APPENDIX A

LOMA LINDA UNIVERSITY

LOMA LINDA, CALIFORNIA 92354
(714) 796-7311 EXT. 3717

SCHOOL OF HEALTH
ADVENTIST HEALTH STUDY

Dear Friend:

You are one of a small group selected from all participants in the Adventist Health Study to help in a special substudy. This substudy is sponsored by the Air Resources Board to measure some effects of the type of air you breathe.

We have greatly appreciated your cooperation and efforts in completing the detailed lifestyle questionnaire which is helping us to determine the possible relationship between various aspects of lifestyle and health status. The enclosed questionnaire will supplement this information with some additional questions.

Most other members in your church are receiving only the back page of this questionnaire which is the first of the yearly hospital history forms being sent to all adult SDAs in California. It is extremely important for you to complete this last page because it is our only means of keeping track of the health status of California SDAs. The few minutes necessary to fill in the entire questionnaire will contribute significantly to new knowledge that may save many lives.

By completing this questionnaire <u>NOW</u>, you will save us the expense and effort of having to contact you personally. Please return the completed questionnaire in the enclosed self-addressed envelop. Thank you for your assistance.

Sincerely yours,

Roland L. Phillips, M.D.
Director

P.S. According to our records, the first copy of this questionnaire which was mailed to you recently has not yet been returned to our office. In case you never received or have not yet completed this questionnaire, we are sending you this second copy of the form to make it convenient for you to complete and return as soon as possible. If you have already returned the form, please accept our sincere thanks and discard this copy.

EXHIBIT 1

1977 RESPIRATORY SYMPTOMS AND RESIDENCE HISTORY QUESTIONNAIRE

Please answer every question.

COUGH

1. Do you usually cough first thing in the morning?
   1 [ ] Yes
   2 [ ] No

2. Do you usually cough at other times during the day or night?
   1 [ ] Yes
   2 [ ] No

3. Do you cough on most days for 3 months or more?
   1 [ ] Yes
   2 [ ] No

4. For how many years have you had a cough?
   1 [ ] Never
   2 [ ] Less than 1 year
   3 [ ] More than 1 but less than 2 years
   4 [ ] 2-5 years
   5 [ ] More than 5 years

SPUTUM

5. Do you usually bring up phlegm, sputum, or mucus from your chest first thing in the morning?
   1 [ ] Yes
   2 [ ] No

6. Do you usually bring up phlegm, sputum, or mucus from your chest at other times during the day or night?
   1 [ ] Yes
   2 [ ] No

7. Do you bring up phlegm, sputum or mucus from your chest on most days for 3 months of the year or more?
   1 [ ] Yes
   2 [ ] No

8. For how many years have you raised phlegm, sputum, or mucus from your chest?
   1 [ ] Never
   2 [ ] Less than 1 year
   3 [ ] More than 1 but less than 2 years
   4 [ ] 2-5 years
   5 [ ] More than 5 years

PLEASE GO TO TOP OF NEXT COLUMN.

WHEEZING

9. Does your breathing ever sound wheezy or whistling?
   1 [ ] Yes
   2 [ ] No

10. Have you ever had attacks of shortness of breath with wheezing?
   1 [ ] Yes
   2 [ ] No

BREATHLESSNESS

11. Are you troubled by shortness of breath when hurrying on level ground or walking up a slight hill?
   1 [ ] Yes
   2 [ ] No

12. Do you get short of breath when walking at a normal pace with other people of your own age on level ground?
   1 [ ] Yes
   2 [ ] No

RESPIRATORY ILLNESS

13. During the PAST YEAR, how often were you unable to do your usual activities because of illnesses such as chest colds, bronchitis, or pneumonia?
   1 [ ] None
   2 [ ] 1 time
   3 [ ] 2-5 times
   4 [ ] More than 5 times

14. Do you think you have ever had any of these chest disorders--asthma, any kind of bronchial condition, or emphysema?
   1 [ ] Yes
   2 [ ] No

15. Has a doctor ever told you that you had asthma, sone kind of bronchial condition, or emphysema?
   1 [ ] No

   IF YES, please check [✓] which conditions.

   2 [ ] Asthma
   4 [ ] Bronchial condition
   8 [ ] Emphysema

EXHIBIT 1

16. How many days per month during the SUMMER (June thru September) are you bothered by stuffy nose or post-nasal drip (i.e. drainage from the back of your nose into your throat)?

1[ ] None
2[ ] 1-5 days
3[ ] 6-10 days
4[ ] 11-20 days
5[ ] 21 days or more

17. How many days per month during the WINTER (October thru May) are you bothered by stuffy nose or post-nasal drip (i.e. drainage from the back of your nose into your throat)?

1[ ] None
2[ ] 1-5 days
3[ ] 6-10 days
4[ ] 11-20 days
5[ ] 21 days or more

Have you EVER regularly smoked cigarettes, pipes, or cigars (aside from possibly trying them once or twice)?

	Yes	No	
18.	1[ ]	2[ ]	Cigarettes
19.	1[ ]	2[ ]	Pipes
20.	1[ ]	2[ ]	Cigars

During the PAST YEAR, how many times have had the following illnesses? (Please ..eck [✓] the appropriate box for EACH illness.)

21. Head cold (e.g. runny nose, sore throat, etc.)

None 1 2 3 or more
0[ ] 1[ ] 2[ ] 3[ ]

22. Chest cold (acute bronchitis e.g. cough and sputum associated with respiratory infection)

None 1 2 3 or more
0[ ] 1[ ] 2[ ] 3[ ]

23. Pneumonia

None 1 2 3 or more
0[ ] 1[ ] 2[ ] 3[ ]

24. How many times was this pneumonia diagnosed by a physician using a chest x-ray?

None 1 2 3 or more
0[ ] 1[ ] 2[ ] 3[ ]

PLEASE GO TO TOP OF NEXT COLUMN.

RESIDENCE

25. Are you usually away from home for more than 2 weeks during the summer (June thru September)?

1[ ] No
2[ ] Yes

26. IF YES, how long are you usually away?

1[ ] 3-4 weeks

2[ ] 5-6 weeks

3[ ] 7-8 weeks

4[ ] 9 weeks or more

27. How many hours per DAY during the work week do you usually spend driving or riding on CROWDED roadways? (Check the nearest category.)

1[ ] None
2[ ] Less than 15 minutes
3[ ] 15 minutes to one hour
4[ ] 2 hours
5[ ] 3 hours
6[ ] 4 hours
7[ ] 5 hours
8[ ] 6 hours or more

28. On a typical WEEKEND, how many hours per day do you spend driving or riding on CROWDED roadways? (Check the nearest category.)

1[ ] None
2[ ] Less than 15 minutes
3[ ] 15-29 minutes
4[ ] 30-59 minutes
5[ ] 1-2 hours
6[ ] 3-4 hours
7[ ] 5-6 hours
8[ ] 6 hours or more

29. How often do you use aerosol sprays (e.g. hair spray, cleaning spray, deodorant, spray paint, etc.)?

1[ ] Daily
2[ ] Several times a week
3[ ] Once a week
4[ ] a few times a month
5[ ] Rarely or never

30. What is your usual or main occupation? (Do not write "retired". If retired or not now working, give your usual occupation when you were working.)

Job Title ______________

Major duties or responsibilities:

____

EXHIBIT 1

How many hours per WEEK, including weekends, do you exercise vigorously or do heavy physical labor (e.g. jogging, tennis, gardening, etc.) in the open air?

31. SUMMER (June thru September)
- 1[ ] None
- 2[ ] 1-7 hours
- 3[ ] 8-14 hours
- 4[ ] 15-21 hours
- 5[ ] 22-28 hours
- 6[ ] 29-35 hours
- 7[ ] 36-42 hours
- 8[ ] More than 42 hours

32. REST OF YEAR (October thru May)
- 1[ ] None
- 2[ ] 1-7 hours
- 3[ ] 8-14 hours
- 4[ ] 15-21 hours
- 5[ ] 22-28 hours
- 6[ ] 29-35 hours
- 7[ ] 36-42 hours
- 8[ ] More than 42 hours

How many hours per WEEK, including weekends, are you outside of buildings?

33. SUMMER (June thru September)
- 1[ ] None
- 2[ ] 1-7 hours
- 3[ ] 8-14 hours
- 4[ ] 15-21 hours
- 5[ ] 22-28 hours
- 6[ ] 29-35 hours
- 7[ ] 36-42 hours
- 8[ ] More than 42 hours

34. REST OF YEAR (October thru May)
- 1[ ] None
- 2[ ] 1-7 hours
- 3[ ] 8-14 hours
- 4[ ] 15-21 hours
- 5[ ] 22-28 hours
- 6[ ] 29-35 hours
- 7[ ] 36-42 hours
- 8[ ] More than 42 hours

35. Have you ever lived for one year or more with someone who smoked?

0[ ] No
[ ] Yes -----> How many years? ____________

36. Have you worked in the same room with someone who smoked?

0[ ] No
[ ] Yes -----> How many years? ____________

37. Have you ever worked where you were exposed much of the time to various types of contaminated air such as chemical fumes, paint fumes, welding, wood or rock dust, etc.

0[ ] No
[ ] Yes -----> How many years? ____________

IF YES, please list:

38. Type of work ____________________

39. Type of contamination ____________________

If you have worked more than 5 miles from home in the past 10 years, please give the work locations and dates

	Started Job: MONTH	YEAR	Ended Job: MONTH	YEAR	TOWN OF WORK	STATE	ZIP CODE PLACE OF WORK
40.							
41.							
42.							
43.							
44.							
45.							

EXHIBIT 1

For each community you have lived in since 1960, please give the information requested below. For large cities, please give the section of the city. If the town was so small that it did not have a post office, give the nearest post office.

Please start with your current place of residence.

	TOWN	STATE	ZIP	Moved to this town MONTH	YEAR
46.	______	______	______	______	______
47.	______	______	______	______	______
48.	______	______	______	______	______
49.	______	______	______	______	______
50.	______	______	______	______	______
51.	______	______	______	______	______

52. For each of the following, please indicate with a [✓] which years you have:

	Never	Prior to 66	1966	67	68	69	70	71	72	73	74	75	76	77
a. Lived in a home with evaporative water cooling.	a[ ]	[ ]	[ ]	[ ]	[ ]	a[ ]	[ ]	[ ]	[ ]	[ ]	a[ ]	[ ]	[ ]	[ ]
b. Lived in a home with refrigerated air-conditioning.	b[ ]	[ ]	[ ]	[ ]	[ ]	b[ ]	[ ]	[ ]	[ ]	[ ]	b[ ]	[ ]	[ ]	[ ]
c. Worked in a building with air-conditioning.	c[ ]	[ ]	[ ]	[ ]	[ ]	c[ ]	[ ]	[ ]	[ ]	[ ]	c[ ]	[ ]	[ ]	[ ]
d. Lived with someone who smoked.	d[ ]	[ ]	[ ]	[ ]	[ ]	d[ ]	[ ]	[ ]	[ ]	[ ]	d[ ]	[ ]	[ ]	[ ]
e. Worked in the same room with someone who smoked.	e[ ]	[ ]	[ ]	[ ]	[ ]	e[ ]	[ ]	[ ]	[ ]	[ ]	e[ ]	[ ]	[ ]	[ ]
f. Worked more than 5 miles from home.	f[ ]	[ ]	[ ]	[ ]	[ ]	f[ ]	[ ]	[ ]	[ ]	[ ]	f[ ]	[ ]	[ ]	[ ]
g. Worked where you were exposed much of the time to various types of contaminated air such as chemical fumes, paint fumes, welding, wood or rock dust, etc.	g[ ] 52	[ ] 53	[ ] 54	[ ] 55	[ ] 56	g[ ] 57	[ ] 58	[ ] 59	[ ] 60	[ ] 61	g[ ] 62	[ ] 63	[ ] 64	[ ] 65

PLEASE CONTINUE ON NEXT PAGE.

ADVENTIST HEALTH STUDY HOSPITAL HISTORY FORM

Please complete each item below by filling in the spaces or checking the appropriate boxes.

1. Have you stayed overnight in a hospital within the last 10 years? ☐ YES ☐ NO

2. **If yes, starting with your most recent hospitalization, complete the information below for each DIFFERENT hospital you have stayed in. Record each hospital only once, even if you have been hospitalized many times in that hospital.**

NAME OF HOSPITAL	CITY	STATE	YEAR OF LAST STAY

3. **If you have stayed in more than 4 different hospitals since 1967, please check here.** ☐

4. Within the past 12 months, have you been told by a doctor (FOR THE FIRST TIME) that you had any of the conditions below? If yes, please indicate where the diagnosis was made.

PEPTIC ULCER ☐ ______ NAME OF DOCTOR OR CLINIC ______ NAME OF HOSPITAL (IF DIAGNOSED IN HOSPITAL) ______ CITY ______ STATE

DIABETES ☐ ______ NAME OF DOCTOR OR CLINIC ______ NAME OF HOSPITAL (IF DIAGNOSED IN HOSPITAL) ______ CITY ______ STATE

ARTHRITIS ☐ ______ NAME OF DOCTOR OR CLINIC ______ NAME OF HOSPITAL (IF DIAGNOSED IN HOSPITAL) ______ CITY ______ STATE

NONE OF THESE ☐

5. Check whether you have had either of these medical procedures done **WITHIN THE LAST 12 MONTHS.** If yes, please indicate where it was done.

SPECIAL UPPER G.I. XRAY (Barium Swallow) ☐ ______ NAME OF DOCTOR OR CLINIC ______ NAME OF HOSPITAL (IF DONE IN HOSPITAL) ______ CITY ______ STATE

SPECIAL XRAY OF LARGE BOWEL (Barium Enema) ☐ ______ NAME OF DOCTOR OR CLINIC ______ NAME OF HOSPITAL (IF DONE IN HOSPITAL) ______ CITY ______ STATE

NEITHER OF THESE ☐

6. **In order to compile meaningful health statistics regarding participants in this study, we need to review your medical records. Your signature below will authorize us to do this.**
I hereby authorize the Adventist Health Study to examine my medical records filed in any hospital, clinic or doctor's office mentioned on this form.

______ YOUR SIGNATURE ______ DATE

7. **If you are under age 25, please check here.** ☐

8. **Please be sure to make any necessary corrections in your name, address and telephone as printed below.**

PLEASE CHECK THE QUESTIONNAIRE TO MAKE SURE THAT YOU HAVE LEFT NO QUESTIONS BLANK WHICH SHOULD HAVE BEEN ANSWERED.
THANK YOU VERY MUCH.

EXHIBIT 2

By Percentage

Loma Linda University

Adventist Health Study
Department of Preventive Medicine
School of Medicine
Loma Linda, CA 92350
714/824-4684

PERCENTAGE RESPONSES FOR LIVING SUBJECTS. 197 Individuals excluded according to exclusion criteria. Some questions also have a built-in selection statement. For these questions the percentages are those with the selection applied.

The n of 4457 for this group is larger than the n of 3914 for the respiratory symptoms cohort used to study air pollution health effects as more individuals were excluded due to failing reliability checks or problems in their residence history making it impossible to estimate ambient air pollution concentrations.

Make sure you are the person named and that address and telephone number are correct.

See * below if this person is no longer living here.

Dear Friend,

In 1977, you were one of a select group from the Adventist Health Study who participated in a special substudy that was sponsored by the California Air Resources Board. Your response to previous questionnaires has been very much appreciated and the results of this substudy have been widely recognized and used by State and National Agencies. Once again we are seeking your cooperation and assistance in completing the following questionnaire for the follow up of this special substudy.

It is important to the scientific validity of the study that all study participants fill out the questionnaire as close as possible to the same point in time. Please take a few minutes now and complete this questionnaire and mail it in the enclosed stamped, return envelope.

All the information will be kept strictly confidential and will be reported only in statistical summaries of large groups of people. Thank you for your important contribution to this research project.

Sincerely yours,

David E. Abbey

David E. Abbey, Ph.D.
Co-director, Adventist Health Study

*

If this person is no longer living at this address, please indicate the person's status and a new address if available or the contact name and address of a close relative and RETURN the uncompleted questionnaire in the enclosed envelope.

[ ] This person is deceased

[ ] This person is now living at a new address:

Name of contact person ______________________

Address ______________________

____

n = 4,457

RESPIRATORY SYMPTOMS AND RESIDENCE HISTORY QUESTIONNAIRE – 1987

INSTRUCTIONS

Please answer EVERY question. For some questions you may not remember the exact detail. If this is the case, then guess as closely as you can.
PLEASE DO NOT LEAVE ANY QUESTIONS BLANK UNLESS ASKED TO SKIP THEM.

Please indicate your answer by placing an X in the appropriate box [ ] or by writing your answer in the space provided.

Check your name, address and telephone number on the first page adding or correcting address and telephone number if necessary. **Make sure you are the person named on the label.**

1. Please enter today's date ____/____/____
   month day year

COUGH

2. Do you usually cough first thing in the morning?

446 [ ] Yes
4009 [ ] No
2 missing

3. Do you usually cough at other times during the day or night?

724 [ ] Yes
3728 [ ] No
5 missing

4. Do you cough on most days for 3 months or more?

474 [ ] Yes
3974 [ ] No
9 missing

5. For how many years have you had a cough?

3350 [ ] Never
333 [ ] Less than 1 year
137 [ ] More than 1 but less than 2 years
206 [ ] 2–5 years
376 [ ] More than 5 years
55 missing

SPUTUM

6. Do you usually bring up phlegm, sputum, or mucus from your chest first thing in the morning?

607 [ ] Yes
3845 [ ] No
4 missing

7. Do you usually bring up phlegm, sputum or mucus from your chest at other times during the day or night?

495 [ ] Yes
956 [ ] No
6 missing

PLEASE GO TO TOP OF NEXT COLUMN

8. Do you bring up phlegm, sputum or mucus from your chest on most days for 3 months of the year or more?

514 [ ] Yes
3931 [ ] No
12 missing

9. For how many years have you raised phlegm, sputum, or mucus from your chest?

3410 [ ] Never
312 [ ] Less than 1 year
123 [ ] More than 1 but less than 2 years
167 [ ] 2–5 years
408 [ ] More than 5 years
37 missing

WHEEZING

10. Does your breathing ever sound wheezy or whistling?

683 [ ] Yes
3766 [ ] No
8 missing

11. Does your chest ever sound wheezy or whistling? (Check No or Yes for each)

No		Yes		
2944	[ ]	1284 [ ]	When you have a cold	229 missing
2996	[ ]	570 [ ]	Occasionally apart from colds	891 m
3271	[ ]	72 [ ]	Most days or nights	1114 missing

12. Have you ever had attacks of shortness of breath with wheezing?

460 [ ] Yes (continue)
3981 [ ] No ⟶ skip to question 15
16 missing

PLEASE GO TO TOP OF NEXT PA...

