DeMocker, U.S. Environmental Protection Agency, Office of Air and Radiation, Office of Policy Analysis and Review. Prepared by J.E. Neumann, M.T. Dickie, and R.E. Unsworth. March 31.

Ito, K. and G.D. Thurston. 1996. Daily PM10/mortality associations: an investigations of at-risk subpopulations. *Journal of Exposure Analysis and Environmental Epidemiology*. Vol. 6(1): 79-95.

Ito, K. 2003. Associations of Particulate Matter Components with Daily Mortality and Morbidity in Detroit, Michigan. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May. pp. 143-156.

Jaffe, D.H., M.E. Singer and A.A. Rimm. 2003. Air pollution and emergency department visits for asthma among Ohio Medicaid recipients, 1991-1996. *Environ Res*. Vol. 91(1): 21-8.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl and T.-C. Lee. 1985. The Theory and Practice of Econometrics. 2nd ed. John Wiley and Sons: New York.

Kennedy. 1990. A Guide to Econometrics. 2nd ed. MIT Press: Cambridge, MA.

Kinney, P.L., K. Ito and G.D. Thurston. 1995. A Sensitivity Analysis of Mortality Pm-10 Associations in Los Angeles. *Inhalation Toxicology*. Vol. 7(1): 59-69.

Klemm, R.J., R.M. Mason, Jr., C.M. Heilig, L.M. Neas and D.W. Dockery. 2000. Is daily mortality associated specifically with fine particles? Data reconstruction and replication of analyses. *J Air Waste Manag Assoc*. Vol. 50(7): 1215-22.

Klemm, R.J. and R. Mason. 2003. Replication of Reanalysis of Harvard Six-City Study. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May. pp. 165-172.

Krewski, D., R. Burnett, M. Goldberg, K. Hoover, J. Siemiatycki, M. Jerrett, M. Abrahamowicz and M. White. 2000. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Health Effects Institute. Cambridge. July. pp. 295.

Krupnick, A.J. and R.J. Kopp. 1988. The Health and Agricultural Benefits of Reductions in Ambient Ozone in the United States. Resources for the Future. Washington, DC. Discussion Paper QE88-10. August.

Krupnick, A.J., W. Harrington and B. Ostro. 1990. Ambient Ozone and Acute Health Effects - Evidence From Daily Data. *Journal of Environmental Economics and Management*. Vol. 18(1): 1-18.

Krupnick, A.J. and M.L. Cropper. 1992. The Effect of Information On Health Risk Valuations. *Journal of Risk and Uncertainty*. Vol. 5(1): 29-48.

Lin, M., Y. Chen, R.T. Burnett, P.J. Villeneuve and D. Krewski. 2002. The influence of ambient coarse particulate matter on asthma hospitalization in children: case-crossover and time-series analyses. *Environ Health Perspect*. Vol. 110(6): 575-81.

Lippmann, M., K. Ito, A. Nádas and R. Burnett. 2000. Association of Particulate Matter Components with Daily Mortality and Morbidity in Urban Populations. Health Effects Institute. Number 95. August.

Loomis, D., M. Castillejos, D.R. Gold, W. McDonnell and V.H. Borja-Aburto. 1999. Air pollution and infant mortality in Mexico City. *Epidemiology*. Vol. 10(2): 118-23.

Manuel, E.H., R.L. Horst, K.M. Brennan, W.N. Lanen, M.C. Duff and J.K. Tapiero. 1982. Benefits Analysis of Alternative Secondary National Ambient Air Quality Standards for Sulfur Dioxide and Total Suspended Particulates, Volumes I-IV. Prepared for U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. Research Triangle Park, NC.

McConnell, R., K. Berhane, F. Gilliland, S.J. London, H. Vora, E. Avol, W.J. Gauderman, H.G. Margolis, F. Lurmann, D.C. Thomas and J.M. Peters. 1999. Air pollution and bronchitic symptoms in Southern California children with asthma. *Environ Health Perspect*. Vol. 107(9): 757-60.

