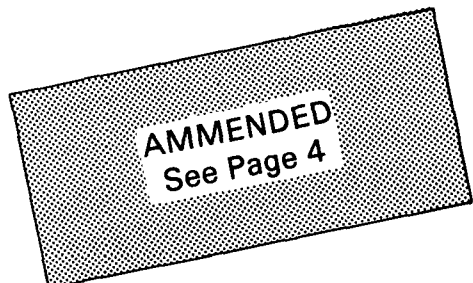




## *Project Summary*



# Effects of Short-Term Intermittent Air Pollutants on Incidence and Severity of Acute Respiratory Disease: Data Collection and Quality Assurance

R. David Flesh, Margaret L. Riha, and Michelle F. Miller

The purpose of this study was to measure the acute response of respiratory disease from peak hourly and daily average exposures to nitrogen dioxide alone and in combination with other pollutants. The study population was made up of families with children attending public elementary schools and living in four California South Coast Air Basin study areas: West Los Angeles/Santa Monica, Garden Grove/Westminster, Glendora/Covina, and Upland/Ontario.

The role and responsibility of the contractor were to collect all background and health data, to process and assure the quality of the data, and to prepare the report. The contractor was assisted by the Statewide Air Pollution Research Center at the University of California, Riverside (UCR), in the data collection tasks.

Families were recruited for participation during the spring and summer months of 1978. Recruitment included face-to-face interviews for the purpose of collecting family background data. Over 3,000 families were interviewed. From these, 1,402 families (about 350 per study area) were selected.

Participation required that families record daily the disease symptoms that they experienced. According to instructions mailed to them, the symptoms were to be recorded on a specially prepared calendar. Each week, the symptoms recorded the previous week were given by telephone to a staff of trained interviewers employed by the contractor. Symptom surveillance began in September 1978 and ended in March 1979.

The background and symptom data were reviewed and edited in the field and in the contractor's main offices prior to data entry. Computer files of the background and symptom data were prepared in three steps. First, the data were transferred from the background and symptom questionnaires to magnetic tapes through the use of an optical reader. Second, supplementary information including updates to the background data and doctors' diagnoses of illnesses reported during symptom surveillance were key-taped into a separate computer file. Third, quality assurance procedures were used to minimize error in the computer files. Maximum allowable error was set at 2%. The computer

files and all questionnaires were then sent to the EPA for analysis.

*This Project Summary was developed by EPA's Health Effects Research Laboratory, Research Triangle Park, NC, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).*

## Introduction

This report covers the health data collection phase of the project.

## Delineating the Study Areas

Four study areas had been tentatively chosen by the EPA for this study in 1976.

In December 1977, the EPA decided to place a second air monitoring station in each of the study areas. The purpose of installing a second station was to permit better characterization of air pollution exposure of families enrolled in the study. An added benefit of having a second station was to expand the portion of each area in which families could be interviewed.

The criterion placed on the locations of the second stations was that they be located no more than two miles from the existing EPA air monitoring stations. An additional criterion stated that the 2-mile area around the new station must be socio-economically comparable to the 2-mile area around the existing station.

The final shape of each of the study areas was a rectangle, measuring two miles by four miles. This was done to improve estimates of human exposure over what would have been obtained if only one air monitoring station had been located in each of the areas and to increase the number of households from which families could be recruited. The study areas were divided into interviewing zones so that the addresses of elementary school families residing in the areas could be plotted on the maps and an orderly schedule of interviewing could be established.

## Background Survey Interviewing

Lists of student and/or family names and addresses were obtained by the contractor. Using EPA-supplied maps, the contractor's clerical staff listed the streets and street number limits which fell into each zone, making proper zone

assignment of each address an easy task.

For a given interviewing zone, the study area with the lowest number of multiple family units determined the number of multiple family dwellings to be interviewed in all four areas. For example, if a particular zone in West Los Angeles had the least number of multiple family dwellings of the four areas, then that number of multiple family addresses would be randomly chosen from the corresponding zones in each of the other three study areas. The same procedure was used for single family dwellings. The zone sample size was, therefore, the sum of the lowest number of multiple family dwellings and the lowest number of remaining addresses making the zone sample size the same for all four study areas. This technique resulted in an initial sample size of 1,294 homes in each study area.

A training session for interviewers was held on May 18 and 19. It began with a presentation given by the contractor on the proper techniques of interviewing in studies of scientific concern. Following this, the interviewers were given copies of the ARD Study Background Questionnaire and interviewers' instructions, which contained information on the background and purpose of the study, a sheet of questions and answers about the study that the interviewers were to use when they contacted the school families for interviews, an interviewing schedule, and maps of the study areas.

After the formal presentations were completed the interviewers practiced administering the questionnaire to each other. Sample interviews were presented and critiqued, and then actual interviews were practiced in the field. The survey began in West Los Angeles on May 22, 1978, the first work day following the training session. The interviewers met with their team leaders late in the afternoon, giving them the questionnaires they completed that day. During this meeting the interviewers received a new supply of blank questionnaires and address sheets for the study area and zone where interviews were to be conducted the next day.