13. Have you had 2 or more such episodes?

393 [ ] Yes
51 [ ] No
16 missing

14. Have you ever required medicine or treatment for the(se) attack(s)?

287 [ ] Yes
157 [ ] No
16 missing

BREATHLESSNESS

15. Are you troubled by shortness of breath when hurrying on level ground or walking up a slight hill?

1439 [ ] Yes
2989 [ ] No
29 missing

16. Do you get short of breath when walking at a normal pace with other people of your own age on level ground?

487 [ ] Yes
3957 [ ] No
13 missing

RESPIRATORY ILLNESS

17. During the PAST YEAR, how often were you unable to do your usual activities because of illnesses such as chest colds, bronchitis, or pneumonia?

3343 [ ] None
829 [ ] 1 time
239 [ ] 2-5 times
37 [ ] More than 5 times
9 missing

18. Do you think you have ever had any of these chest disorders—asthma, any kind of bronchial condition, or emphysema?

1016 [ ] Yes
3437 [ ] No
4 missing

19. Has a doctor ever told you that you had asthma, some kind of bronchial condition, or emphysema?

3558 [ ] No 899 missing
IF YES, please check [X] which conditions
335 [ ] Asthma 4122 missing
634 [ ] Bronchial condition 3823 missing
73 [ ] Emphysema 4384 missing

PLEASE GO TO TOP OF NEXT COLUMN

THE FOLLOWING QUESTION REQUESTS ADDITIONAL INFORMATION. PLEASE READ IT CAREFULLY. WE NEED TO KNOW IF YOU HAVE HAD THESE AND OTH CONDITIONS AND THE AGE OF FIRST DIAGNOSIS.

20. Has a doctor ever told you that you had asthma, some kind of bronchial condition, pneumonia, emphysema, or any other serious respiratory condition?

3106 [ ] No 1351 missing
IF YES, check all that apply and give age of first diagnosis

340 [ ] Asthma 1011 missing
_______ age of first diagnosis

526 [ ] Attacks of bronchitis 825 missing
_______ age of first diagnosis

178 [ ] Chronic bronchitis 1173 missing
_______ age of first diagnosis

707 [ ] Pneumonia 644 missing
_______ age of first diagnosis

71 [ ] Emphysema 1280 missing
_______ age of first diagnosis

90 [ ] Other (specify) 1261 missing
_______ age of first diagnosis

21. Do you currently have asthma that has been confirmed by a doctor?

4309 [ ] No
137 [ ] Yes
11 missing

22. IF YES, are you currently taking medication for asthma?

1745 [ ] No
92 [ ] Yes
2620 missing

23. Before starting school (up to 7 years of age) do you think you had more or less than the average number of colds for children your age?

824 [ ] much less
998 [ ] less
2314 [ ] about the same
226 [ ] more
47 [ ] much more
48 missing

PLEASE GO TO TOP OF NEXT PAGE

24. How many **days per month** during the SUMMER (June through September) are you bothered by stuffy nose or post-nasal drip (i.e. drainage from the back of your nose into your throat)?

2312 [ ] None
823 [ ] 1-5 days
336 [ ] 6-10 days
296 [ ] 11-20 days
672 [ ] 21 days or more
18 missing

25. How many **days per month** during the WINTER (October through May) are you bothered by stuffy nose or post-nasal drip (i.e. drainage from the back of your nose into your throat)?

1790 [ ] None
1155 [ ] 1-5 days
492 [ ] 6-10 days
331 [ ] 11-20 days
667 [ ] 21 days or more
22 missing

SMOKING HISTORY

26. Have you EVER regularly smoked cigarettes, cigars or a pipe (aside from possibly trying them once or twice?)

3770 [ ] No ➡ skip to question 34
680 [ ] Yes
7 missing

27. At what age did you first start smoking regularly?

________ age in years

28. Are you currently smoking?

680 [ ] No
0 [ ] Yes ➡ skip to question 30

29. At what age did you stop smoking?

_______ age in years

30. Have you ever regularly smoked cigars?

618 [ ] No
58 [ ] Yes ➡ **How many years?**_______
4 missing

31. Have you ever regularly smoked a pipe?

607 [ ] No
69 [ ] Yes ➡ **How many years?**_______
4 missing

32. Approximately how many years, **in total**, have you regularly smoked **cigarettes** (not counting the times when you had quit)?

______ years

PLEASE GO TO TOP OF NEXT COLUMN

33. **During most of the time that you regularly smoked, how many cigarettes did you usually smoke each day?**

6 [ ] None
97 [ ] 1-4 per day
197 [ ] 5-14 (1/2 pack) per day
223 [ ] 15-24 (1 pack) per day
68 [ ] 25-34 (1 1/2 packs) per day
58 [ ] 35-44 (2 packs) per day
8 [ ] 45-54 (2 1/2 packs) per day
20 [ ] over 2 1/2 packs per day
3 missing

CHILDHOOD EXPOSURE TO TOBACCO SMOKE

34. **Did your natural mother smoke when she was pregnant with you?**

4201 [ ] Definitely NO
145 [ ] Don't think so
48 [ ] Probably
51 [ ] Definitely YES
12 missing

35. **During any time in your life have you EVER lived for six months or more with someone who smoked?**

2299 [ ] No ➡ skip to question 47
2153 [ ] Yes ➡ continue with question 36
5 missing

36. **When you were a child or teenager (up to 18 years), did you ever live for six months or more with someone who smoked?**

671 [ ] No ➡ skip to question 41
1478 [ ] Yes
4 missing

37. **During what ages of your childhood did you live for six months or more with someone who smoked? (Check all that apply)**

987 [ ] 0-5 years of age
1108 [ ] 6-12 years of age
1110 [ ] 13-18 years of age

38. **During what ages of your childhood did your mother smoke cigarettes (check all that apply).**

1277 [ ] Not at all
112 [ ] 0-5 years of age
143 [ ] 6-12 years of age
143 [ ] 13-18 years of age

PLEASE GO TO TOP OF NEXT PAGE

39. During your CHILDHOOD (up to 18 years) which of the following persons smoked in your home for six months or more? **Check all that apply and for each person estimate the total number of years** that you were exposed to their tobacco smoke.

176 [ ] Mother ➡ ________ YEARS
1265 [ ] Father ➡ ________ YEARS
350 [ ] Others ➡ ________ YEARS

40. **As a child or teenager, during the** majority of these years that you **lived** with someone who smoked tobacco, how many hours per day **on the average** were you exposed to tobacco smoke?

239 [ ] None
463 [ ] Less than 1 hour per day
503 [ ] 1-2 hours per day
150 [ ] 3-5 hours per day
76 [ ] 6-8 hours per day
22 [ ] 9 or more hours per day
34 missing

ADULT EXPOSURE TO TOBACCO SMOKE

41. As an **adult** (19 years of age or over), have you ever **lived** for six months or more with someone who smoked?

728 [ ] No ➡ skip to question 47
1313 [ ] Yes, in the past
108 [ ] Yes, currently
4 missing

42. What was your age as an **adult** when you **first** lived with someone who smoked?

_______ **age in years**

43. What was your age as an **adult** when you **last** lived with someone who smoked?

_______ **age in years**

44. During your **ADULT** years (19 years and older), which of the following persons have smoked in your home for six months or more? **Check all that apply and for each person estimate the total number of years** that you were exposed to their tobacco smoke.

519 [ ] Spouse ➡ ________ YEARS
352 [ ] Others ➡ ________ YEARS

PLEASE GO TO TOP OF NEXT COLUMN

45. **As an adult**, during the majority of these years that you **lived** with someone who smoked tobacco, how many hours per day **on the average** were you exposed to tobacco smoke?

20 [ ] None
72 [ ] Less than 1 hour per day
194 [ ] 1-2 hours per day
310 [ ] 3-5 hours per day
124 [ ] 6-8 hours per day
64 [ ] 9 or more hours per day
2 missing

46. Are cigarettes **currently** smoked in your home?

726 [ ] No
58 [ ] Yes ➡ Approximately how many per day?
2 missing

____

WORK EXPOSURE TO TOBACCO SMOKE

47. Have you **ever** worked where someone smoked in the same room or enclosed space in which you **worked**?

2543 [ ] No ➡ skip to question 52
1721 [ ] Yes, in the past
185 [ ] Yes, currently
8 missing

48. Approximately how many years **in total** have you ever **worked** where someone smoked in the same room or enclosed space in which you worked.

_________ **years**

49. What was your age when you **first worked** with someone who smoked?

_________ **age in years**

50. What was your age when you **last worked** with someone who smoked?

_________ **age in years**

51. During the years that you **worked** where someone smoked tobacco in the same room or enclosed space as you worked, how many hours per day **on the average** were you exposed to tobacco smoke?

10 [ ] None
259 [ ] Less than 1 hour per day
420 [ ] 1-2 hours per day
442 [ ] 3-5 hours per day
769 [ ] 6 or more hours per day
6 missing

PLEASE GO TO TOP OF NEXT PAGE

52. Please estimate the total number of hours per day on the average that you are CURRENTLY exposed to someone else's tobacco smoke. Then do the same for exposure DURING THE LAST TEN YEARS. (Include time in the home, work, in automobiles, public transport and social situations that you are/were exposed to other people's tobacco smoke).

AVERAGE CURRENT EXPOSURE

2927 [ ] None
1120 [ ] Less than 1 hour per day
166 [ ] 1-2 hours per day
100 [ ] 3-5 hours per day
88 [ ] 6-8 hours per day
29 [ ] 9 or more hours per day
27 missing

AVERAGE DURING LAST 10 YEARS

2176 [ ] None
1423 [ ] less than 1 hour per day
362 [ ] 1-2 hours per day
230 [ ] 3-5 hours per day
164 [ ] 6-8 hours per day
39 [ ] 9 or more hours per day
63 missing

ACTIVITIES

53. How many hours per DAY during the work week do you usually spend driving or riding on CROWDED roadways? (Check the nearest category.)

1597 [ ] None
741 [ ] Less than 15 minutes
1536 [ ] 15 minutes to one hour
378 [ ] 2 hours
86 [ ] 3 hours
43 [ ] 4 hours
21 [ ] 5 hours
42 [ ] 6 hours or more
13 missing

54. On a typical WEEKEND, how many hours per day do you spend driving or riding on CROWDED roadways? (Check the nearest category.)

1318 [ ] None
603 [ ] Less than 15 minutes
752 [ ] 15-29 minutes
803 [ ] 30-59 minutes
731 [ ] 1-2 hours
193 [ ] 3-4 hours
31 [ ] 5-6 hours
15 [ ] 6 hours or more
11 missing

PLEASE GO TO TOP OF NEXT COLUMN

55. In total, approximately how many hours per week do you usually spend driving or riding on any type of roadway?

879 [ ] Never ride or drive on a weekly basis

________ hours per week riding or driving on roadways

3578 missing

56. How often do you use aerosol sprays (e.g. hair spray, cleaning spray, deodorant, spray paint, etc.)?

1096 [ ] Daily
641 [ ] Several times a week
397 [ ] Once a week
704 [ ] A few times a month
1606 [ ] Rarely or never
13 missing

57. How long are you usually away from home during the Summer (June through September)?

3194 [ ] 2 weeks or less
984 [ ] 3-4 weeks
144 [ ] 5-6 weeks
44 [ ] 7-8 weeks
49 [ ] 9 weeks or more
42 missing

How many hours per WEEK, including weekends, do you exercise vigorously or do heavy physical labor (e.g. jogging, tennis, gardening, etc.) in the open air?

58. SUMMER (June through September)

981 [ ] None
1585 [ ] 1-3 hours
968 [ ] 4-7 hours
485 [ ] 8-14 hours
197 [ ] 15-21 hours
110 [ ] 22-28 hours
39 [ ] 29-35 hours
49 [ ] 36-42 hours
35 [ ] More than 42 hours
8 missing

59. REST OF YEAR (October through May)

1010 [ ] None
1801 [ ] 1-3 hours
925 [ ] 4-7 hours
391 [ ] 8-14 hours
149 [ ] 15-21 hours
58 [ ] 22-28 hours
46 [ ] 29-35 hours
37 [ ] 36-42 hours
27 [ ] More than 42 hours
13 missing

PLEASE GO TO TOP OF NEXT P.

How many hours per WEEK, including weekends, are you outside of buildings?

60. SUMMER
(June through September)

113 [ ] None
1756 [ ] 1-7 hours
1102 [ ] 8-14 hours
572 [ ] 15-21 hours
375 [ ] 22-28 hours
199 [ ] 29-35 hours
134 [ ] 36-42 hours
189 [ ] More than 42 hours
17 missing

61. REST OF YEAR
(October through May)

148 [ ] None
2123 [ ] 1-7 hours
1066 [ ] 8-14 hours
476 [ ] 15-21 hours
278 [ ] 22-28 hours
131 [ ] 29-35 hours
89 [ ] 36-42 hours
126 [ ] More than 42 hours
20 missing

62. During your usual daily activities how much time do you usually spend close to any sources of combustion, such as heavy traffic, gas powered equipment, lawn mowers, gas stoves or ranges?

1050 [ ] None
995 [ ] Less than 15 minutes
754 [ ] 15-29 minutes
602 [ ] 30-59 minutes
602 [ ] 1-2 hours
172 [ ] 3-4 hours
72 [ ] 5-6 hours
175 [ ] 6 hours or more
35 missing

EMPLOYMENT

63. What is your current employment status?

525 [ ] Unemployed
332 [ ] Working part time
1249 [ ] Working full time
1767 [ ] Retired and not working
432 [ ] Retired and working part time
92 [ ] Retired but working full time
→ IF RETIRED, please give date of retirement

60 missing

____/____
month year

PLEASE GO TO TOP OF NEXT COLUMN

64. What is your usual or main occupation? (Do not write "retired". If retired or not now working, give your usual occupation when you were working.)

Job Title________________

Type of business or industry:

____

Major duties or responsibilities:

____

65. How many years have you been employed in this occupation?

____________ years

66. Are you currently employed in your usual or main occupation?

2480 [ ] No
1799 [ ] Yes → skip to question 68
178 missing

67. When were you last employed in your usual or main occupation?

____/____
month year

68. What type of air cooling system do you currently have at your work place?

2096 [ ] I am not currently working
536 [ ] None
102 [ ] Evaporative cooler (swamp cooler)
1532 [ ] Refrigerating type (air conditioner)
58 [ ] Both
133 missing

69. Since March 1977 what type of air cooling system/systems have you had at your place or places of work?

1220 [ ] I have not worked since March 1977 → skip to question 73
607 [ ] None
136 [ ] Evaporative cooler (swamp cooler)
2176 [ ] Refrigerating type (air conditioner)
190 [ ] Both
128 missing

PLEASE GO TO TOP OF NEXT PAGE

70. Since March 1977, have you changed your occupation or your location of work?

1854 [ ] No
1189 [ ] Yes
64 missing

71. Since March 1977, have you worked for 1 month or more at a location more than 5 miles from home?

1636 [ ] No ⟹ skip to question 73
1425 [ ] Yes
46 missing

72. If you have worked more than 5 miles from home since March 1977, give the work locations and dates.

Started Job: MONTH	YEAR	Ended Job: MONTH	YEAR	TOWN OF WORK	STATE	ZIP CODE OF WORK PLACE

73. Have you moved or been away from home for more than one month since March 1977?

2723 [ ] No ⟹ please skip to question 75
1154 [ ] Yes, I have moved
36 [ ] Yes, I have been away for more than 1 month.
_________ How many months?
44 missing

74. For each community in which you have lived or stayed for one month or more since March 1977, please give the information requested below. For large cities, please give the section of the city. Please start with your residence in 1977 and work towards the present.

TOWN	STATE	ZIP	When did you start living in this town? MONTH	YEAR

PLEASE GO TO TOP OF NEXT PAGE

75. Have you EVER WORKED (for 3 months or more) where you were exposed much of the time as a part of your job to any of the following? If so, write in the approximate number of years you were exposed to each agent and the year when last exposed.