McDonnell, W.F., D.E. Abbey, N. Nishino and M.D. Lebowitz. 1999. Long-term ambient ozone concentration and the incidence of asthma in nonsmoking adults: the AHSMOG study. *Environ Res*. Vol. 80(2 Pt 1): 110-21.

Moolgavkar, S.H., E.G. Luebeck, T.A. Hall and E.L. Anderson. 1995. Air Pollution and Daily Mortality in Philadelphia. *Epidemiology*. Vol. 6(5): 476-484.

Moolgavkar, S.H., E.G. Luebeck and E.L. Anderson. 1997. Air pollution and hospital admissions for respiratory causes in Minneapolis St. Paul and Birmingham. *Epidemiology*. Vol. 8(4): 364-370.

Moolgavkar, S.H. 2000a. Air pollution and hospital admissions for diseases of the circulatory system in three U.S. metropolitan areas. *J Air Waste Manag Assoc*. Vol. 50(7): 1199-206.

Moolgavkar, S.H. 2000b. Air Pollution and Daily Mortality in Three U.S. Counties. *Environ Health Perspect*. Vol. 108(8): 777-784.

Moolgavkar, S.H. 2000c. Air Pollution and Hospital Admissions for Chronic Obstructive Pulmonary Disease in Three Metropolitan Areas in the United States. *Inhalation Toxicology*. Vol. 12(Supplement 4): 75-90.

Moolgavkar, S.H. 2003. Air Pollution and Daily Deaths and Hospital Admissions in Los Angeles and Cook Counties. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May. pp. 183-198.

Mrozek, J.R. and L.O. Taylor. 2002. What Determines the Value of Life? A Meta-Analysis. *Journal of Policy Analysis and Management*. Vol. 21: 253-270.

National Center for Health Statistics. 1999. National Vital Statistics Reports. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Washington, DC. Volume 47, Number 19. June 30.

National Research Council. 2002. Estimating the Public Health Benefits of Proposed Air Pollution Regulations. The National Academies Press: Washington, D.C.

Norris, G., S.N. YoungPong, J.Q. Koenig, T.V. Larson, L. Sheppard and J.W. Stout. 1999. An association between fine particles and asthma emergency department visits for children in Seattle. *Environ Health Perspect*. Vol. 107(6): 489-93.

O'Connor, R.M. and G.C. Blomquist. 1997. Measurement of Consumer-Patient Preferences Using a Hybrid Contingent Valuation Method. *Journal of Health Economics*. Vol. 16: 667-683.

Ostro, B., M. Lipsett, J. Mann, H. Braxton-Owens and M. White. 2001. Air pollution and exacerbation of asthma in African-American children in Los Angeles. *Epidemiology*. Vol. 12(2): 200-8.

Ostro, B.D. 1987. Air Pollution and Morbidity Revisited: A Specification Test. *Journal of Environmental Economics and Management*. Vol. 14: 87-98.

Ostro, B.D. and S. Rothschild. 1989. Air Pollution and Acute Respiratory Morbidity - an Observational Study of Multiple Pollutants. *Environ Res*. Vol. 50(2): 238-247.

Ostro, B.D., M.J. Lipsett, M.B. Wiener and J.C. Selner. 1991. Asthmatic Responses to Airborne Acid Aerosols. *Am J Public Health*. Vol. 81(6): 694-702.

Ostro, B.D., M.J. Lipsett, J.K. Mann, H. Braxtonowens and M.C. White. 1995. Air Pollution and Asthma Exacerbations Among African-American Children in Los Angeles. *Inhalation Toxicology*. Vol. 7(5): 711-722.

Owings, M.F. and L. Lawrence. 1999. Detailed Diagnoses and Procedures, National Hospital Discharge Survey, 1997. National Center for Health Statistics. Hyattsville, MD. Vital Health Statistics, Series 13, No. 145. December.