Families who could not be contacted on a previous visit were reassigned to the interviewers. Every day, the interviewers would first attempt to make callbacks in previous zones before making contacts with families in newly assigned zones.

Interviewing was rotated from study area to study area to avoid day of the week effects and to maximize the chances of finding the families at home. A "round" of interviewing was said to have been completed after each four days of the survey. The final round (Round 12) of interviewing ended on August 1. The total numbers of interviews completed in each of the study areas during the 12 rounds of interviewing were as follows: 740 in West Los Angeles/Santa Monica, 754 in Garden Grove/Westminster, 824 in Glendora/Covina, and 842 in Upland/Ontario. Telephone validation of 10 percent of the interviews completed by each interviewer was conducted through Round 12 by non-interviewing personnel. No significant differences were found between the initial interviews and the validation interviews.

## Selecting Families for Participation

As the face-to-face interviews of families progressed through the 12 rounds, the questionnaires were edited for completeness and inconsistencies. The answers given were checked to determine eligibility, also. Five criteria, developed by the EPA, had to be met for families to qualify for participation.

1. Have access to working telephone.
2. Have lived in the study area for one or more years.
3. Do not plan to move within 12 months.
4. Agree to participate in the weekly telephone surveillance.
5. Have one or more elementary school children living in the home.

Initially, priority for selection was to take the following factors into consideration:

1. Presence of the pre-school children in the family.
2. Length of residence in the study area:
  - a. Three or more years
  - b. One to three years
3. Parents living the home:
  - a. Both parents
  - b. Female parent
  - c. Male parent
4. Type of fuel used for cooking:
  - a. Gas
  - b. Electricity

5. Type of air conditioning in the home:
  - a. Central
  - b. Room
  - c. None
6. Type of dwelling unit:
  - a. Single Family
  - b. Multi-family
7. Presence of older family members (e.g., grandparents) living in the home.

It was planned that the study areas would be matched in terms of these factors as well as in terms of the socioeconomic characteristics. However, as the responses to the Background Questionnaires were being coded, it quickly became clear that the study area could not be completely matched and that all of the factors could not be accommodated if the selection process was to yield the quota of at least 350 families in each study area. Therefore, a decision was made by the EPA to use 80% of the families who cook with gas and 20% of the families who cook with electricity in each study area where this combination was possible (in West Los Angeles, Garden Grove, and Glendora) and a 50%:50% ratio of gas to electricity where the 80%:20% ratio was not possible (in Upland). The EPA also decided that 2-parent families with no pre-schoolers, living in single family dwellings, should be given lowest selection priority and should be randomly sampled to complete the quota of 350 panel families in each study area. Other factors, including the type of air conditioning and the presence of older family members living in the home, had to be dropped completely as selection criteria, since too few homes in the West Los Angeles study area had air conditioning and too few school families contained older people.

A total of 1,400 families (4 study areas x 350 families per study area) were selected and randomly distributed into 14 groups in preparation for weekly telephone surveillance. Each group was assigned to a different telephone interviewer.

### **Weekly Telephone Interviewing**

In August 1978, the contractor and subcontractor each recruited seven telephone interviewers and two alternates in preparation for weekly symptom surveillance. All interviewers had previously been members of the background survey teams. A training session was

held on September 6 and 7 with the interviewers, alternates, and supervisors in attendance. An EPA instruction manual on how to complete the ARD Study Symptom Questionnaire was distributed to each of the interviewers prior to the briefing. A review of the instructions was followed by a demonstration on conducting the weekly symptom surveillance calls. After a period of questions and answers, the interviewers practiced conducting weekly interviews.

Pre-surveillance telephone calls were conducted during the week of September 11. During the presurveillance calls, an interviewer ascertained, without directly asking, whether or not the respondent still wished to participate in the panel. Respondents were asked if they had received their calendars and if they had any questions relating to the use of the calendar. Next, a day and time for the weekly telephone calls was established with each respondent. Finally, all information on the background of the respondent's family was updated.

Families which stated, during the pre-surveillance call, that they were no longer interested in participating in the study, were replaced by eligible families which were selected as alternates. A withdrawn family that cooked with gas was replaced with another family from the same study area that also cooked with gas. This was done to maintain the existing balance of selection factors and the quota of 350 panel families in each study area.

The families began to keep records of symptoms on their calendars on September 10. The first week of symptom surveillance telephoning took place during the week of September 18. Each interviewer was provided with 100 pre-printed ARD Study Symptom Questionnaires, one for each of the families to which the interviewer was assigned.

All weekly calls to families were made on Mondays and Tuesdays whenever possible. The records of contacts with all families were reviewed at regular intervals by the field supervisors to ascertain how many of the families' surveillance calls had to be made on days other than Monday or Tuesday, how many weeks information was taken from an alternate respondent, and how many weeks the family was unable to be contacted. If a family reported on an alternate day, used an alternate respondent, and/or failed to give a report

entirely, for more than eight weeks, the family was withdrawn from the study.