3394 [ ] check here if no regular exposure to any of the below

Number of Years of Exposure	Year when last Exposed	
______	______	Radiation, X-ray
		DUSTS
______	______	Sand or rock dust
______	______	Asbestos
______	______	Talc, graphite
______	______	Fiberglass, rock wool
______	______	Sawdust
______	______	Metal dust
______	______	Road, soil dust
		Other dusts (specify dates & dusts)
______	______	______________
______	______	______________
______	______	______________
		FUMES
______	______	Paint fumes
______	______	Formaldehyde
______	______	Solvents
______	______	Insecticides
______	______	Resins or glues
______	______	Diesel fumes
______	______	Freon or refrigerants
______	______	Auto exhaust
______	______	Solder (or flux) fumes
______	______	Welding fumes/nitrogen oxides
		Other fumes or airborne contaminants (specify dates & fumes)
______	______	______________
______	______	______________
______	______	______________

PLEASE GO TO TOP OF NEXT COLUMN

76. Please specify types of jobs or industries where you had the above exposures.

Industries	Job Duties
______________	______________
______________	______________
______________	______________

77. Would you rate your exposures to dusts at work (past or present) as

1964 [ ] None
1244 [ ] Mild
381 [ ] Moderate
77 [ ] Severe
791 missing

78. Would you rate your exposures to fumes or airborne contaminants at work (past or present) as

2070 [ ] None
1103 [ ] Mild
393 [ ] Moderate
92 [ ] Severe
799 missing

RESIDENCE

79. How many years have you lived in your present home?

______ years

80. What type of building do you live in?

3508 [ ] Single family home detached from any other house
200 [ ] Single family home attached to one or more houses (for example, a townhouse, duplex, triplex)
374 [ ] Mobile home or trailer
345 [ ] Apartment, condominium with more than 3 units.
[ ] Other, please specify____________
30 missing

81. How old is your residence?

59 [ ] Less than 1 year
47 [ ] 1 year old
126 [ ] 2-3 years old
714 [ ] 4-10 years old
802 [ ] 11 to 20 years old
2679 [ ] More than 20 years old
30 missing

82. What is the size of the living area of your home (in square feet)?

____________ sq ft

83. How many bedrooms are there in your home?

____

PLEASE GO TO TOP OF NEXT PAGE

84. How many **bathrooms** are there in your home?

____

85. Where do you park your car(s) when you are home?

1772 [ ] In an attached garage
465 [ ] In an attached carport
764 [ ] Garage or carport detached from your living quarters
1246 [ ] In a driveway, next to living quarters
172 [ ] Other ~~have no car~~
38 missing

86. What type of air cooling system do you have in your home?

1573 [ ] None
400 [ ] Evaporative cooler (swamp cooler)
2386 [ ] Refrigerating type (air conditioner)
75 [ ] Both
23 missing

87. How many years have you had this type of air cooling system in your home?

_______ years

88. What type of air cooling system did you have in your home in March 1977?

2055 [ ] None
517 [ ] Evaporative cooler (swamp cooler)
1797 [ ] Refrigerating type (air conditioner)
29 [ ] Both
59 missing

HEATING AND COOKING

89. How frequently are the following **heating** systems used in your home during the **winter** months? Please check **never** if not used.

Never	Monthly or less	Weekly	Daily	
2003	217	243	1993	Gas forced air furnace
3765	82	73	537	Gas wall furnace
3950	34	49	424	Gas floor furnace
4375	13	7	62	Gas space heater
4386	30	18	23	Kerosene space heater
3088	554	538	277	Fireplace
3906	66	63	422	Wood stove
				Other (specify)
4079	51	53	274	Electric
4431	0	0	25	Oil
4453	0	1	3	Coal
4425	3	3	26	Other fuel
4448	1	0	8	Other non-fuel
4439	0	3	15	Hot water heated by gas

PLEASE GO TO TOP OF NEXT COLUMN

90. How many months of the year do you heat your home?

________ **months**

91. How many **YEARS** during your childhood, (up to 18 years), were each of the following types of heating used in your home? Please check none if never used.

None	1-5	6-10	11-15	16+	
2654	257	217	223	1106	GAS
2039	368	378	340	1332	WOOD
2296	367	329	300	1165	COAL or OIL
4403	20	11	6	17	OTHER fuel
4326	39	21	16	54	Other non-fuel

92. How many **YEARS** during your **adult** life, (19 years and over), have each of the following types of heating been used in your home? Please check none if never used.

None	1-5	6-10	11-20	21+	
499	117	135	323	3383	GAS
3032	456	235	201	533	WOOD
3520	380	167	136	254	COAL or OIL
4414	17	8	3	15	OTHER fuel
4088	108	75	58	128	Other non-

93. Have you ever lived in a home with a gas cooking stove?

290 [ ] No ➡ skip to question 100
2139 [ ] Yes, CURRENTLY
2019 [ ] Yes, IN THE PAST.
Year last used
19_______
9 missing

If you answered CURRENTLY to question 93, answer questions 94–98 for CURRENT use only

If you answered IN THE PAST to question 93, answer questions 94–98 in terms of PAST USE.

94. How is (was) your gas cooking stove lighted?

252 [ ] Light by hand
520 [ ] Electric ignition
3298 [ ] Pilot light ➡ how many pilots in range and oven?
88 missing

____

PLEASE GO TO TOP OF NEXT P.

95. How often do (did) you use an exhaust fan or range hood when food is (was) being prepared on the stove?

1922 [ ] Rarely or never
1576 [ ] Occasionally (when kitchen is smoky or for odors)
316 [ ] At least half the time that the stove is on
264 [ ] Always/almost always, whenever stove is on
80 missing

96. On the average, how many hours per DAY is (was) COOKING done with a GAS stove in your home?

171 [ ] None
1441 [ ] Less than 1 hour per day
2108 [ ] 1-2 hours per day
363 [ ] More than 2 hours per day
75 missing

97. On the average, how many hours per WEEK is (was) BAKING done with a GAS oven in your home?

392 [ ] None
1193 [ ] Less than 1 hour per week
1705 [ ] 1-2 hours per week
675 [ ] 3-5 hours per week
109 [ ] 6 or more hours per week
84 missing

98. During the winter, how frequently is (was) the range or stove used to help heat your house?

3349 [ ] Never
371 [ ] Monthly or less
150 [ ] Weekly
209 [ ] Daily
79 missing

99. For how many years have you lived in a home where a gas cooking stove was used?

_______ years

100. Do you have a gas water heater?

776 [ ] No

IF YES, where is it located?

1505 [ ] Inside the home or in a closet inside the home
1059 [ ] In the garage
44 [ ] Carport
1034 [ ] In a closet accessed from outside the home
[ ] Other (specify) ________________

39 missing

PLEASE GO TO TOP OF NEXT COLUMN

DIET AND MISCELLANEOUS

101. How often do you currently eat meat, poultry or fish when you are following your usual routine?

1233 [ ] Never
535 [ ] Less than once per MONTH
571 [ ] 1-2 times per MONTH
865 [ ] 1-2 times per WEEK
654 [ ] 3-4 times per WEEK
261 [ ] 5-6 times per WEEK
268 [ ] Once per DAY
56 [ ] More than once per DAY
14 missing

102. Has your use of meat, poultry or fish changed since 1976?

1004 [ ] Never used in 1976 or after
1683 [ ] Decreased since 1976
1439 [ ] No change since 1976
293 [ ] Increased since 1976
38 missing

103. How often do you currently drink DECAFFEINATED coffee when you are following your usual routine. Please note that the choices refer to the number of "TIMES" you use coffee (NOT the number of cups).

2420 [ ] Never
807 [ ] Less than once per WEEK
444 [ ] 1-3 times per WEEK
159 [ ] 4-6 times per WEEK
401 [ ] Once per DAY
137 [ ] 2 times per DAY
41 [ ] 3 times per DAY
15 [ ] 4 times per DAY
5 [ ] 5 times per DAY
2 [ ] Over 5 times per DAY
26 missing

104. How often do you currently drink REGULAR (NOT decaffeinated) coffee when you are following your usual routine. Please note that the choices refer to the number of "TIMES" you use coffee (NOT the number of cups).

2690 [ ] Never
486 [ ] Less than once per WEEK
272 [ ] 1-3 times per WEEK
177 [ ] 4-6 times per WEEK
478 [ ] Once per DAY
188 [ ] 2 times per DAY
95 [ ] 3 times per DAY
27 [ ] 4 times per DAY
15 [ ] 5 times per DAY
11 [ ] Over 5 times per DAY
18 missing

PLEASE GO TO TOP OF NEXT PAGE

105. If you drink coffee, on the average how many CUPS do you usually drink at one sitting? Don't forget refills at coffee shops or restaurants.

2409 [ ] Never drink coffee
1227 [ ] 1 Cup
674 [ ] 2 Cups
86 [ ] 3 Cups
13 [ ] 4 or more cups
48 missing

106. If you drink coffee, what size cup do you usually use? The average cup contains 8 ounces, but some cups are larger or smaller.

2352 [ ] Never drink coffee
735 [ ] 6 ounce cup or smaller
1155 [ ] 8 ounce cup
106 [ ] 10 ounce cup
11 [ ] 12 ounce cup
5 [ ] 16 ounce cup or larger
93 missing

107. Has your use of **DECAFFEINATED** coffee changed since 1976?

2242 [ ] Never used in 1976 or after
614 [ ] Decreased since 1976
960 [ ] No change since 1976
531 [ ] Increased since 1976
110 missing

108. Has your use of **REGULAR** coffee changed since 1976?

2365 [ ] Never used in 1976 or after
1008 [ ] Decreased since 1976
748 [ ] No change since 1976
251 [ ] Increased since 1976
85 missing

109. How often do you currently drink **soft drinks containing caffeine** (such as Pepsi, Coca-Cola, Dr. Pepper, Mountain Dew, etc.)?

1956 [ ] Never
923 [ ] Less than once per MONTH
600 [ ] 1-2 times per MONTH
466 [ ] 1-2 times per WEEK
220 [ ] 3-4 times per WEEK
101 [ ] 5-6 times per WEEK
114 [ ] Once per DAY
68 [ ] More than once per DAY
9 missing

110. Has your use of **soft drinks containing caffeine** changed since 1976?

1764 [ ] Never used in 1976 or after
1174 [ ] Decreased since 1976
175 [ ] No change since 1976
275 [ ] Increased since 1976
69 missing

PLEASE GO TO TOP OF NEXT COLUMN

111. What is your best estimate of your present weight in normal indoor clothing WITHOUT shoes?

_________ pounds

112. What is your sex?

1572 [ ] Male
2882 [ ] Female
3 missing

113. What is your date of birth?

_____/_____/_____
month day year

114. What is your current affiliation with the Seventh-day Adventist (SDA) Church?

4130 [ ] Current baptized member
16 [ ] Affiliated, but never baptized
264 [ ] Former SDA
32 [ ] No affiliation
15 missing

115. Have you been hospitalized since January 1, 1983?

3095 [ ] No
1334 [ ] Yes
28 missing

116. Since January 1, 1983, has a doctor told you (FOR THE FIRST TIME) that you had a tumor or cancer of any kind?

4121 [ ] No
323 [ ] Yes ➜ please give location or type
13 missing of tumor ____________________

[redacted] and date of diagnosis ____/____
month year

117. Since January 1, 1983 has a doctor told you (FOR THE FIRST TIME) that you had a Heart Attack or "Coronary" (Myocardial Infarction)?

4277 [ ] No
151 [ ] Yes ➜ Date of diagnosis ____/____
29 missing month year

MEN, This is the end of your questionnaire.

Please check the questionnaire to make sure that you have not left any questions blank which should have been answered.

Thank you very much for completing the questionnaire. Please fold it and mail it in the return envelope as soon as possible.

WOMEN, please continue with the WOMEN ONLY questions on the attached pink page.

WOMEN, please answer all the questions on both sides of this pink page.

118. Are you currently pregnant or nursing a baby?

2827 [ ] No
8 [ ] Yes
47 missing

119. Has your uterus been removed surgically?

1557 [ ] No
1271 [ ] Yes ➡ AT WHAT AGE _____
54 missing

120. Have one or both ovaries been removed surgically?

1937 [ ] No
99 [ ] Not sure, but had surgery near my ovaries.
62 [ ] Had surgery on ovaries but not sure whether one or both were removed
526 [ ] Both ovaries removed ➡ AT WHAT AGE(S)__________
197 [ ] Only one ovary removed ➡ AT WHAT AGE__________
61 missing

121. Are you currently having menstrual periods?

50 [ ] Yes, regularly
71 [ ] Yes, irregularly
353 [ ] Menstrual periods have completely stopped due to natural change of life
1066 [ ] Past surgery stopped my menstrual periods
142 missing

122. Have you EVER taken birth control pills (oral contraceptives)?

2286 [ ] No ➡ skip to question 128
538 [ ] Yes
58 missing

123. When did you **FIRST** take birth control pills _____/_____
month year

124. When did you **LAST** take birth control pills _____/_____
month year

125. Approximately how many **years, in total**, did you take birth control pills (sum all years of use, but do not include gaps in between different periods of use)?
_________ **years**

126. What was the brand name of the **LAST** birth control pill that you used?____________________

127. For how many **years** did you use this particular brand? ________ **years**

PLEASE GO TO TOP OF NEXT PAGE

APPENDIX B

Appendix B

Final Report

Description of Respiratory Symptoms Algorithms, Severity Scores and Reliability Checks.

Exhibit 1 contains an English description of the computer algorithms used to classify individuals as having none, possible, or definite symptoms for each respiratory symptoms complex - overall AOD, chronic bronchitis (cough type), chronic bronchitis (sputum type), emphysema, and asthma. Individuals having definite symptoms of one or more of the respiratory complexes were defined as having definite airway obstructive disease. The question numbers in parenthesis in Exhibit 1, referred to the 1987 AHSMOG questionnaire, see Appendix A. Exhibit 2 contains the computer statements of the algorithms, again question numbers refer to the 1987 questionnaire, a code of "-1" means "missing." Table 1 contains the frequency distributions in 1977 and 1987 for each grade of each respiratory symptoms complex for those individuals who completed both the 1977 and 1987 questionnaires.

Table 2 contains gender and age specific rates for each type of respiratory symptoms complex for:

- prevalence in 1977
- prevalence in 1987 cumulative incidence rate = number of new cases between 1977 and 1987 divided by number who did not have definite symptoms in 1977
- reversal rate = number of individuals who had definite symptoms in 1977 who did not have definite symptoms in 1987, divided by number of individuals who had definite symptoms in 1977
- persistent prevalence rate = definite symptoms in 1977 and 1987 divided by number of individuals answering both questionnaires.

An earlier version of this Table was used to conduct reliability checks on questionnaire reporting of respiratory symptoms. The questionnaires of individuals reporting a reversal of asthma were carefully examined to determine if reliable responses were obtained that were consistent between the 1977 and 1987 questionnaires. Individuals showing inconsistent responses such as a "Yes" response to "Has a doctor every told you that you have asthma," in 1977 but a "No" response to that question in 1987 were excluded from the analysis cohort. Individuals reporting a reversal of emphysema were also excluded, as clinically, emphysema is thought to be irreversible.

Table 1 Frequency Distributions for Respiratory Symptoms in 1977 and 1987 for Those Subjects Completing both the 1977 and 1987 Questionnaire.

Frequency Distributions for Respiratory Symptoms in 1977 and 1987 for Those Subjects Completing both the 1977 and 1987 Questionnaires

Chronic Bronchitis (Cough type) 1977:

Grade	Frequency	Percent
0	3560	79.9
1	587	13.2
2	82	1.8
3	62	1.4
4	95	2.1
5	61	1.4
Missing	10	0.2
Totals	4457	100.0

Chronic Bronchitis (Cough type) 1987:

Grade	Frequency	Percent
0	3356	75.3
1	746	16.7
2	80	1.8
3	82	1.8
4	105	2.4
5	85	1.9
Missing	3	0.1
Totals	4457	100.0

Table 1 **Frequency Distributions for Respiratory Symptoms in 1977 and 1987 for Those Subjects Completing both the 1977 and 1987 Questionnaire.** (continued...)

Chronic Bronchitis (Sputum type) 1977:

Grade	Frequency	Percent
0	3586	80.5
1	488	10.9
2	103	2.3
3	99	2.2
4	116	2.6
5	51	1.1
Missing	14	0.3
Totals	4457	100.0

Chronic Bronchitis (Sputum type) 1987:

Grade	Frequency	Percent
0	3427	76.9
1	638	14.3
2	93	2.1
3	93	2.1
4	122	2.7
5	73	1.6
Missing	11	0.2
Totals	4457	100.0

Asthma 1977:

Grade	Frequency	Percent
0	4151	93.1
1	76	1.7
2	107	2.4
3	79	1.8
4	38	0.9
Missing	6	0.1
Totals	4457	100.0

Table 1 Frequency Distributions for Respiratory Symptoms in 1977 and 1987 for Those Subjects Completing both the 1977 and 1987 Questionnaire. (continued...)

Asthma 1987:

Grade	Frequency	Percent
0	3969	89.1
1	222	5.0
2	105	2.4
3	88	2.0
4	73	1.6
Missing	0	0.0
Totals	4457	100.0

Emphysema 1977:

Grade	Frequency	Percent
0	4401	98.7
1	21	0.5
2	15	0.3
3	15	0.3
Missing	5	0.1
Totals	4457	100.0

Emphysema 1987:

Grade	Frequency	Percent
0	4380	98.3
1	24	0.5
2	26	0.6
3	27	0.6
Missing	0	0.0
Totals	4457	100.0

Table 1 Frequency Distributions for Respiratory Symptoms in 1977 and 1987 for Those Subjects Completing both the 1977 and 1987 Questionnaire. (continued...)

AOD Collapsed Variable 1977:

Grade	Frequency	Percent
0	3062	68.7
1	710	15.9
2	669	15.0
Missing	16	0.4
Totals	4457	100.0

AOD Collapsed Variable 1987:

Grade	Frequency	Percent
0	2784	62.5
1	932	20.9
2	733	16.4
Missing	8	0.2
Totals	4457	100.0

Table 2. Prevalence, Incidence, Reversal and Persistent Prevalence (Yes in '77 and '87) of Self-Reported Respiratory Symptoms Complexes for All Males and Females, 1977-1987.