Parker, J.D. and D.M. Makuc. 2001. Methodologic Implications of Allocating Multiple Race Data to Single Race Categories. Division of Health Utilization and Analysis, National Center for Health Statistics. Hyattsville, MD. Unpublished manuscript.

Pereira, L.A.A., D. Loomis, G.M.S. Conceicao, A.L.F. Braga, R.M. Arcas, H.S. Kishi, R.M. Singer, G.M. Bohm and P.H.N. Saldiva. 1998. Association between air pollution and intrauterine mortality in Sao Paulo, Brazil. *Environmental Health Perspectives*. Vol. 106(6): 325-329.

Peters, A., D.W. Dockery, J.E. Muller and M.A. Mittleman. 2001. Increased particulate air pollution and the triggering of myocardial infarction. *Circulation*. Vol. 103(23): 2810-5.

Pope, C.A. 1991. Respiratory hospital admissions associated with PM10 pollution in Utah, Salt Lake, and Cache Valleys. *Arch Environ Health*. Vol. 46(2): 90-7.

Pope, C.A., D.W. Dockery, J.D. Spengler and M.E. Raizenne. 1991. Respiratory Health and Pm10 Pollution - a Daily Time Series Analysis. *American Review of Respiratory Disease*. Vol. 144(3): 668-674.

Pope, C.A., J. Schwartz and M.R. Ransom. 1992. Daily Mortality and PM10 Pollution in Utah Valley. *Archives of Environmental Health*. Vol. 47(3): 211-217.

Pope, C.A., M.J. Thun, M.M. Namboodiri, D.W. Dockery, J.S. Evans, F.E. Speizer and C.W. Heath. 1995. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am J Respir Crit Care Med*. Vol. 151(3): 669-674.

Pope, C.A., 3rd. 2000. Particulate matter-mortality exposure-response relations and threshold [see comments]. *Am J Epidemiol*. Vol. 152(5): 407-12.

Pope, C.A., 3rd, R.T. Burnett, M.J. Thun, E.E. Calle, D. Krewski, K. Ito and G.D. Thurston. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama*. Vol. 287(9): 1132-41.

Popovic, J.R. 2001. 1999 National Hospital Discharge Survey: annual summary with detailed diagnosis and procedure data. *Vital Health Stat 13*. Vol. (151): i-v, 1-206.

Ransom, M.R. and C.A. Pope. 1992. Elementary School Absences and PM(10) Pollution in Utah Valley. *Environmental Research*. Vol. 58(2): 204-219.

Rosamond, W., G. Broda, E. Kawalec, S. Rywik, A. Pajak, L. Cooper and L. Chambless. 1999. Comparison of medical care and survival of hospitalized patients with acute myocardial infarction in Poland and the United States. *Am J Cardiol*. Vol. 83(8): 1180-5.

Rossi, G., M.A. Vigotti, A. Zanobetti, F. Repetto, V. Gianelle and J. Schwartz. 1999. Air pollution and cause-specific mortality in Milan, Italy, 1980-1989. *Arch Environ Health*. Vol. 54(3): 158-64.

Rowe, R.D. and L.G. Chestnust. 1986. Oxidants and Asthmatics in Los Angeles: A Benefits Analysis -- Executive Summary. Prepared for U.S. Environmental Protection Agency, Office of Policy Analysis. Prepared by Energy and Resource Consultants, Inc. Washington, DC. EPA-230-09-86-018. March.

Russell, M.W., D.M. Huse, S. Drowns, E.C. Hamel and S.C. Hartz. 1998. Direct medical costs of coronary artery disease in the United States. *Am J Cardiol*. Vol. 81(9): 1110-5.

Saldiva, P.H.N., A. Lichtenfels, P.S.O. Paiva, I.A. Barone, M.A. Martins, E. Massad, J.C.R. Pereira, V.P. Xavier, J.M. Singer and G.M. Bohm. 1994. Association Between Air Pollution and Mortality Due to Respiratory Diseases in Children in Sao Paulo, Brazil - a Preliminary Report. *Environ Res*. Vol. 65(2): 218-225.