Only during the Week 1 were families which withdrew from the study replaced by alternates. After Week 1, no further replacements were made.

Any changes in background information previously supplied by families were recorded as they occurred during the 12 of surveillance. A special form was used for this purpose.

After every six weeks of symptom surveillance, i.e., after Weeks 6, 12, and 18, the 14 groups of families were randomly reassigned to the interviewers. In accordance with the instructions of the EPA Project Officer, no interviewer was assigned the same group of families she had interviewed previously.

During Weeks 12 and 24 random samples of 10% of the completed interviews from each interviewer's group of families was selected for validation. The selected families who had been interviewed on Monday or Tuesday of Weeks 12 and 24 were reinterviewed on Wednesday or Thursday of the same week by the field supervisors. The integrity of the validation procedure was enhanced by the exchange of assignments between the field supervisors, i.e., the prime contractor's field supervisor validated the interviews conducted by the subcontractor's interviewers and vice versa. Once again, no significant differences were found between the initial interviews and the validation interviews.

The final session (Week 26) of symptom surveillance telephone calls was conducted on March 12 and 13. The numbers of families who participated through the 26 weeks of the study were as follows: 316 (90.3%) in West Los Angeles, 290 (82.9%) in Garden Grove/Westminster, 300 (85.7%) in Glendora/Covina, and 310 (88.6%) in Upland/Ontario.

### **Assuring the Accuracy of the Data Files**

As a preliminary step in testing the accuracy of the data files, computer range checks were performed for all questions in order to verify that all answers were within a valid range. Invalid responses were checked, recoded, and corrected on the data tapes.

The quality assurance testing procedure employed sampling inspection tables based on a Poisson distribution, which sets the probability at 90% that less than 2.0% of the forms have one or

more errors. For a quantity of between 3,001 and 4,000 Background Questionnaires, 510 questionnaires had to be drawn at random, and checked against the data from those questionnaires that exist in the computer files. According to this sampling plan, the maximum number of acceptable errors was six. For the purpose of counting errors, an error was considered to exist if one or more responses in the computer file of a particular questionnaire differed from the corresponding responses on the actual questionnaire.

A 100% check on the Background Questionnaires was conducted as a penalty for failing the sample test. Corrections to the background data files were made and a new random sample of 510 questionnaires was selected and compared to the computer files twice before the computer file passed the test. The background data files, after passing the quality assurance test, were submitted to EPA on two magnetic tapes.

For a quantity of between 1,001 and 2,000 Symptom Questionnaires a random sample of 380 family records had to be selected. This was done for each of the 26 weeks of data. The computer files corresponding to the random sample of 380 questionnaires were allowed to have a maximum of four errors in order to be within the acceptable levels of the sampling plan. An error existed when a family record in the computer files did not exactly correspond to the actual questionnaire for that family.

The first three weeks sampled failed with a great many errors and the decision was made to perform a 100% check of all 26 weeks of symptom data before doing any further random sampling. All 26 weeks were 100% checked by com-

paring the computer files to the original written forms.

Near the end of the contractor's authorized level of effort, random samples were drawn from each week of surveillance. Weeks 1, 4, 6, 13, 20, 21, 22, 23, 25, and 26 passed the test, since they had four or less errors per sample. The remaining weeks were again subjected to 100% error checks and corrections were again made to the computer files to eliminate the errors encoun-

tered. The symptom data files were submitted to the EPA on 26 magnetic tapes, one for each week of symptom surveillance. EPA continued processing the weekly surveillance forms that had failed the quality assurance test until they also passed.

This contract was for collection of health data only. HURL, RTP, USEPA is in the process of analyses and interpretation of this data set.

*R. David Flesh, Margaret L. Riha, and Michelle F. Miller are will Del Green Associates, Inc., San Diego, CA 92131.*

*Wilson B. Riggan is the EPA Project Officer (see below).*

*The complete report, entitled "Effects of Short-Term Intermittent Air Pollutants on Incidence and Severity of Acute Respiratory Disease: Data Collection and Quality Assurance," (Order No. PB 82-129 479; Cost: \$21.00, subject to change) will be available only from:*

*National Technical Information Service  
5285 Port Royal Road  
Springfield, VA 22161  
Telephone: 703-487-4650*

*The EPA Project Officer can be contacted at:  
Health Effects Research Laboratory  
U.S. Environmental Protection Agency  
Research Triangle Park, NC 27711*

United States  
Environmental Protection  
Agency

Center for Environmental Research  
Information  
Cincinnati OH 45268

Postage and  
Fees Paid  
Environmental  
Protection  
Agency  
EPA 335



Official Business  
Penalty for Private Use \$300

PS 0000324  
U S ENVIR PROTECTIUN AGENCY  
REGION 5 LIBRARY  
230 S DEARBORN STREET  
CHICAGO IL 60604