		All Males - Age in 1977			All Females - Age in 1977		
		25-54	55-74	75 +	25-54	55-74	75 +
	n* =	644	707	61	1130	1230	137
Cough COPD	Prevalence in 1977	0.047	0.061	0.016	0.059	0.065	0.088
	Prevalence in 1987	0.062	0.095	0.066	0.074	0.074	0.066
	Cum Incidence Rate	0.038	0.068	0.067	0.040	0.045	0.032
	Reversal Rate	0.433(30)	0.488(43)	1.000(1)	0.373(67)	0.506(79)	0.583(12)
	Yes in '77 & '87	0.027	0.031	0.000	0.037	0.032	0.036
Sputum COPD	Prevalence in 1977	0.095	0.089	0.082	0.070	0.063	0.081
	Prevalence in 1987	0.102	0.102	0.148	0.075	0.084	0.074
	Cum Incidence Rate	0.066	0.061	0.107	0.040	0.049	0.040
	Reversal Rate	0.557(61)	0.476(63)	0.400(5)	0.468(79)	0.390(77)	0.545(11)
	Yes in '77 & '87	0.042	0.047	0.049	0.037	0.038	0.037
Asthma	Prevalence in 1977	0.033	0.021	0.000	0.047	0.031	0.022
	Prevalence in 1987	0.048	0.047	0.016	0.074	0.045	0.044
	Cum Incidence Rate	0.018	0.026	0.016	0.031	0.014	0.030
	Reversal Rate	0.048(21)	0.000(15)	****	0.038(53)	0.000(38)	0.333(3)
	Yes in '77 & '87	0.031	0.021	0.000	0.045	0.031	0.015
Emphysema	Prevalence in 1977	0.003	0.003	0.000	0.002	0.007	0.000
	Prevalence in 1987	0.006	0.011	0.016	0.006	0.015	0.015
	Cum Incidence Rate	0.003	0.009	0.016	0.004	0.009	0.015
	Reversal Rate	0.000(14)	0.000(2)	****	0.000(2)	0.000(8)	****
	Yes in '77 & '87	0.003	0.003	0.000	0.002	0.007	0.000
All AOD	Prevalence in 1977	0.141	0.126	0.082	0.127	0.111	0.103
	Prevalence in 1987	0.165	0.162	0.180	0.163	0.143	0.140
	Cum Incidence Rate	0.093	0.098	0.043	0.084	0.075	0.082
	Reversal Rate	0.400(90)	0.393(89)	0.400(5)	0.294(143)	0.316(136)	0.357(14)
	Yes in '77 & '87	0.085	0.077	0.049	0.090	0.076	0.066

**** Denominator = 0   ( ) = Denominator   * varies slightly among diseases

Algorithms for Respiratory Symptoms

The algorithms are first written out in English and then given in computer statements.

NOTE: [ ] establish precedence in logic statements. OR always means AND/OR. Question numbers refer to 1987 Questionnaire.

Chronic Bronchitis (Cough type)

None:

Grade 0 - Severity Score 0

No cough first thing in morning (Q2=2) AND no cough at other times (Q3=2) AND no cough on most days for 3 months or more (Q4=2) AND never had a cough for years (Q5=1).

Possible:

Grade 1 - Severity Score 3 (one of 3 sets of symptoms a, b, or c)

a) [Cough first thing in morning (Q2=1) OR cough at other times] (Q3=1) AND no cough on most days for 3 months or more (Q4=2).

b) No cough on most days for 3 months or more (Q4=2) AND [cough for more than 1 but less than 2 years (Q5=3) OR cough for 2 to 5 years (Q5=4) or cough for more than 5 years (Q5=5)].

c) [No cough first thing in morning (Q2=2) OR no cough at other times (Q3=2)] AND [cough less than a year (Q5=2) OR cough more than 1 but less than 2 years (Q5=3)].

Definite: cough greater than 2 years (Q5= >3) AND Grade 2, 3, 4, or 5.

Grade 2 - Severity Score 4

[Cough first thing in morning (Q2=1) OR cough at other times (Q3=1)] AND cough on most days for 3 months or more (Q4=1).

Grade 3 - Severity Score 6

Cough first thing in morning (Q2=1) AND cough at other times (Q3=1) AND cough on most days for 3 months or more (Q4=1).

Grade 4 - Severity Score 8

[Cough first thing in morning (Q2=1) OR cough at other times (Q3=1)] AND cough on most days for 3 months or more (Q4=1) AND shortness of breath with light to moderate activity (Q15=1) AND no shortness of breath with normal walking (Q16=2).

Grade 5 - Severity Score 10

[Cough first thing in morning (Q2=1) OR cough at other times (Q3=1)] AND cough on most days for 3 months or more (Q4=1) AND shortness of breath with normal walking (Q16=1).

Chronic Bronchitis (Sputum Type)

None:

Grade 0 - Severity Score 0

No sputum first thing in morning (Q6=2) AND no sputum at other times of day (Q7=2) AND no sputum for most days for 3 months or more (Q8=2) AND never raise sputum for any length of time (Q9=1).

Possible:

Grade 1 - Severity Score 3 (one of 3 sets of symptoms a, b, or c)

a) [Sputum first thing in morning (Q6=1) OR sputum at other times of day (Q7=1)] AND no sputum for most days for 3 months or more (Q8=2).

b) [Sputum more than 1 but less than 2 years (Q9=3) OR sputum 2 to 5 years (Q9=4) OR sputum more than 5 years (Q9=5)] AND no sputum for most days for 3 months or more (Q8=2).

c) [Sputum less than 1 year (Q9=2) OR sputum more than 1 but less than 2 years (Q9=3)] AND [no sputum first thing in morning (Q6=2) OR no sputum at other times of day (Q7=2)].

Definite:

Sputum production for more than 2 years (Q9=3) AND Grade 2, 3, 4, or 5.

Grade 2 - Severity Score 4

[Sputum first thing in morning (Q6=1) OR sputum at other times of day (Q7=1)] AND sputum on most days for 3 months of year or more (Q8=1).

Grade 3 - Severity Score 6

[Sputum first thing in morning (Q6=1) AND sputum at other times of day (Q7=1)] AND sputum on most days for 3 months of year or more (Q8=1).

Grade 4 - Severity Score 8

[Sputum first thing in morning (Q6=1) OR sputum at other times of day] (Q7=1) AND sputum on most days for 3 months of year or more (Q8=1) AND shortness of breath with light to moderate activity (Q15=1) AND no shortness of breath with normal walking (Q16=2).

Grade 5 - Severity Score 10

[Sputum first thing in morning (Q6=1) OR sputum at other times of day (Q7=1)] AND sputum on most days for 3 months of year or more (Q8=1) AND shortness of breath with normal walking (Q16=1).

Asthma

None:

Grade 0 - Severity Score 0

Doctor has never diagnosed asthma (Q19A NE 1 OR Q20A NE 1) AND breathing never sounds wheezy (Q10=2) AND have never had attacks of shortness of breath with wheezing (Q12=2). (NE = not equal).

Possible:

Grade 1 - Severity Score 3 (one of the following two conditions)

a) Doctor has never diagnosed asthma (Q19A NE 1 or Q20A NE 1) AND breathing has sounded wheezy (Q10=1) AND have had attacks of shortness of breath with wheezing (Q12-1).

b) Doctor has diagnosed asthma (Q19A EQ 1 OR Q20A EQ 1) AND breathing has not sounded wheezy (Q10=2) AND have had no attacks of shortness of breath with wheezing(Q12=2).

Definite:

Doctor has diagnosed asthma (Q19A=1 OR Q20A=1) AND grade 2, 3, or 4.

Grade 2 - Severity Score 5

Breathing has sounded wheezy (Q10=1) OR have had attacks of shortness of breath with wheezing (Q12=1).

Grade 3 - Severity Score 8

[Breathing has sounded wheezy (Q10=1) OR have had attacks of shortness of breath with wheezing] (Q12=1) AND shortness of breath with light to moderate activity (Q15=1) AND no shortness of breath with normal walking (Q16=2).

Grade 4 - Severity Score 10

[Breathing has sounded wheezy (Q10=1) OR have had attacks of shortness of breath with wheezing] (Q12=1) AND shortness of breath with normal walking (Q16=1).

Emphysema

None:

Grade 0 - Severity Score 0

Doctor has never diagnosed emphysema (Q19C NE 1 OR Q20E NE 1) AND not short of breath with light to moderate activity (Q15=2) AND not short of breath with normal walking (Q16=2).

Possible:

Grade 1 - Severity Score 3

Doctor has diagnosed emphysema (Q19C EQ 1 or Q20E EQ 1) AND not short of breath with light to moderate activity (Q15=2) AND not short of breath with normal walking (Q16=2).

Definite:

Doctor has diagnosed emphysema (Q19C=1 OR Q20E=1) AND grade 2 or 3.

Grade 2 - Severity Score 8

Short of breath with light to moderate activity (Q15=1) AND no shortness of breath with normal walking (Q16=2).

Grade 3 - Severity Score 10

Short of breath with normal walking (Q16=1).

COMPUTER ALGORITHMS FOR RESPIRATORY SYMPTOMS

```
GRADE 0   EACH PERSON IS GIVEN A GRADE OF 0 FOR THAT VARIABLE IF THEY
          DO NOT FIT INTO ANY OF THE OTHER CATAGORIES
```

<u>CHRONIC BRONCHITIS (COUGH TYPE), VARIABLE = CCOPD</u>

```
MISSING
IF (Q2 LT '1' AND Q3 LT '1' AND Q4 LT '1' AND Q5 LT '1') CCOPD = -1

IF ((Q2 LT '1' AND Q3 LT '1') OR Q4 LT '1') CCOPD = -1

GRADE 1  (POSSIBLE)
IF ((Q2 EQ '1' OR Q3 EQ '1') AND Q4 EQ '2') CCOPD = 1

IF  ((Q4 EQ '2' AND Q5 EQ '3') OR Q5 EQ '4' OR Q5 EQ  '5') CCOPD = 1

IF ((Q2 EQ '2' OR Q3 EQ '2') AND (Q5 EQ '2' OR Q5 EQ '3')) CCOPD = 1

GRADE 2  (DEFINITE)
IF (Q5 GT '3' AND (Q2 EQ '1' OR Q3 EQ '1') AND Q4 EQ '1') CCOPD = 2

GRADE 3  (DEFINITE)
IF (Q5 GT '3' AND Q2 EQ '1' AND Q3 EQ '1' AND Q4 EQ '1') CCOPD = 3

GRADE 4  (DEFINITE)
IF (Q5 GT '3' AND (Q2 EQ '1' OR Q3 EQ '1') AND Q4 EQ '1' AND Q15 EQ
  '1' AND Q16 EQ '2') CCOPD = 4

GRADE 5  (DEFINITE)
IF (Q5 GT '3' AND (Q2 EQ '1' OR Q3 EQ '1') AND Q4 EQ '1'
    AND Q16 EQ '1') CCOPD = 5
```

<u>CHRONIC BRONCHITIS (SPUTUM TYPE), VARIABLE = SCOPD</u>

```
MISSING
IF (Q6 LT '1' AND Q7 LT '1' AND Q8 LT '1' AND  Q9 LT '1') SCOPD = -1
IF ((Q6 LT '1' AND Q7 LT '1') OR Q8 LT '1') SCOPD = -1

GRADE 1  (POSSIBLE)
IF ((Q6 EQ '1' OR Q7 EQ '1') AND Q8 EQ '2') SCOPD = 1
IF (Q9 GT '2' AND Q8 EQ '2') SCOPD = 1
IF ((Q9 EQ '2' OR Q9 EQ '3') AND (Q6 EQ '2' OR Q7 EQ '2')) SCOPD = 1

GRADE 2  (DEFINITE)
IF (Q9 GT '3' AND (Q6 EQ '1' OR Q7 EQ '1') AND Q8 EQ '1') SCOPD = 2

GRADE 3  (DEFINITE)
IF (Q9 GT '3' AND Q6 EQ '1' AND Q7 EQ '1' AND Q8 EQ '1') SCOPD = 3
```

```
GRADE 4  (DEFINITE)
IF (Q9 GT '3' AND (Q6 EQ '1' OR Q7 EQ '1') AND Q8 EQ '1' AND Q15 EQ
   '1' AND Q16 EQ '2') SCOPD = 4

GRADE 5  (DEFINITE)
IF (Q9 GT '3' AND (Q6 EQ '1' OR Q7 EQ '1') AND Q8 EQ '1'
     AND Q16 EQ '1') SCOPD = 5
```

<u>ASTHMA, VARIABLE = ACOPD</u>

```
MISSING
IF (Q19 LT '1' AND Q19A LT '1' AND Q19B LT '1' AND Q19C LT '1' AND
     Q20 LT 1 AND Q20A LT 1 AND Q20B LT 1 AND Q20C LT 1 AND Q20D LT 1
     AND Q20E LT 1 AND Q20F LT 1) ACOPD = -1
IF (Q10 LT '1' AND Q12 LT '1') ACOPD = -1

GRADE 1  (POSSIBLE)
IF ((Q19A NE '1' AND Q20A NE  1 ) AND (Q10 EQ '1' AND Q12 EQ '1'))
      ACOPD = 1
IF (Q19A EQ '1' OR Q20A EQ 1) ACOPD = 1

GRADE 2  (DEFINITE)
IF ((Q19A EQ '1' OR Q20A EQ  1 ) AND (Q10 EQ '1' OR Q12 EQ '1'))
      ACOPD = 2

GRADE 3  (DEFINITE)
IF ((Q19A EQ '1' OR Q20A EQ  1 ) AND (Q10 EQ '1' OR Q12 EQ '1')
      AND Q15 EQ '1' AND Q16 NE '1') ACOPD = 3

GRADE 4  (DEFINITE)
IF ((Q19A EQ '1' OR Q20A EQ  1 ) AND (Q10 EQ '1' OR Q12 EQ '1')
      AND Q16 EQ '1')  ACOPD = 4
```

<u>EMPHYSEMA, VARIABLE = EMPH</u>

```
MISSING
IF ((Q19 LT '1' AND Q19A LT '1'  AND Q19B LT '1' AND Q19C LT '1')
      AND Q20 LT 1 AND Q20A LT 1 AND Q20B LT 1 AND Q20C LT 1 AND
      Q20D LT 1 AND Q20E LT 1 AND Q20F LT 1) EMPH = -1

GRADE 1  (DEFINITE)
IF ((Q19C EQ '1' OR Q20E EQ 1) AND Q15 NE '1' AND Q16 NE '1')
      EMPH = 1

GRADE 2  (DEFINITE)
IF ((Q19C EQ '1' OR Q20E EQ 1) AND Q15 EQ '1' AND Q16 NE '1')
      EMPH = 2

GRADE 3  (DEFINITE)
IF ((Q19C EQ '1' OR Q20E EQ 1) AND Q16 EQ '1') EMPH = 3
```

APPENDIX C

Appendix C

Notes on the NO_2 Indoor Adjustment Algorithm

Following the introduction, this appendix contains eight (8) exhibits:

Exhibit 1. "Mnemonic Variable Names Used in Indoor Adjustment Equations."

Exhibit 2. "Candidate Questions from the GRI Questionnaire."

Exhibit 3. "Candidate Questions from the 1987 AHSMOG Questionnaire."

Exhibit 4. "Coding of Original Variable Used in GRI Study."

Exhibit 5. "Coding of Mnemonic Variables from 1987 AHSMOG Questions and Manner of Employment in Regression Equations."

Exhibit 6. "Final Model Equations for Adjusting NO_2 Mean Concentration."

Exhibit 7. "Comparison of Frequency Distributions for Candidate Regression Variables Between AHSMOG and GRI Study Subjects."

Exhibit 8. Scatterplots of GRI regression residuals from adjusted mean concentration NO_2 regression equations.

Appendix C

Notes on the NO_2 Indoor Adjustment Algorithm

Introduction:

Estimating average exposure to NO_2 from fixed site monitors of ambient concentrations can be confounded by indoor variations in ambient NO_2 levels. The thought was that maybe if appropriate adjustments could be made to the ambient mean concentration levels, some of these variations could be explained and actual exposures more accurately reflected. In the development stages of the 1987 AHSMOG Questionnaire, the researchers at Loma Linda University became aware of the work of the Gas Research Institute (GRI) study in estimating to NO_2 exposure of individuals living in southern California. Selected questions from the GRI questionnaire thought to be most influential in influencing exposure were incorporated into the 1987 AHSMOG Questionnaire. The questions chosen were modified from the personal interview format of the GRI study to the mailed questionnaire format of the AHSMOG study. The changes made to the GRI questions may be seen by comparing Exhibit 2 with Exhibit 3, using the mnemonic variable names to link the two together. The major difference to be noted is that whereas the GRI study involved two days, the AHSMOG study involved (then) ten years. For example, the GRI interview asked the number of minutes one spend doing certain activities which influenced exposure, whereas for the AHSMOG study the average number of hours per week, over a ten year period was estimated by subjects.

The housing characteristics and personal activity questions borrowed from the GRI study were not available for the 1977 AHSMOG Questionnaire, so adjustments made on the bases of these questions could only be calculated backwards as far as the last move a person made to their present location. Because the regression models could only be applied since the last move, the number of months that NO_2 could be adjusted for any individual that had moved was reduced, and in many cases, significantly so.

The tables and exhibits that follow are to enable the reader to reconstruct the method of developing and applying the indoor adjustments to NO_2 ambient mean concentration. The adjustment model for the AHSMOG study was developed using questions from the GRI study and values of NO_2 measured by personal monitors for 582 people in age groups 20 years of age and older from the GRI study, who had usable data. There was a total of 650 subjects in the GRI study. Those younger than 20 years old were not included in the comparisons since all subjects in the AHSMOG study were 25 years or older in 1976. Age of subjects in the GRI study was only available by decade; hence the lower point of 20 for the GRI study. The process of developing the model equations has been described in paper 10 and is partially recounted with modifications in the following two paragraphs.

A regression estimation procedure was used to adjust ambient concentrations estimates for the 3,914 AHSMOG study participants of the respiratory symptoms cohort who completed the 1977 and 1987 questionnaires using data collected on building characteristics and individual activity patterns in 1987. The ambient concentration estimates for the AHSMOG study participants consisted of monthly averages interpolated from fixed site monitoring stations to the zip code centroids of subjects' residences. The GRI data consisted of a similar questionnaire data, plus additional personal and ambient NO_2 exposure data, were collected on a sample of 582 subjects from Los Angeles and Orange Counties (Spengler, et al, 1992). The regression estimation equations were formed for this sample of 582 subjects and then applied to the AHSMOG cohort.