Samet, J., S. Zeger, F. Dominici, F. Curriero, I. Coursac, D. Dockery, J. Schwartz and A. Zanobetti. 2000. The National Morbidity, Mortality, and Air Pollution Study. Health Effects Institute. Cambridge, MA. Report No. 94. May.

Samet, J.M., S.L. Zeger, J.E. Kelsall, J. Xu and L.S. Kalkstein. 1997. Air Pollution, Weather, and Mortality in Philadelphia 1973-1988. Health Effects Institute. Cambridge, MA. March.

Schwartz, J. 1993. Particulate Air Pollution and Chronic Respiratory Disease. *Environ Res*. Vol. 62: 7-13.

Schwartz, J., D. Slater, T.V. Larson, W.E. Pierson and J.Q. Koenig. 1993. Particulate air pollution and hospital emergency room visits for asthma in Seattle. *Am Rev Respir Dis*. Vol. 147(4): 826-31.

Schwartz, J. 1994a. PM(10) Ozone, and Hospital Admissions For the Elderly in Minneapolis St Paul, Minnesota. *Archives of Environmental Health*. Vol. 49(5): 366-374.

Schwartz, J. 1994b. Air Pollution and Hospital Admissions For the Elderly in Detroit, Michigan. *American Journal of Respiratory and Critical Care Medicine*. Vol. 150(3): 648-655.

Schwartz, J. 1994c. Air Pollution and Hospital Admissions For the Elderly in Birmingham, Alabama. *American Journal of Epidemiology*. Vol. 139(6): 589-598.

Schwartz, J. 1994d. What Are People Dying of On High Air Pollution Days. *Environmental Research*. Vol. 64(1): 26-35.

Schwartz, J., D.W. Dockery, L.M. Neas, D. Wypij, J.H. Ware, J.D. Spengler, P. Koutrakis, F.E. Speizer and B.G. Ferris. 1994. Acute Effects of Summer Air Pollution On Respiratory Symptom Reporting in Children. *Am J Respir Crit Care Med*. Vol. 150(5): 1234-1242.

Schwartz, J. 1995. Short term fluctuations in air pollution and hospital admissions of the elderly for respiratory disease. *Thorax*. Vol. 50(5): 531-538.

Schwartz, J., D.W. Dockery and L.M. Neas. 1996. Is Daily Mortality Associated Specifically With Fine Particles. *Journal of the Air & Waste Management Association*. Vol. 46(10): 927-939.

Schwartz, J. 2000a. Harvesting and long term exposure effects in the relation between air pollution and mortality [see comments]. *Am J Epidemiol*. Vol. 151(5): 440-8.

Schwartz, J. 2000b. The distributed lag between air pollution and daily deaths. *Epidemiology*. Vol. 11(3): 320-6.

Schwartz, J. 2000c. Assessing confounding, effect modification, and thresholds in the association between ambient particles and daily deaths. *Environ Health Perspect*. Vol. 108(6): 563-8.

Schwartz, J. and L.M. Neas. 2000. Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren. *Epidemiology*. Vol. 11(1): 6-10.

Schwartz, J. and A. Zanobetti. 2000. Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. *Epidemiology*. Vol. 11(6): 666-72.

Schwartz, J., F. Laden and A. Zanobetti. 2002. The concentration-response relation between PM(2.5) and daily deaths. *Environ Health Perspect*. Vol. 110(10): 1025-9.

Schwartz, J. 2003. Daily Deaths Associated with Air Pollution in Six US Cities and Short-Term Mortality Displacement in Boston. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May. pp. 219-226.

Sheppard, L., D. Levy, G. Norris, T.V. Larson and J.Q. Koenig. 1999. Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987-1994. *Epidemiology*. Vol. 10(1): 23-30.

Sheppard, L. 2003. Ambient Air Pollution and Nonelderly Asthma Hospital Admissions in Seattle, Washington, 1987-1994. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May. pp. 227-230.

Smith, D.H., D.C. Malone, K.A. Lawson, L.J. Okamoto, C. Battista and W.B. Saunders. 1997. A national estimate of the economic costs of asthma. *Am J Respir Crit Care Med*. Vol. 156(3 Pt 1): 787-93.

Smith, R.L., D. Spitzner, Y. Kim and M. Fuentes. 2000. Threshold dependence of mortality effects for fine and coarse particles in Phoenix, Arizona. *J Air Waste Manag Assoc*. Vol. 50(8): 1367-79.

Spix, C., J. Heinrich, D. Dockery, J. Schwartz, G. Volksch, K. Schwinkowski, C. Collen and H.E. Wichmann. 1993. Air pollution and daily mortality in Erfurt, east Germany, 1980-1989. *Environ Health Perspect*. Vol. 101(6): 518-26.

Stanford, R., T. McLaughlin and L.J. Okamoto. 1999. The cost of asthma in the emergency department and hospital. *Am J Respir Crit Care Med*. Vol. 160(1): 211-5.

Stieb, D.M., R.T. Burnett, R.C. Beveridge and J.R. Brook. 1996. Association between ozone and asthma emergency department visits in Saint John, New Brunswick, Canada. *Environmental Health Perspectives*. Vol. 104(12): 1354-1360.

Thurston, G.D., K. Ito, C.G. Hayes, D.V. Bates and M. Lippmann. 1994. Respiratory hospital admissions and summertime haze air pollution in Toronto, Ontario: consideration of the role of acid aerosols. *Environ Res*. Vol. 65(2): 271-290.

Tolley, G.S. and et al. 1986. Valuation of Reductions in Human Health Symptoms and Risks. Prepared for U.S. Environmental Protection Agency. January.

U.S. Bureau of the Census. 1997. Statistical Abstract of the United States: 1997. 117 ed. Washington, DC.

U.S. Bureau of the Census. 2002. Statistical Abstract of the United States: 2001. Washington DC.

U.S. Department of Education. 1996. The Condition of Education 1996, Indicator 42: Student Absenteeism and Tardiness. National Center for Education Statistics. Washington DC.

U.S. EPA. 1986. Review of the National Ambient Air Quality Standards for Particulate Matter: Updated Assessment of Scientific and Technical Information Addendum to the 1982 OAQPS Staff Paper. U.S. EPA, Office of Air Quality Planning and Standards. Research Triangle Park, NC. EPA 450/05-86-012.

U.S. EPA. 1994. Documentation for Oz-One Computer Model (Version 2.0). Prepared for U.S. EPA, Office of Air Quality Planning and Standards. Prepared by Mathtech, Inc., under Contract No. 68D30030, WA 1-29. Research Triangle Park, NC. August.

U.S. EPA. 1996. Review of National Ambient Air Quality Standards for Ozone: Assessment of Scientific and Technical Information. OAQPS Staff Paper. U.S. EPA, Office of Air Quality Planning and Standards. Research Triangle Park, NC. EPA-452\R-96-007. June.

U.S. EPA. 1997. The Benefits and Costs of the Clean Air Act: 1970 to 1990. U.S. EPA, Office of Air and Radiation, Office of Policy, Planning and Evaluation. Washington, DC. EPA 410-R-97-002. October.

U.S. EPA. 1999a. An SAB Advisory: The Clean Air Act Section 812 Prospective Study Health and Ecological Initial Studies. Prepared by the Health and Ecological Effects SubCommittee (HEES) of the Advisory Council on the Clean Air Compliance Analysis, Science Advisory Board, U.S. Environmental Protection Agency. Washington, DC. EPA-SAB-Council-ADV-99-005. February.