The exposure data on the 582 subjects involved two consecutive 24-hour samples for each of the subjects on randomly selected days during the years 1987 and 1988. Outdoor ambient monitors and bedroom micro-environmental monitors were placed in each study participant's home and were set to sample over the 48-hour period of personal monitoring. The data from the bedroom micro-environmental monitors were not used in forming the regression equations, since such data were not available for the AHSMOG cohort. The study participants wore passive diffusion badges which enabled estimation of average concentration of NO_2 during each of the 24-hour measuring periods and for the 48-hour sampling duration. Study participants kept activity diaries and the household characteristics were reported.

Stepwise multiple regression procedures were used to select the best candidate predictor variables from the large pool of possible common candidate variables shown in Exhibit 1. Several a priori modifications were made to the variables before they were applied in the models. Variables we considered to be potential sources of indoor residential NO_2 were treated as interactive terms and were multiplied by the fraction of time spent indoors. Variables that might modify indoor exposure from outdoor ambient air (e.g., air conditioning) were treated as interactive product terms of the fraction of time spent indoors and the outdoor ambient NO_2 concentration. Finally, the hours spent outside were treated as a product of average outdoor concentration multiplied by the fraction of time spent outside.

Since a number of the candidate variables applied only to homes with gas ranges, separate regression models were used for individuals living in homes with gas ranges. In initial stepwise selection procedures for individuals living in homes with gas ranges, variables describing the heating type used in the home were entered. Hence we considered two separate regression models for individuals with gas ranges -- one for the "heating season" and one for the "non-heating" season. The "heating season" was defined as the months of December, January, February, and March; the Generalized F test was used to determine if season specific regression models were needed for subjects living in homes with gas ranges. The Generalized F test indicated a statistically significant improvement of the fit for the season specific models ($p < 0.05$). Based on this result three separate multiple regression prediction models were constructed: the first for individuals living in homes without gas ranges; the second for individuals living in homes with gas ranges for the heating season; and a third model for individuals living in homes with gas ranges for the non-heating season. The three regression models are shown in Exhibit 6.

Paper 10, from which the substance of the above abstract was taken, speaks of two variables called FOM (the fraction of time spent outdoors multiplied by the ambient mean concentration) and FIM (the fraction of time spent indoors multiplied by the ambient mean concentration). These variable names correspond to HRSOUTM and FF, respectively, in this appendix.

The details of how the AHSMOG variables were coded and the algorithms which were used to apply the regression prediction equations to the AHSMOG cohort are contained in Exhibit 5, "Coding of Mnemonic Variables from 1987 AHSMOG Questions and Manner of Employment in Regression Equations." In brief the method was to translate the questionnaire data for each individual into codes for each candidate variable for each month that one lived in one's latest residence. Then this file was combined with the month by month average exposure levels for ambient mean concentration of NO_2 as interpolated from fixed site monitoring stations to compute the monthly average adjusted mean concentration of NO_2, and these monthly values summed for the period the data was available. Flags were included in the output file to indicate the interpolation quality of NO_2 data. A description of each of the exhibits follows.

Exhibit 1, Mnemonic Variable Names Used in Indoor Adjustment Equations, shows the relationship between the GRI variable names used in developing the regression equations and the mnemonic variable names used in the actual adjustment equations. Also in this exhibit are the question numbers from the 1987 AHSMOG Questionnaire which were used in the construction of the corresponding mnemonic variable. The "M" at the end of a variable name means it was modified, typically, multiplied by the fraction of time spent indoors, if it involved indoor activities, and by the fraction of time spent outdoors, if it involved outdoor activities. One mnemonic variable, FF was the product of the fraction of time spent indoors times the ambient mean concentration of NO_2. The details of exactly how these variables were modified are given in Exhibit 6. Exhibit 1 is the key to tying together the GRI variable names (Exhibit 4) with the rest of the exhibits and tables.

Exhibit 2, Candidate Questions from the GRI Questionnaire, shows the actual questions asked in the GRI questionnaire which were used to form variables for use in the regression equations. Where two questions were combined to form a group of dummy variables, the second question does not have the variable names to its left, as do the single question variables. For example, the question, "What is the main heating system in your home?" which has no mnemonic variable labels to its left, is combined with the question which precedes it, namely, "What is the main type of fuel used to heat your home?" which has mnemonic variables to its left. The manner of creating the GRI variables from the questions is given in Exhibit 4.

Exhibit 3, Candidate Questions from the 1987 AHSMOG Questionnaire, shows the actual questions asked in the 1987 AHSMOG questionnaire which were used to form variables for use in the regression equations. Where two questions were combined to form a variable, or a group of dummy variables, the second question does not have a title above it. For example, the variable name HRSOUTM is based on the combination of questions 60 and 61 under the main question, "How many hours per WEEK, including weekends, are you outside of buildings?" The manner of creating the mnemonic variables from the questions is given in Exhibit 5. The numbers to the left of the response boxes in the questionnaire are the frequencies of responses for each of these questions. Note that the questions do not add up the same number. Some questions allow multiple responses, such as question 89, and some questions were to be skipped (see the note at the end of question 93).

Exhibit 4, Coding of Original Variable Used in GRI Study, lists the variables used by IES in forming the regression equations, and gives a brief description of how the candidate questions shown in Exhibit 2, were converted to these variables. The first two variables shown came from the monitored values, not from the questionnaire. FCONC7 was the average hourly NO_2 concentration monitored outside the house, while FCONC12 was the two-day average personal NO_2 exposure derived from the passive diffusion badges worn on the individual. Note that the units are parts per billion (ppb) rather than the parts per hundred million (pphm) used in the regression equations displayed in Exhibit 6. The regression coefficients in Exhibit 6 which had to be modified accordingly are there indicated by an asterisk (*).

Exhibit 5, Coding of Mnemonic Variables from 1987 AHSMOG Questions and Manner of Employment in Regression Equations, lists the sources of information used in the formation of each of the mnemonic variables. The question numbers given in this exhibit refer to the questions in the 1987 AHSMOG questionnaire which are displayed in Exhibit 3. The first set of variables were common to all three regression equations, or were used to modify other variables. The second set of variables were used only in the seasonal gas regression equations. The third set of variables were candidates for the regression equations but did not come into any final model. Note that none of the variables described in this exhibit end with an "M." The modifications which caused them to have an appended "M" are given in Exhibit 6. The section following the variable descriptions explains further details of how residence history affected the application of the models, how the appropriate model was ascertained, and how the seasons were determined.

Exhibit 6, Final Model Equations for Adjusting NO_2 Mean Concentration, displays the actual equations used in the computations of adjusted mean NO_2 concentration. The explanation of the "M" suffix to the variable names given in Exhibit 5 are shown as equations. These are primarily modifications to reflect the proportion of time spent indoors or outdoors. There are three models. One model is a year-round model for those having neither gas heating nor gas ranges. The second model is the winter gas model, and the third is the non-winter gas model. The seasons are described in the final section of Exhibit 5. Also there is a note at the end of Exhibit 6 which explains the effect of modifying the coefficients in the original model, which used parts per billion (ppb), for use with parts per hundred million (pphm) concentrations of NO_2. The coefficients which were modified are indicated by an asterisk (*).

Exhibit 7. Comparison of Frequency Distributions for Candidate Regression Variables Between AHSMOG and GRI Study Subjects, is a table which shows how the distributions of responses in the two studies compare. Categorical variables are compared with a chi-square test and the statistics are given immediately following the distribution. The continuous variables are compared with a t-test and the results given following the distribution. The continuous variables were also broken into categories for ease of comparison of the respective distributions and the results of further chi-square tests are also shown. For single variable categorical variables, such as HRSCOOKM, the numbers under the "Value" headings are the values of the categorical variables before they were modified by fraction of time indoors or outdoors factor. For nominal variables, such as OVENHETM, the numbers under the "Value" headings are zero (0) for "No", and one (1) for "Yes." For sets of dummy variables, the values given are merely a convenient way to number the responses. In these cases, every variable except the reference variable, was coded to be a one for exactly the response that it represented and a zero otherwise, as described in Exhibit 5. The reference dummy category in the sets of dummy variables is indicated with an asterisk (*). Exhibit 7 concludes with a note that though the respective distributions differ, the primary point to observe is that the ranges of values experienced by the AHSMOG cohort lie within the corresponding ranges experienced by the GRI subjects.

Exhibit 8, contains two figures which plot the residuals from the multiple regression prediction models as a function of outdoor ambient concentration of NO_2. Figure 1, shows the residuals from the original nonseasonal prediction equation for individuals with gas ranges. These were from nonseasonal prediction equations for individuals with gas ranges. Figure 2, shows the residuals from the original nonseasonal non-gas model. Nonseasonal prediction models were used for these plots since we were primarily concerned about geographical generalizability. Both figures show equal scatter of residuals for all levels of ambient concentration indicating that the prediction models should provide equally reliable estimates for individuals living in low or high areas of ambient NO_2. This is important since many of the AHSMOG study subjects lived in low ambient NO_2 areas. The models used for the figures are slightly different from the final models of Exhibit 6 for two reasons. The final models for individuals with gas ranges were seasonal, and the final models for each of the equations, individuals were required only to have nonmissing values for those variables which were selected to be in the final model. This meant that more individuals were included in the final models than in the models used for the figures which excluded all individuals missing any of the candidate variables. The regression equations from which these residuals came, for Figure 1, the nonseasonal gas model, based on ambient concentration units in pphm:

$$
\begin{aligned}
PERSONC = {} & 0.427*PILOTM \\
& + 0.736*OVENHETM \\
& + 1.127*FORCHETM \\
& + 1.121*WALLHETM \\
& + 0.919*OTHRHETM \\
& + 1.165*HRSOUTM \\
& + 0.473*FF
\end{aligned}
$$

And for Figure 2, the nonseasonal non-gas model, based on ambient concentration units in pphm:

$$
\begin{aligned}
PERSONC = {} & 0.334*WATHEAT \\
& + 1.046*HRSOUTM \\
& + 0.0142*HRSDRIVE \\
& + 0.624*FF
\end{aligned}
$$

Again, the regression coefficients of the second model differ slightly from those of the final prediction equations due to differing numbers of individuals excluded because of missing values. Fewer individuals were excluded for the final regression equations because the candidate variables were restricted to only those used. The final regression equations for individuals with gas ranges were season specific, but even here, the values of the coefficients in the final regression equations agree closely with the common coefficients in the nonseasonal model (see Exhibit 6).

Exhibit 1. Mnemonic Variable Names Used in Indoor Adjustment Equations

Mnemonic Variable Names	GRI Variable Names	1987 AHSMOG Quest. No.
PERSCONC	FCONC12	
PILOTM	HCQ1214AM	93, 94
NOPILOTM	HCQ1214BM (reference dummy)	93, 94
HODALWAM	HCQ1215AM (reference dummy)	95
HODSOMEM	HCQ1215BM	95
HODNEVRM	HCQ1215CM	95
WATHEAT	HCQ19	100
OVENHETM	HCQ1211M	98
HRSCOOKM	HCQ17M	96, 97
EVAPCOLM	HCQ18AM	86, 88
REFCOLM	HCQ18BM	86, 88
NOCOLM	HCQ18CM (reference dummy)	86, 88
OLDHOMEM	HCQ3M	81
FORCHETM	HCQ89AM	89
WALLHETM	HCQ89BM	89
OTHRHETM	HCQ89CM	89
ELECHETM	(not explicitly defined) (reference dummy)	89
HRSOUTM	HSUM67M	60, 61
EMPLOY	PCQ1	63
HRSDRIVE	PCQ10	55
FOUT	FOUT	
FF	FF	
HAMBCONC	FCONC7	

Exhibit 2. Candidate Questions from the GRI Questionnaire

GRI QUESTIONS

Variables	Original Questions from the Questionnaires
	Personal Characteristics Questionnaire
EMPLOY	Do you have an occupation or major activity where you spend at least 20 hours per week, outside the home, on a regular basis? If yes, what is it?
HRSDRIVE	Yesterday, how many minutes did you spend traveling?
	Home Characteristics Questionnaire
OLDHOMEM	In about what year was your home originally built? Since 1980 1970's 1960's 1950's 1940's Before 1940 don't know
EVAPCOLM REFCOLM NODCOLM	Do you use any air conditioning in your house" If yes, what type is it? Yes, evaporative (desert cooler) Yes, refrigeration (central air) Yes, refrigeration (window units) Yes, cent. refrig. & window Yes, cent. refrig. & evap. Yes, window refrig. & evap. No, none don't know
FORCHETM WALLHETM OTHRHETM	What is the main type of fuel used to heat your home? Gas, from the utility company through pipes Bottled, tank or LP gas Electricity Fuel oil, kerosene Solar Something else, specify ______ None don't know What is the main heating system in your home? Forced-air (central warm-air furnace with ducts to individual rooms) Built-in electric units Wall furnace Floor furnace Portable space heaters Steam or hot water system Other, specify ______ Not applicable Don't know

PILOTM
NOPILOTM

What type of range or stove do you cook on?

Gas
Electric
Combination of gas and electric (i.e., gas burners/electric oven
Other, specify ________
Doesn't have range
Don't know

Does your range/stove/oven have a continuously burning gas pilot light?

Yes
No
Not applicable
Don't know

HODALWAM
HODSOMEM
HODNEVRM

What type of range or stove do you cook on?

Gas
Electric
Combination of gas and electric (i.e., gas burners/electric over)
Other, specify ________________
Doesn't have range
Don't know

Do you use an exhaust fan or hood when food is being prepared on the stove?

Yes, always
Yes, sometimes
No or almost never
Hood or vent without fan
Not applicable
Don't know

HRSCOOKM

In an average week, 7 days, how many days would you and any other household members use the stove burners at all for preparing breakfast?
For a typical meal, how many burners would you say are used at once?
How many minutes would the burners usually be on?
On how many days per week would the oven be used for cooking breakfast?
And for how many minutes would it be on?
How about for lunch, dinner, and any other time during the week?

Meal	Stove			Oven		Office Use Only
	days/wk	burners/meal	Min. on	days/wk	Min. on	Total Burner Min.
Breakfast						
Lunch						
Dinner						
Other						

OVENHETM	What types of range or stove do you cook on? Gas Electric Combination of gas and electric (i.e., gas burners/electric oven) Other, specify___________________ Doesn't have range Don't know
	During the winter, is the range or stove ever used to help heat your house? Yes No Not applicable Don't know
WATHEAT	Do you have a gas water heater within your living area (including garage and basement)?
	Derived Variables
HRSOUTM	Derived from the time/activity diary.
PERSCONC	Derived from the field measurement.
HAMBCONC	Derived from the field measurement.

EXHIBIT 3

Appendix C

Candidate Questions from 1987 AHSMOG Questionnaire

(n=3914)

HRSDRIVE

55. In total, approximately how many hours per week do you usually spend driving or riding on any type of roadway?

742[ ] Never ride or drive on a weekly basis 3172 missing

_______ hours per week riding or driving on roadways $\bar{x}=6.1$ $\sigma=6.2$

HRSOUTM

How many hours per WEEK including weekends, are you outside of buildings?

60. SUMMER (June through September)

86[ ] None
1526[ ] 1-7 hours
970[ ] 8-14 hours
507[ ] 15-21 hours
336[ ] 22-28 hours
182[ ] 29-35 hours
27[ ] 36-42 hours
167[ ] More than 42 hours
13 missing

61. REST OF YEAR (October through May)

121[ ] None
1848[ ] 1-7 hours
961[ ] 8-14 hours
412[ ] 15-21 hours
250[ ] 22-28 hours
115[ ] 29-35 hours
86[ ] 36-42 hours
106[ ] More than 42 hours
15 missing

EMPLOY

EMPLOYMENT

63. What is your current employment status?

458 [ ] Unemployed
296 [ ] Working part time
1118 [ ] Working full time
1508 [ ] Retired and not working
402 [ ] Retired and working part time
[ ] Retired but working full time

80 → IF RETIRED, please give date of retirement

____/____
month year

52 missing

OLDHOME

81. How old is your residence?

55 [ ] Less than 1 year
40 [ ] 1 year old
118 [ ] 2-3 years old
621 [ ] 4-10 years old
694 [ ] 11 to 20 years old
2361 [ ] More than 20 years old

EVAPCOLM, REFCOLM, NOCOLM

86. What type of air cooling system do you have in your home?

1422[ ] None
352[ ] Evaporative cooler (swamp cooler)
2057[ ] Refrigerating type (air conditioner)
4[ ] Both
19 missing

88. What type of air cooling system did you have in your home in March 1977?

1853[ ] None
441[ ] Evaporative cooler (swamp cooler)
1557[ ] Refrigerating type (air conditioner)
24[ ] Both
53 missing

FORCHETM, WALLHETM, OTHRHETM, ELECHETM

HEATING AND COOKING

89. How frequently are the following heating systems used in your home during the winter months? Please check never if not used.

Never	Monthly or less	Weekly	Daily	
	190	221	1735	Gas forced air furnace
	70	68	468	Gas wall furnace
	33	42	365	Gas floor furnace
	11	7	54	Gas space heater
	28	18	21	Kerosene space heater
	497	476	232	Fireplace
	58	61	375	Wood stove
1	46	46	245	electric
			23	oil
		1	2	coal
	3	2	23	other fuel
	1	0	7	non-fossil
	2	0	1	steam
		3	13	hot water/gas

PILOTM, NOPILOTM

'ave you ever lived in a home
with a gas cooking stove?

265[ ] No ⟶ skip to question 100
1886[ ] Yes. CURRENTLY
1756[ ] Yes. IN THE PAST.
Year last used
19______
7 missing

If you answered CURRENTLY to question 93, answer questions 94–98 for CURRENT use only.

If you answered IN THE PAST to question 93, answer questions 94–98 in terms of PAST USE.

94. How is (was) your gas cooking stove lighted?

224[ ] Light by hand
463[ ] Electric ignition
2874[ ] Pilot light ⟶ how many pilots in range and oven? ______
81 missing

HODALWAM, HODSOMEM, HODNEVRM

95. How often do (did) you use an exhaust fan or range hood when food is (was) being prepared on the stove?

684 [ ] Rarely or never
1380 [ ] Occasionally (when kitchen is smoky or for odors)
284 [ ] At least half the time that the stove is on
225 [ ] Always/almost always, whenever stove is on
69 missing

HRSCOOKM

96. On the average, how many hours per DAY is (was) COOKING done with a GAS stove in your home?

144 [ ] None
1290 [ ] Less than 1 hour per day
1839 [ ] 1–2 hours per day
303 [ ] More than 2 hours per day
66 missing

97. On the average, how many hours per WEEK is (was) BAKING done with a GAS oven in your home?

321 [ ] None
1052 [ ] Less than 1 hour per week
1513 [ ] 1–2 hours per week
584 [ ] 3–5 hours per week
99 [ ] 6 or more hours per week
73 missing

OVENHETM

98. During the winter, how frequently is (was) the range or stove used to help heat your house?

2944 [ ] Never
328 [ ] Monthly or less
133 [ ] Weekly
172 [ ] Daily
65 missing

WATHEAT

100. Do you have a gas water heater?

689 [ ] No
IF YES, where is it located?
1323 [ ] Inside the home or in a closet inside the home
921 [ ] In the garage
36 [ ] Carport
911 [ ] In a closet accessed from outside the home
[ ] Other (specify) ______
34 missing

Exhibit 4. Coding of Original Variables Used in GRI Study

Note that the original data was received at Integrated Environmental Services (IES) from Harvard University already coded into a computer readable form. Missing values were coded with the standard SAS "." format. Individuals in the analysis were required to be at least 20 and have a nonmissing value for person exposure (FCONC12). If an individual was missing a question or variable which was to be used in the particular analysis, all his or her data were excluded from those analyses by SAS. The responses of "don't know" and "not applicable" were also excluded from the analyses. For purposes of analysis, winter is assumed to be the months from December through March, inclusive.

FCONC7 The average hourly outdoor ambient NO_2 concentration in parts per billion (ppb).

FCONC12 The two day average personal NO_2 exposure. This was derived from the field measured data. It was a continuous variable with units in parts per billion (ppb).

PCQ10 Hours driving per week. This was a continuous variable. It was based on question 10 of the Personal Characteristics Questionnaire, which asked, "Yesterday, how many minutes did you spend traveling?" This was converted to hours per week.

HSUM67 Hours outside of buildings per week. This was a continuous variable derived from the time/activity diary for each of the two days the individual participated in the study. (It was based on the time activity diary in connection with questions 6 and 7 on the Personal Characteristics Questionnaire.) It was recoded to a categorical variable, HSUM67M, as follows: HSUM67M = 0, if HSUM67 was less than 0.5 hours. HSUM67M = 4, if HSUM67 was at least 0.5 hours but less than 7.5 hours. HSUM67M = 11, if HSUM67 was at least 7.5 hours but less than 14.5 hours. HSUM67M = 18, if HSUM67 was at least 14.5 hours but less than 21.5 hours. HSUM67M = 25, if HSUM67 was at least 21.5 hours but less than 28.5 hours. HSUM67M = 32, if HSUM67 was at least 28.5 hours but less than 35.5 hours. HSUM67M = 39, if HSUM67 was at least 35.5 hours but less than 42.5 hours. HSUM67M = 46, if HSUM67 was greater than or equal to 42.5 hours.

PCQ1 A categorical variable indicating whether person had a major activity outside the home. It was assigned a value of one (1) if the person was employed or otherwise occupied at least twenty (20) hours per week or more outside the home, it was assigned a value of zero (0) otherwise.

HCQ3 A categorical variable based indicating whether house was built before 1970. This variable was based on question 3 (decade of house construction) of the House Characteristics Questionnaire. If the house was built before 1970 this variable was assigned a value of one (1). It was assigned a value of zero (0) otherwise.

HCQ18A
HCQ18B
HCQ18C These were categorical variables indicating the type of cooling system used in the home. This variable was based on question 18 in the House Characteristics Questionnaire. HCQ18A was assigned a value of one (1) if either an evaporative cooling system was used or if both an evaporative and a refrigerative cooling system were used. These would be categories "evaporative (desert cooler)", "cent. refrig. & evap.", and, "window refrig. & evap." HCQ18A was assigned a value of zero (0) otherwise. HCQ18B was assigned a value of one (1) if a refrigerative type of cooling system was the only type used. This would be answers from the categories "refrigeration (central air)", "refrigeration (window units)", and, "cent. refrig. & window". HCQ18B was assigned a value of zero (0) otherwise. HCQ18C was assigned a value of one (1) if no cooling system was used, it was assigned a value of zero (0) otherwise. The net effect was that exactly one of these variables would be assigned a value of one (1) and the other two would be assigned a value of zero (0). HCQ18C was intended to be the reference category.

HCQ89A
HCQ89B
HCQ89C These categorical variables indicated the type of heating system used. They were created from questions 8 (primary heating fuel) and 9 (primary heating system) of the House Characteristics Questionnaire. The heating system was considered to be gas if the primary heating fuel was either "gas, from the utility company through pipes" or "bottled, tank or LP gas". Other fuel sources besides "electricity", were considered to be "other heating systems". HCQ89A was assigned a value of one (1) if the response indicated a gas fuel source (question 8) and a "forced-air (central warm-air furnace with ducts to individual rooms)" heating system (question 9) was used to heat the home, it was assigned a value of zero (0) otherwise. HCQ89B was assigned a value of one (1) if a gas fuel source (question 8) with a wall furnace or a floor furnace (question 9) was used to heat the home, it was assigned a value of zero (0) otherwise. HCQ89C was assigned a value of one (1) if any other type of nonelectric fuel (question 8) was used, or any other heating system (besides forced-air, wall furnace, floor furnace, or built-in electric units) was used, it was assigned a value of zero (0) otherwise. This meant that the reference category, electric heating system, was indicated by all three of these variables being assigned the value of zero (0). Other responses, such as "not applicable", "don't know", or missing values, were excluded from the analyses.

H1214A
H1214B

These categorical variables indicated whether a gas stove with or without a pilot light was used. They were based on question 12 (stove type) and 14 (pilot light) of the House Characteristics Questionnaire. H1214A was assigned a value of one (1) if a gas or combination stove (question 12) was used, and it had a "continuously burning pilot light" (question 14); it was assigned a value of zero (0) otherwise. H1214B was assigned a value of one (1) if a gas or combination stove was used (question 12), and it did not have a "continuously burning pilot light" (question 14); it was assigned a value of zero (0) otherwise. In case either question 12 or 14 were missing, both categorical variables were assigned a missing value.

H1215A
H1215B
H1215C

These categorical variables indicated whether or not a hood was used with the gas stove. These variables were based on questions 12 (stove type) and 15 (hood use) of the House Characteristics Questionnaire. H1215A was assigned a value of one (1) if the individual had a gas or combination stove and always used an exhaust hood when using the stove, it was assigned a value of zero (0) otherwise. H1215B was assigned a value of one (1) if the individual had a gas or combination stove and sometimes used an exhaust hood when using the stove, it was assigned a value of zero (0) otherwise. H1215C was assigned a value of one (1) if the individual had a gas or combination stove and never used an exhaust hood when using the stove, it was assigned a value of zero (0) otherwise. There were also categories for "hood or vent with no fan", "not applicable", and "don't know", but these were not used in the analyses.

HCQ17

A continuous variable indicating the number of hours cooking and baking per day. This variable was based on question 17 of the House Characteristics Questionnaire, which asked how many minutes the stove and/or oven was on per meal, how many burners were used, and how many times per week it would be used. The total burner minutes were converted to hours and the value was recoded as follows: HCQ17M was assigned a value of zero (0) if either HCQ17 had a value of zero (0), or if the stove type (question 12) was not "gas", or "combination". HCQ17M was assigned a value of one (1) if HCQ17 was more than zero (0) but less than one (1). HCQ17M was assigned a value of two (2) if HCQ17 was at least one (1) but not more than two (2). HCQ17M was assigned a value of three (3) if HCQ17 was more than two (2). This variable was not relevant to individuals who had neither a gas nor combination type stove, so in such cases this variable was assigned a value of zero. It was missing only if the person had a gas or combination type stove.

HCQ1211

A categorical variable used to indicate whether or not a gas stove was used to heat the home. It was based on questions 12 (stove type) and 11 (stove to heat) of the House Characteristics Questionnaire. Question 12 asked about the type of stove used in the home. The categories were: "Gas", "Electric", "Combination of gas and electric (i.e. gas burners/electric oven)", "Other, specify _______", "Doesn't have range", and "Don't know". Question 11 asked whether the stove was used to heat the home during the winter. The possible responses were: "Yes", "No", "Not applicable", and "Don't know". It was assigned a value of one (1) if a person had a gas stove, or a combination gas and electric stove, and used it to heat the home, it was assigned a value of zero (0) otherwise. Missing values for either question 12 or 11 caused this variable to be assigned a missing value.

HCQ19 A categorical variable used to indicate whether or not the person had a gas water heater within the living area. It was assigned a value of one (1) if a person had a gas water heater within the living area (including garage and basement), it was assigned a value of zero (0) otherwise.

Exhibit 5. Coding of Mnemonic Variables from 1987 AHSMOG Questions and Manner of Employment in Regression Equations

The value for PERSCONC was computed for an individual for a month only if none of the variables in the relevant equation were missing, and the individual was living in his/her final residence.

The variables used in the models are defined as follows:

PERSCONC The adjusted personal average hourly concentration for a given month. The units are pphm of NO_2.

AMBCONC The cumulative exposure to NO_2 in a month. The quantity was parts per hundred million (pphm) for each hour summed over the number of hours in the month. The units were pphm-hr. The data was supplied by the California Air Resources Board covering the period from April 1, 1977 to March 31, 1987.

HAMBCONC The average hourly concentration of NO_2 in a given month, computed as AMBCONC divided by the number of hours in that month. Units were pphm.

HRSOUT The number of hours outdoors per week. This was the midpoint of the 7-hour time interval which the respondent chose as representative of him-/herself. There were separate questions for summer (question 60) and for non-summer (question 61). The summer response was assigned for the months from June through September, inclusive. The non-summer response was assigned for the other months, namely, October through May, inclusive. The possible categories and the values assigned for modelling computations were as follows:

Questionnaire Category	Assigned Value
None	0
1 - 7 hours per week	4
8 - 14 hours per week	11
15 - 21 hours per week	18
22 - 28 hours per week	25
29 - 35 hours per week	32
36 - 42 hours per week	39
More than 42 hours per week	42

FOUT The fraction of time outdoors, computed as HRSOUT divided by the number of hours in a week, i.e., 168 hours. This variable was unitless.

FIN The fraction of time indoors, computed as 1 - FOUT. This variable was also unitless.

WATHEAT	A dummy variable representing the presence of a gas water heater inside the home (question 100). It was assigned a value of one (1), if the respondent indicated he/she had a gas water heater inside the home or inside a closet in the home (response 2). It was assigned the value of zero (0) for all other responses. Nonresponse was counted as missing.
HRSDRIVE	The number of hours per week driving or riding on roadways. There were two ways to respond to this question (question 55). The respondent could check the response that he/she never rode or drove on a weekly basis, or the respondent could write on the line provided the number of hours per week spent riding or driving on roadways. If the respondent checked the first response (never ride or drive on a weekly basis), the value of HRSDRIVE was taken as zero (0), regardless of whether a numerical value was written in the blank provided. Nonresponse for a part was taken as missing only for that part and not for the other part because generally a person answered only one of the parts.
EMPLOY	A dummy variable indicating that the respondent was employed (question 63). Responses besides unemployed (response 1), or retired and not working (response 4), were counted as working. Nonresponse was counted as missing. If the person retired during the study period and was not working, The variable was assigned a value of one (1) for the months before the retirement date, it was assigned zero (0) from the month of retirement and onward. If only the year of retirement was given, the month was taken as April of that year. If no retirement date was indicated, a default data was assigned which was midway through the study period (month 60).
FORCHET WALLHET ELECHET OTHRHET	Dummy variables indicating the type of heating system used to heat the home (question 89). FORCHET represented gas forced air heating. WALLHET represented gas wall or floor heating. ELECHET was any sort of electric heating, it was taken as the reference category and was not explicitly coded in the final model. OTHRHET was other non-electric heating. These were assigned together so that at most one had a value of one (1) when use of that type of heating was indicated, and a value of zero (0), otherwise. Nonresponse was missing. In case of multiple systems, the most frequently used was selected. In cases of ties, the system was selected in the order: FORCHET, WALLHET, ELECHET, and OTHRHET.

NOTE: The following variables were computed only for those individuals who either currently (response 2), or in the past (response 3), lived in a home with a gas cooking stove (question 93). Namely, the following variables were relevant only to the seasonal gas models.

PILOT	A dummy variable indicating that the respondent's gas cooking stove was lighted by a pilot light (question 94). It was assigned the value of one (1) only if the respondent indicated his/her gas cooking stove was lighted by a pilot light (response 3). It was assigned the value of zero (0) for all other responses. Nonresponse was counted as missing.

OVENHET A dummy variable indicating that the respondent used his/her range or stove to help heat the house during the winter on a weekly or daily basis (question 98). It was assigned the value of one (1) if the respondent indicated a frequency of either weekly (response 3), or daily (response 4). It was assigned the value of zero (0) for all other responses. Nonresponse was counted as missing.

HRSCOOK The number of hours cooking (question 96) and baking (question 97, adjusted) per day. This was the sum of hours cooking per day and the quotient of hours baking per week divided by the number of days in a week (7). Each question was a scale of limited values offering the following options. For hours cooking: (1) none, (2) less than 1 hour per day, (3) 1 - 2 hours per day, or (4) more than 2 hours per day. For hours baking: (1) none, (2) less than 1 hour per week, (3) 1 - 2 hrs/wk, (4) 3 - 5 hrs/wk, or (5) 6 or more hrs/wk.

QUESTIONNAIRE CATEGORY

Hours Cooking	Contribution
None	0
Less than 1 hr/day	1
1-2 hrs/day	2
More than 2 hrs/day	3

Hours Baking	Contribution
None	0.00
Less than 1 hr/wk	0.07
1.0-2.0 hrs/wk	0.21
2.1-5.9 hrs/wk	0.57
6 or more hrs/wk	1.00

The combined score was rescaled to (0) none, (1) less than one hr/day, (2) one to two hrs/day, or (3) more than two hours per day. If either response was missing, the combined score was counted as missing.

The following list of variables were created and exist in the computer file but did not enter the final models.

HODNEVR
HODSOME
HODALWA

Dummy variables indicating how often the respondent used his/her range hood or exhaust fan if he/she had a gas stove (question 95). At most one of these variables was assigned a one (1), the rest were assigned zero (0). The relevant categories were HODNEVR, rarely or never use the hood (response 1), HODSOME, use the hood occasionally or use it about half the time (response 2 or 3), or, HODALWA, always or almost always use the hood (response 4), this was the reference category. Nonresponse was counted as missing.

EVAPCOL
REFCOL
NOCOL

Dummy variables indicating the type of air cooling system in the respondent's home. There were three questions which had a bearing on these variables. The type of cooling system in the home in 1987 (question 86), the number of years the respondent had had this type of cooling system (question 87), and the type of cooling system in the home in March 1977 (question 88).

If the type of cooling system in 1987 was the same as that in March 1977, it was assumed to be the same for the whole period of the study. If the type of cooling system in 1987 was different from that in March 1977, the study month of change was calculated from question 87 as April of the year of change. The variables were assigned the March 1977 values prior to the study month of change, and the 1987 values after and including the study month of change. If the number of years having the current system was not available to calculate the study month of change, month 60 was arbitrarily assigned. If there was no response for the 1987 system, these dummy variables for the months after and including the month of change were missing. Similarly, if the response for the March 1977 system was missing, these dummy variables were counted missing for the study months before the month of change. NOCOL was the reference category.

OLDHOME

A dummy variable indicating the home was more than twenty years old (question 81). If more than twenty years old was indicated (response 6) this variable was assigned a one (1). To other responses it was assigned zero (0). Nonresponse was taken as missing.

Adjusted individual monthly exposures were accumulated only from the time the individual began living in his/her last residence. If a person moved to his/her final residence within the study period, the first month for the calculations of adjusted mean concentration (cumulative exposure) began in the month following the indicated study month of move. However, if a person was in his/her final residence at the beginning of the study, the calculations of adjusted mean concentration were calculated from the first month of the study period. Regardless of the transition flags mentioned below as being set for various changes in the model, no values were calculated prior to the study months corresponding to living in one's final residence.

Final residence date and location was determined by his/her residence history (based on question 74) and the length of time indicated as living in his/her present residence (question 79). This time was taken as the more recent of either the date of the study month following the move to the last zip code, or the study month corresponding to April of the year of moving to the present residence, calculated from the number of year having lived at the present residence (see NOTE below). Months away from home were counted as missing. If the number of years having lived at the present residence was missing, the study month of the move was taken. No one was missing a residence history.

The appropriate model (seasonal gas, or nonseasonal non-gas) was chosen based on whether the respondent had never (response 1), currently (response 2), or ever (response 3), lived in a home with a gas cooking stove (question 93), and the number of month he/she heated his/her home (question 90). If the respondent had never lived in a home with a gas cooking stove (response 1), the non-gas model was chosen. If the respondent had lived in a home with a gas cooking stove in the past (response 3), the year of last use was asked (i.e. 19__). If April of the year indicated following response 3, was in the study period, it was taken as the study month of stove change. For the months before this change month the seasonal gas model was used, while for the change month and after the non-gas model was used. If the year of stove change was missing, it was assumed not to have fallen in the study period and the non-gas model was used for the whole period. This was equivalent to assuming that an individual had not replaced his stove since moving to his final residence.