U.S. EPA. 1999b. The Clean Air Act Amendments (CAAA) Section 812 Prospective Study of Costs and Benefits (1999): Advisory by the Health and Ecological Effects Subcommittee on Initial Assessments of Health and Ecological Effects; Part 1. Prepared by the Health and Ecological Effects SubCommittee (HEES) of the Advisory Council on the Clean Air Compliance Analysis, Science Advisory Board, U.S. Environmental Protection Agency. Washington, DC. EPA-SAB-Council-ADV-99-012. July 28.

U.S. EPA. 2002a. Air Quality Criteria for Particulate Matter, Third External Review Draft. National Center for Environmental Assessment, Office of Research and Development. Research Triangle Park, NC. EPA 600/P-99/002aC. April 2002.

U.S. EPA. 2002b. AQS Coding Manual and Data Dictionary: APPENDICES (draft). Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency. Research Triangle Park, NC. May 9.

Vedal, S., J. Petkau, R. White and J. Blair. 1998. Acute effects of ambient inhalable particles in asthmatic and nonasthmatic children. *American Journal of Respiratory and Critical Care Medicine*. Vol. 157(4): 1034-1043.

Viscusi, K. and J.E. Aldy. 2003. The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World. AEI-Brookings Joint Center for Regulatory Studies. Washington, DC. January.

Viscusi, W.K., W.A. Magat and J. Huber. 1991. Pricing Environmental Health Risks - Survey Assessments of Risk - Risk and Risk - Dollar Trade-Offs For Chronic Bronchitis. *Journal of Environmental Economics and Management*. Vol. 21(1): 32-51.

Viscusi, W.K. 1992. Fatal Tradeoffs: Public and Private Responsibilities for Risk. Oxford University Press: New York.

Wang, X., H. Ding, L. Ryan and X. Xu. 1997. Association between air pollution and low birth weight: a community- based study. *Environ Health Perspect*. Vol. 105(5): 514-20.

Watson, W. and J. Jaksch. 1982. Air Pollution: Household Soiling and Consumer Welfare Losses. *Journal of Environmental Economics and Management*. Vol. 9: 248-262.

Weisel, C.P., R.P. Cody and P.J. Lioy. 1995. Relationship between summertime ambient ozone levels and emergency department visits for asthma in central New Jersey. *Environ Health Perspect*. Vol. 103 Suppl 2: 97-102.

Whittemore, A.S. and E.L. Korn. 1980. Asthma and Air Pollution in the Los Angeles Area. *Am J Public Health*. Vol. 70: 687-696.

Wittels, E.H., J.W. Hay and A.M. Gotto, Jr. 1990. Medical costs of coronary artery disease in the United States. *Am J Cardiol*. Vol. 65(7): 432-40.

Woodruff, T.J., J. Grillo and K.C. Schoendorf. 1997. The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States. *Environmental Health Perspectives*. Vol. 105(6): 608-612.

Woods & Poole Economics Inc. 2001. Population by Single Year of Age CD. CD-ROM. Woods & Poole Economics, Inc.

World Health Organization (WHO). 2003. Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide: Report on a WHO Working Group. World Health Organization. Bonn, Germany. EUR/03/5042688. January.

Yu, O., L. Sheppard, T. Lumley, J.Q. Koenig and G.G. Shapiro. 2000. Effects of Ambient Air Pollution on Symptoms of Asthma in Seattle-Area Children Enrolled in the CAMP Study. *Environ Health Perspect*. Vol. 108(12): 1209-1214.

Zanobetti, A., J. Schwartz, E. Samoli, A. Gryparis, G. Touloumi, R. Atkinson, A. Le Tertre, J. Bobros, M. Celko, A. Goren, B. Forsberg, P. Michelozzi, D. Rabczenko, E. Aranguez Ruiz and K. Katsouyanni. 2002. The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement. *Epidemiology*. Vol. 13(1): 87-93.

Zeger, S.L., F. Dominici and J. Samet. 1999. Harvesting-resistant estimates of air pollution effects on mortality. *Epidemiology*. Vol. 10(2): 171-5.