The season for the gas model was determined by rounding the number of months the home was heated (question 90) to a whole number (half months were rounded up). If the rounded number of months was even, the months covered by these months, centered on February 1, were taken as winter months. If the rounded number of months was odd, the months covered by these months, centered on January 15, were taken as winter months. Other months were considered non-winter months. If the number of months the home was heated was missing, no calculations for any month were made, all were counted as missing.

For each month, the algorithm combined the unadjusted monthly exposure to NO_2 for that month, and the above variables for that same month, to compute the adjusted NO_2 exposure for that month. The final model equations, for YEAR-ROUND NON-GAS, for WINTER GAS, and for NON-WINTER GAS, are displayed in Table I.

Exhibit 6. Final Model Equations for Adjusting NO_2 Mean Concentration

Year-Round Non-Gas Model:

PERSCONC = 0.3709270 · WATHEAT*
+ 1.033490 · HRSOUTM
+ 0.0153869 · HRSDRIVE*
+ 0.623695 · FF

Winter Gas Model:

PERSCONC = 0.443312414 · PILOTM*
+ 0.934460406 · OVENHETM*
+ 1.461719812 · FORCHETM*
+ 1.664617218 · WALLHETM*
+ 0.679355432 · OTHRHETM*
+ 1.28409792 · HRSOUTM
+ 0.35685334 · FF

Non-Winter Gas Model:

PERSCONC = 0.482013368 · PILOTM*
+ 0.467557022 · OVENHETM*
+ 0.197041620 · HRSCOOKM*
+ 0.96733273 · HRSOUTM
+ 0.323246108 · EMPLOY*
+ 0.010750374 · HRSDRIVE*
+ 0.61787524 · FF

The variables ending in "M" were modified from their original form as follows:

PILOTM = FIN · PILOT
OVENHETM = FIN · OVENHET
FORCHETM = FIN · FORCHET
WALLHETM = FIN · WALLHET
OTHRHETM = FIN · OTHRHET
HRSCOOKM = FIN · HRSCOOK
HRSOUTM = FOUT · HAMBCONC

Another created variable was:

FF = FIN · HAMBCONC

Where,

FIN = 1 - FOUT
FOUT = HRSOUT/(number of hours in a week, i.e., 168)
HAMBCONC = AMBCONC/(number of hours in the particular month)

*The original equations were based on ambient concentrations in parts per billion (ppb). Our data was in parts per hundred million (pphm). Those coefficients which were not modified by the hourly average ambient concentration (HAMBCONC) needed to be divided by 10 to get the coefficients shown. The coefficients which were divided by 10 are indicated by the asterisk.

Exhibit 7. Comparison of Frequency Distributions for Candidate Regression Variables Between AHSMOG and GRI Study Subjects

Note: Below are listed all the candidate variables considered for the regression models. Even though the names of the modified variables are given, the values for the original, unmodified values from the questionnaires are listed when there is more than one dummy variable involved. The reference category of a set of dummy variables is indicated by an asterisk (*).

Pilot Light in Gas Stove[1] (HCQ1214AM & HCQ1214BM = PILOTM & NOPILOTM)

VALUE		AHSMOG	%	GRI	%
1	Gas stove & pilot light	2874	73.4%	290	49.8%
2	Gas stove & no pilot light*	687	17.6%	125	21.5%
	Nongas stove	265	6.8%	161	27.7%
.	Missing	88	2.2%	6	1.0%
		3914	100.0%	582	100.0%

Chi-Square = 26.815
d.f. = 1
p < 0.001

Daily Hours Cooking and Baking with Gas[2] (HCQ17M = HRSCOOKM)

VALUE		AHSMOG	%	GRI	%
0	Gas stove & no cooking	120	3.1%	6	1.0%
1	Gas stove & 0-0.9 hrs. cooking	23	0.6%	119	20.4%
2	Gas stove & 1-2 hrs. cooking	1281	32.7%	111	19.1%
3	Gas stove & more than 2 hrs. cooking	2143	54.8%	76	13.1%
	Nongas stove	265	6.8%	161	27.7%
.	Missing or not applicable	82	2.1%	109	18.7%
		3914	100.0%	582	100.0%

Chi-Square = 1167.739
d.f. = 3
p < 0.001

Gas Stove Hood[1] (HCQ1215AM, HCQ1215BM, HCQ1215CM = HODALWAM, HODSOMEM, HODNEVRM)

VALUE		AHSMOG	%	GRI	%
1	Gas stove & always use hood*	225	5.7%	51	8.8%
2	Gas stove & sometimes use hood	1664	42.5%	143	24.6%
3	Gas stove & never use hood or N/A**	1684	43.0%	221	38.0%
	Nongas stove	265	6.8%	161	27.7%
.	Missing	76	1.9%	6	1.0%
		3914	100.0%	582	100.0%

Chi-Square = 34.247
d.f. = 2
$p < 0.001$

** This category also includes the hood without exhaust fan and the not applicable category of the GRI questionnaire.

Gas Stove Oven Used to Heat Home[1] (HCQ1211M = OVENHETM)

VALUE		AHSMOG	%	GRI	%
0	Gas stove & not used to heat	3272	83.6%	353	60.7%
1	Gas stove & used to heat home	305	7.8%	60	10.3%
	Nongas stove	265	6.8%	161	27.7%
.	Missing or not applicable	72	1.8%	8	1.4%
		3914	100.0%	582	100.0%

Chi-Square = 16.044
d.f. = 1
$p < 0.001$

Gas Water Heater Inside Home[1] (HCQ19 = WATHEAT)

VALUE		AHSMOG	%	GRI	%
0	No gas water heater inside home	2557	65.3%	258	44.3%
1	Gas water heater inside home	1323	33.8%	316	54.3%
.	Missing	34	0.9%	8	1.4%
		3914	100.0%	582	100.0%

Chi-Square = 94.4026
d.f. = 1
$p < 0.001$

Home Cooling System[1] (HCQ18AM, HCQ18BM, HCQ18CM = EVAPCOLM, REFCOLM, NOCOLM)

VALUE		AHSMOG	%	GRI	%
1	Evaporative & Both Cooling	416	10.6%	34	5.8%
2	Refrigerative Cooling	2058	52.6%	268	46.0%
3	No cooling*	1422	36.3%	274	47.1%
.	Missing	18	0.5%	6	1.0%
		3914	100.0%	582	100.0%

Chi-Square = 31.4097
d.f. = 2
$p < 0.001$

Home Age[1] (HCQ3M = OLDHOMEM)

VALUE		AHSMOG	%	GRI	%
0	Home less than 20 years old	1528	39.0%	156	26.8%
1	Home at least 20 years old	2361	60.3%	371	63.7%
.	Missing	25	0.6%	55	9.5%
		3914	100.0%	582	100.0%

Chi-Square = 18.4669
d.f. = 1
$p < 0.001$

Home Heating System[1] (HCQ89AM, HCQ89BM, HCQ89CM = FORCHETM, WALLHETM, OTHRHETM)

VALUE		AHSMOG	%	GRI	%
1	Gas forced air heating	2026	51.8%	277	47.6%
2	Gas wall and/or floor heating	881	22.5%	212	36.4%
3	Other (nonelectric) heating	595	15.2%	30	5.2%
4	Electric heating*	174	4.4%	57	9.8%
.	Missing	238	6.1%	6	1.0%
		3914	100.0%	582	100.0%

Chi-Square = 101.7254
d.f. = 3
$p < 0.001$

Hours Per Week Outdoors[2] (HSUM67M = HRSOUTM)

VALUE		AHSMOG	%	GRI	%
0	None	82	2.1%	14	2.4%
4	1-7 hrs/wk outdoors	1805	46.1%	81	13.9%
11	8-14 hrs/wk outdoors	944	24.1%	151	25.9%
18	15-21 hrs/wk outdoors	440	11.2%	123	21.1%
25	22-28 hrs/wk outdoors	285	7.3%	67	11.5%
32	29-35 hrs/wk outdoors	140	3.6%	52	8.9%
39	36-42 hrs/wk outdoors	103	2.6%	23	4.0%
46	More than 42 hrs/wk outdoors	99	2.5%	52	8.9%
.	Missing	16	0.4%	19	3.3%
		3914	100.0%	582	100.0%

Chi-Square = 282.052
d.f. = 7
$p < 0.001$

Employment Status[1] (PCQ1 = EMPLOY)

VALUE		AHSMOG	%	GRI	%
0	Not employed	1832	46.8%	195	33.5%
1	Employed	1883	48.1%	387	66.5%
.	Missing	199	5.1%	0	0.0%
		3914	100.0%	582	100.0%

Chi-Square = 50.4598
d.f. = 1
$p < 0.001$

Hours Per Week Driving on Crowded Roadways[3] (PCQ10 = HRSDRIVE)

VALUE		AHSMOG	%	GRI	%
3	0- 5.9 hrs/wk driving or riding	2688	68.7%	233	40.0%
9	6-11.9 hrs/wk driving or riding	829	21.2%	167	28.7%
15	12-17.9 hrs/wk driving or riding	240	6.1%	94	16.2%
21	18-23.9 hrs/wk driving or riding	66	1.7%	41	7.0%
27	24-29.9 hrs/wk driving or riding	24	0.6%	23	4.0%
33	30-35.9 hrs/wk driving or riding	18	0.5%	9	1.5%
39	36-41.9 hrs/wk driving or riding	6	0.2%	3	0.5%
45	42-47.9 hrs/wk driving or riding	3	0.1%	6	1.0%
51	More than 48 hrs/wk driving or riding	8	0.2%	5	0.9%
.	Missing	32	0.8%	1	0.2%
		3914	100.0%	582	100.0%

Chi-Square = 302.868
d.f. = 8
$p < 0.001$

Mean	4.89	10.14
Standard Deviation	5.75	9.87
Number of Cases	3882	581

t = 12.508
d.f. = 640
s.e. = 0.420
$p < 0.001$

Ambient Mean NO_2 Concentration[3] (FCONC7 = HAMBCONC)

VALUE		AHSMOG	%	GRI	%
0	0-3.9 ppb ambient concentration	314	8.0%	6	1.0%
8	4-11.9 ppb	84	2.1%	73	12.5%
16	12-19.9 ppb	446	11.4%	115	19.8%
24	20-27.9 ppb	782	20.0%	139	23.9%
32	28-35.9 ppb	1023	26.1%	122	21.0%
40	36-43.9 ppb	420	10.7%	49	8.4%
48	44-51.9 ppb	447	11.4%	36	6.2%
56	52-59.9 ppb	354	9.0%	7	1.2%
64	60-67.9 ppb	12	0.3%	12	2.1%
72	68-75.9 ppb	10	0.3%	6	1.0%
80	76+ ppb	22	0.6%	6	1.0%
.	Missing	0	0.0%	11	1.9%
		3914	100.0%	582	100.0%

Mean	30.86	27.16
Standard Deviation	15.63	14.88
Number of Cases	3914	571

t = 5.316
d.f. = 4483
s.e. = 0.696
$p < 0.001$

Chi-square = 327.482
d.f. = 10
$p < 0.001$

Ages, sex by decade[4]

VALUE	AHSMOG	%	GRI	%
20-29 years old, male	22	0.6%	48	8.2%
30-39 years old, male	130	3.3%	81	13.9%
40-49 years old, male	284	7.3%	32	5.5%
50-59 years old, male	463	11.8%	31	5.3%
60 or more years old, male	515	13.2%	55	9.5%
20-29 years old, female	30	0.8%	71	12.2%
30-39 years old, female	289	7.4%	87	14.9%
40-49 years old, female	490	12.5%	36	6.2%
50-59 years old, female	719	18.4%	54	9.3%
60 or more years old, female	972	24.8%	72	12.4%
. Missing	0	0.0%	15	2.6%
	3914	100.0%	582	100.0%

Chi-square = 760.020
d.f. = 9
$p < 0.001$

NOTE: Three (3) are missing both age and sex, seven (7) males are missing age, and five (5) females are missing age in GRI study.

[1]Recoded with dummy variables.
[2]Values used in regression computations.
[3]Used as a continuous variable in regression computations.
[4]Continuous in AHSMOG analyses, discrete for GRI. The age data was by decade for the GRI study.

*Reference category

NOTE: Although there are statistically significant differences in the distributions of a number of the variables above; the important thing for application of GRI based regressions to the AHSMOG data set is that the range of values experienced by AHSMOG study participants lie within those experienced by GRI study participants. This occurs for all of the above variables. There are fewer GRI subjects in the lowest ambient exposure category than would be desirable. This was partially overcome by using an interactive regression model where certain variable values such as fraction of time spent indoors/outdoors were multiplied by ambient concentration.

Exhibit 8. Figure 1. Scatterplot of Residuals Versus Ambient Outdoor NO_2 for Preliminary Non-Seasonal Prediction Equation for Individuals with a Gas Range in the Home.

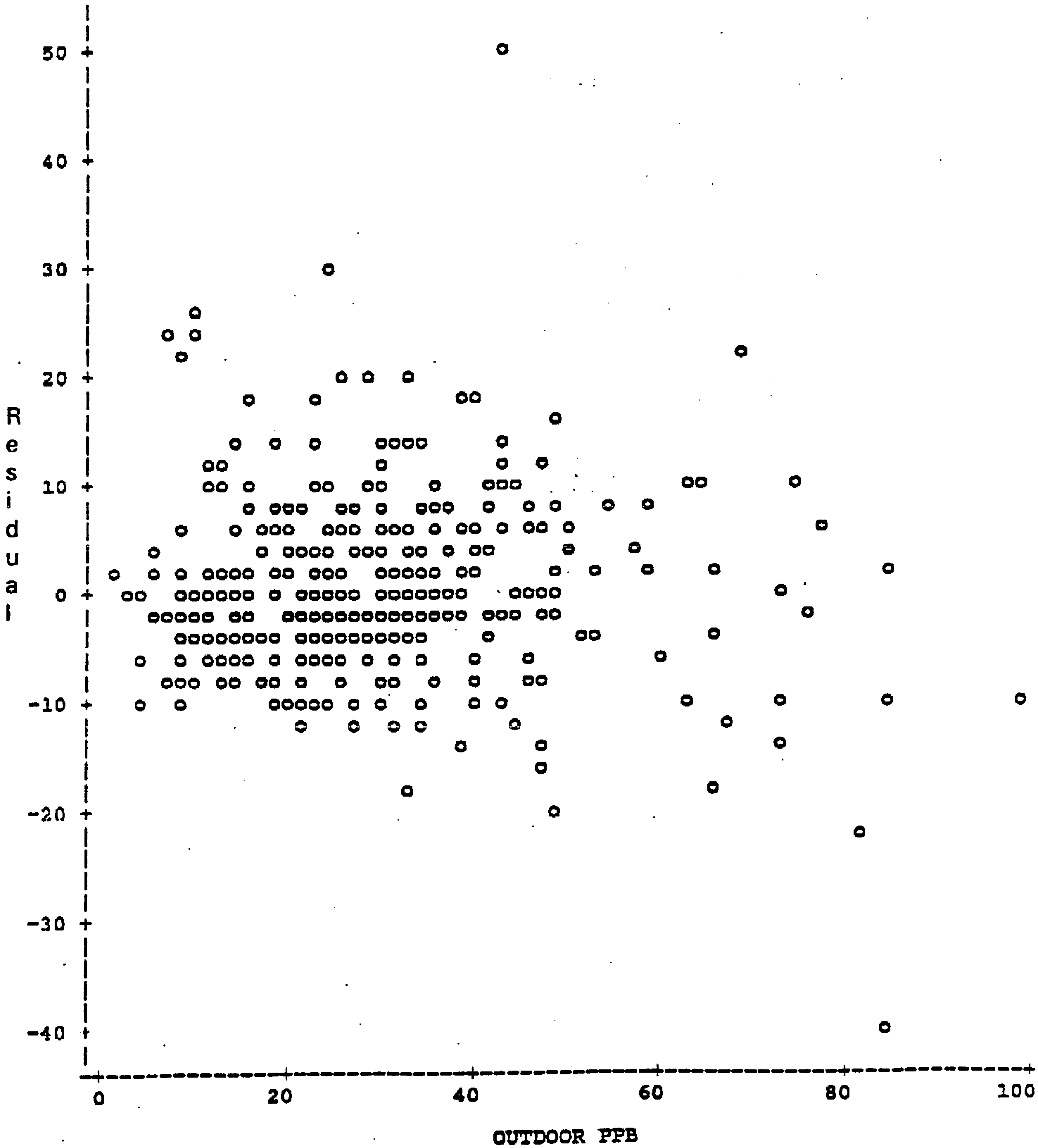
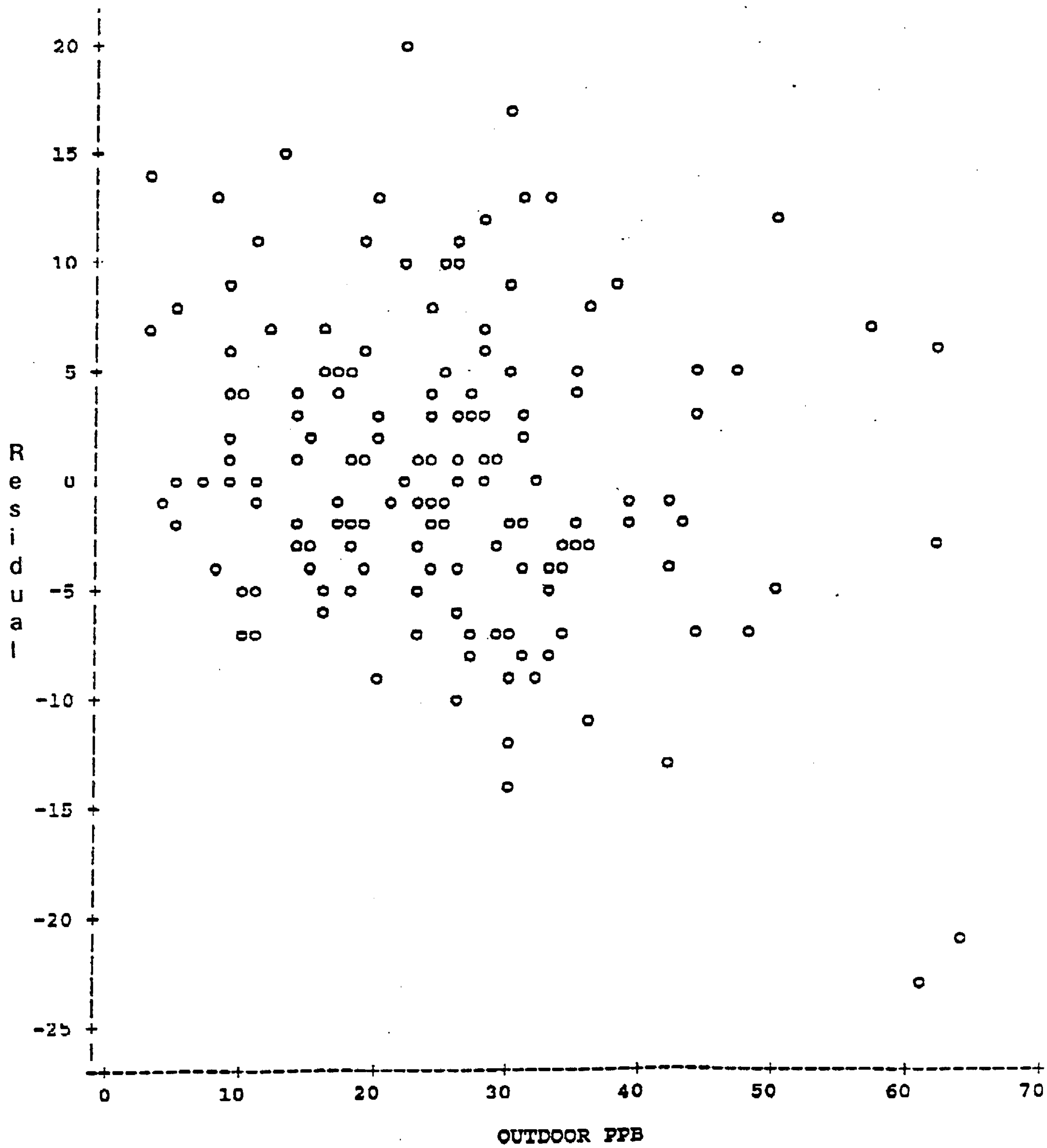


Exhibit 8. Figure 2. Scatterplot of Residuals Versus Ambient Outdoor NO_2 for Preliminary Non-Seasonal Prediction Equation for Individuals without a Gas Range in the Home.



APPENDIX D

Appendix D

Indirect Estimation Methods for PM2.5

§ 1. Daily Visibility Estimates

This chapter describes the procurement of daily visibility data for the years 1966 through 1986 at ten California airports and how the visibility estimates (beta extinction coefficients) were estimated. Pechan and Associates, a firm experienced in working with airport visibility data, provided the daily extinction coefficient estimates. They also formed estimates of mean concentration of PM2.5 and exceedance frequencies for various cutoffs of PM2.5. The estimates of PM2.5 provided by Pechan & Associates, however, were never used for health effects analyses since they were based on national regression equations established by Trijonis. We determined that these national regression equations were not suitable for specific sites in California and formed site and season specific regression equations for estimating PM2.5 from visibility data as described in later sections. Pechan & Associates obtained the visibility data from the National Climatic Data Center. The ten airports for which visibility data were provided are listed in Table 1.

Visibility measures are recorded by airport controllers by recording the distance and miles to the furthest of a set of discrete markers that they can see. These observations are made several times during the day. Pechan & Associates abstracted the readings for three different hours of each day - 10:00 a.m., 1:00 p.m. and 4:00 p.m. For a day to be included, at least two of these three values must be present. Days that did not have at least two valid measures were counted as missing. Missing days were assumed to have the average value for the month.

The following rules and formula were used to determine the beta extinction coefficients for each day. The national equation developed by Trijonis for estimating PM2.5 is also given. The rules were developed at a meeting of Pechan and Associates on January 5 and 6, 1984 with fine particulate experts, Dr. John Trijonis and Dr. Rudolf Husar.

§ 2. Rules for Calculating Beta Extinction Coefficients

a) Eliminate the observation's measure if there is any precipitation or fog for that reading.

b) Eliminate the day if there are not at least two (good) readings for that day.

c) Calculate the daily arithmetic mean for visibility in miles, V, based on at least two values/day.

d) Apply the relative humidity correction, C from Table 2, to the visibility observation, V, by letting $V' = V/C$, where C varies depending on the observation's corresponding relative humidity range.

e) For each day, calculate the fine particulate concentration, $B = 18.7/Vz'$. (B has no intrinsic meaning.)

f) Calculate FP, based on Trijonis' equation as a function of B.

$$FP = 8.4 + 7.5 * B$$

B-V Relationship

The relationship between the extinction coefficient (β_{ext}) expressed in units of km^{-1} and visual range in km (V) is expressed by the equation:

$$\beta_{ext} = \frac{K}{V\ (in\ km)} \tag{1}$$

Although the Koschmieder formula theoretically defines K = 3.91 based on a 2% contrast detection threshold, both John Trijonis and Rudy Husar in the visibility meetings on January 5 and 6, 1984 at Pechan offices, agreed that K should be 3.0. The support for 3.0 is that airport visibility data underestimates true visual range (defined as the distance at which the contrast for a perfectly black target is reduced to 2%), so that a 5% contrast detection threshold is assumed rather than a 2% threshold.

Since airport visibility data are in miles, rather than convert to km, it was determined that we should use the equation:

$$B = \beta_{ext}\ (in\ (10\ km)^{-1}\ \ units) = \frac{K'}{V'\ (in\ mi),} \tag{2}$$

where $K' = 18.7$ and B is expressed in $(10\ km)^{-1}$ units.

The derivation is as follows:

$$\beta_{ext}\ (in\ (10\ km^{-1}\ units) = \frac{3.0}{V(10\ km)}$$

$$= \frac{3.0}{.1609 \times V'\ (in\ mi)},\ (since\ 1\ mi = 1.609\ km\ or\ .1609\ (10\ km)$$

$$= \frac{18.7}{V'\ (in\ mi)} \quad (in\ (10\ km)^{-1}\ units) \tag{3}$$

§ 3. Data Provided to Loma Linda

Using the above procedures, Pechan and Associates provided Loma Linda University with daily extinction coefficient values for each of the 10 airports for all days of the years 1966 through 1986 for which the airport had valid data.

§ 4. Specification of Airport Areas for Visibility Based Fine Particulate Estimates.

Meetings were held at the California Air Resources Board (CARB) offices in Sacramento and at Loma Linda University in the Spring of 1988 to decide with zip code areas could be assigned to each of the 10 airports listed in Table 1 for the purpose of estimating fine particulate ambient concentrations based on visibility data recorded at the airports. Participants in the meetings included Bart Ostro, Tony Van Curen, John Moore and David Abbey. Zip code areas assigned to each airport were given an A or B quality rating to reflect a subjective judgement of the relative accuracy of fine particulate ambient concentrations estimated from the airport. The results of these meetings were documented on four (4) overlays of Western Economic Research zip code maps, copies of which are retained at LLU and at CARB. In addition to this the zip code areas for Bakersfield and Fresno are listed. A complete listing of all of the zip codes for each airport is given in Table 3. Table 4 gives counts of the number of AHSMOG study subjects who completed the 1977 Questionnaire according to residence during, or since, 1966, within airport areas. All subjects who had lived for one month or more outside an airport area were excluded. Table 5 applies the exclusion criteria of the AHSMOG incidence cohort to those included from Table 4 to obtain the AHSMOG cancer incidence subcohort of 3,308 for whom PM2.5 estimates could be formed from airport visibility data.

§ 5. Using Airport Visibility Measures to Estimate PM2.5 and PM10.

The following sections, 6 through 11, describe the estimation methods that were used to form fine particulate ambient concentration estimates based on airport visibility measures.

Two fractions of fine particulates were estimated - those < 10 microns - PM10, and those < 2.5 microns - PM2.5. Because estimation of PM10 based on Total Suspended Particulates (TSP) was found to be superior to estimation of PM10 based on airport visibility measures only the PM2.5 estimates were utilized for health effects analyses.

§ 6. Selection of Stations for Airport Areas

Table 6 shows the fixed site monitoring stations for PM2.5 and PM10 which were selected for each airport area. This Table also shows the years for which PM2.5 and PM10 data were available, the total number of observations used, and the distance of the monitoring station from the airport. The PM2.5 and PM10 data monitored at fixed site monitoring stations on every sixth day, were matched with the daily airport visibility data. Some of the airport areas - Ontario for PM2.5, and Ontario and San Diego for PM10, were divided into sub-regions to allow different regression estimation equations to be used in the different sub-regions. The definition of these sub-regions is as follows: The Ontario basin was divided into three sub-regions - a western region, a central region, and an eastern region. The western region had an eastern boundary on a zip code map corresponding to the Los Angeles/San Bernardino county line and then followed the San Bernardino/Orange county boundary line. The central Ontario region had an eastern boundary corresponding to the division between zip codes 92366 and 92376 which is Rialto, and then the division between zip codes 92335 and 92316, which is Bloomington. Then the boundary followed a San Bernardino/Riverside county boundary to the south end of the Ontario airport basin. Everything in the Ontario east of this eastern boundary to the central sub-region consisted of the eastern sub-region. The eastern sub-region included portions of Riverside, Redlands, and San Bernardino urban areas. For PM10 San Diego was also divided into two sub-regions primarily for testing consistency of regression equations *(see notes in Table 6)*. The zip codes contained in each subregion are given in Table 3. For PM2.5 data no sub-regions of San Diego were formed.

§ 7. Priority Scheme for Matching of Monitored PM2.5 and PM10 Data With Airport Visibility Data.

For some days more than one source of PM2.5 or PM10 data was available. In this case the following priority rule was used to select the most "preferred" data. If more than one station with data was available the closest station was generally used, though in some cases CARB staff deemed a more distant station more reliable than others, and these were used in preference. Different methods, projects, or agencies may have been responsible for generating the PM2.5 or PM10 data. Codes for these are defined in CARB document ARB 440, Appendix D, Revision 1, May 1, 1979. After selection of a preferred station, consideration was given to the preferred method, project, and agency in that order of priority. For PM2.5 the priorities were: Method 53, then 63; all PM2.5 was one project -- Project 22; for agency - A,I, then F. For PM10 priorities were: Method 65, then 61, then 56; Project--11, then 22, then 31; Agency -- A, I, then F.

§ 8. Exclusion of Outliers and Suspect Data.

Scatterplots of PM2.5 versus the daily extinction coefficient β, and PM10 versus β were made for the purpose of detecting bivariate outliers. In addition to examining the data for bivariate outliers the data were constrained to satisfy the inequality that PM2.5 must be $\leq$ PM10 which in turn must be $\leq$ TSP. Any data that did not satisfy this inequality were coded as missing, in which case the priority algorithm sought to replace the missing data with another observation of lower priority which did satisfy the inequality.

§ 9. Detection of Seasons and Tests for Pooling of Seasonal Data.

After removing the outliers, which in most cases, were greater than 3 standardized residuals from the regression equation, the residuals from the refitted regression equation were plotted by month to determine it season specific regressions might be needed. For a number of regression sets, nonrandom patterns of regression residuals by month were noted and a "data suspected" season was determined by visual examination. General F tests were conducted to determine if the season specific regressions could be pooled into an overall regression for the entire year. The results of these F tests are shown in Table 7. Before performing the general F test, two-tailed F tests were conducted to determine if the assumption of equality of residual of mean square errors for the season specific regressions were satisfied. Results of the two-tailed F tests are also shown in Table 7. When the two-tailed F tests were statistically significant ($p < 0.05$), a log transformation was used for both the fine particulate and the daily β extinction coefficients in order to stabilize the variance. Then the general F test was conducted on the transformed data. If no season was detected from the data, an imposed season of April through October, versus November through March, was tried. For PM2.5 the results of the general F test indicated that season specific regressions were needed for every area with the exception of San Diego, for which there was insufficient data to test for a seasonal effect. For PM10, most areas also required season specific regressions. The exceptions for PM10 were Bakersfield, Ontario East sub-region, and Stockton. As for PM2.5, San Diego lacked sufficient data by month for PM10 to determine whether or not there were seasonal effects.

Previous work by Tony Van Curen at CARB had indicated that Sacramento followed an April through October, versus November through March season. For all areas where we had failed to detect a seasonal effect or where the season that we detected from the data differed from these seasons, the April through October/November through March seasons were tried. If the resultant regressions improved the R^2 the April through October/November through March seasons were adopted.

§ 10. Pooling Over Areas

In addition to attempting to pool seasons for airport areas we also attempted to pool adjacent airport areas. General F tests were conducted to determine if adjacent areas could be pooled. These F tests were conducted on a seasonal basis using the same seasons for both areas if seasonal specific regressions were required. In some cases, this required adjusting the season of one or the other area. The results of the general F tests for pooling areas is shown in Table 8. If adjacent areas had sufficient data points to give reliable area specific regressions we kept them separate even though the general F tests indicated that they could have been pooled.

§ 11. Final Regression Estimation Equations.

The final regression equations for estimating PM2.5 or PM10 from the daily β extinction coefficients are given in Table 9. The symbol b in that Table is used to denote the daily β extinction coefficient. A split halves R^2 is also given in Table 9, this was computed by using a random half of each regression subset to predict the second half and correlating predicted with observed values in the second half. The split halves R^2 gives an unbiased estimate of the actual R^2 that might be achieved for new applications of data. A split halves procedure conducted over all the paired data rather than each specific regression indicated a correlation between estimated and observed PM2.5 of 0.82.

§ 12. Estimating PM2.5 From Both Visibility and TSP

Table 9 also gives the R^2 of a multiple regression of PM2.5 on β extinction coefficient and TSP. By comparing the R^2 with β extinction coefficient alone one can see how much the prediction is improved by adding TSP. In every case there is some improvement in the multiple R^2 by adding TSP, however, a decision not to use TSP to predict PM2.5 was made, as to do so would have restricted the estimation of PM2.5 to every sixth day when TSP was measured and also to years when TSP was adequately monitored. Note that the β extinction coefficient is available on airports on a daily basis for each of the years 1966 - 1986.

§ 13. Could an Overall National Regression Have Been Used For PM2.5?

To determine whether or not the overall national regression equation of Trijonis, used by Pechan and Associates, could have been used to estimate PM2.5, we computed 95% confidence intervals for the intercept and slope of the area specific non-seasonal regression equations. Note that non-seasonal regression equations had to be used as the Trijonis equation was non-seasonal. The results are shown in Table 10. There were only two areas where the Trijonis equation coefficients fell within the confidence intervals. These were Ontario central and Sacramento. However, both of these areas required seasonal specific regressions. Hence, we were unable to use the national regression equations of Trijonis.

§ 14. Could PM2.5 Have Been Estimated from TSP?

To determine whether or not PM2.5 could have been better estimated from TSP than from visibility data, site and seasonal regression equations were formed for estimating PM2.5 from TSP and for estimating PM2.5 from visibility for those years and areas of California where all three types of data were available. Table 11 of this appendix gives the R^2 for estimating PM2.5 from TSP and from visibility data, as measured with the β extinction coefficient, for each season/area and compares the average R^2 over all season areas. The average R^2 for estimating PM2.5 from TSP was 0.414; that for estimating PM2.5 from visibility, 0.577. Hence, the decision was made to estimate PM2.5 from visibility.

§ 15. Estimating Cumulative Ambient PM2.5 for Study Participants.

Using the regression estimation equations of Table 9, estimates of 24 hour mean ambient concentrations of PM2.5 were formed from the daily β extinction coefficients for each day of the years 1966 through 1986. Monthly means were then formed as well as exceedance frequencies in excess of each of the cutoffs 20, 30, and 40 $\mu g/m^3$. In forming monthly means the average of the estimated PM2.5 values over the days where β was observed was computed. The percentage of days where β was observed in excess of each of the cut-offs 20, 30, and 40 $\mu g/m^3$ was converted to an exceedance frequency statistic by multiplying by the number of hours (days) in a month. Units of hours were used for comparability with gaseous pollutants.

Monthly values were assigned to each zip code in the area of the appropriate regression set and these were then applied to the residence histories of study participants for the same period 1966 through 1986. Work location zip codes were not used. Monthly statistics for study participants were cumulated over years according to three time periods - 1966 through March 1977, April 1977 through December 1986, and 1966 through 1986.

Table 1. Ten California Airports For Which Visibility Data Was Obtained

1. Bakersfield (Meadows)
2. Fresno (Hammer)
3. Los Angeles City (LAX)
4. Los Angeles (Long Beach - Dougherty)
5. Sacramento (Executive)
6. San Diego (Lindbergh Field)
7. San Francisco (Alameda Naval Air Station)
8. San Jose (Sunnyvale - Moffett)
9. Stockton (Stockton Metro)
10. Ontario (Ontario International)

Table 2. Relative Humidity Correction Table

Relative Humidity Range	Correction, C
0 to <35	1.20
35 to <45	1.14
45 to <55	1.04
55 to <65	1.00
65 to <73	0.92
73 to <78	0.83
78 to <83	0.69
83 to <88	0.61
88 to <93	0.32
93 to <100	0.15

Quality 'A' Zip codes for PM2.5 from Visibility Regressions
by Raoul Burchette

ALAMEDA
(01)
94102
94103
94104
94105
94107
94108
94110
94111
94112
94114
94117
94124
94127
94131
94134
94501
94601
94602
94603
94605
94606
94607
94608
94609
94610
94611
94612
94613
94614
94618
94619
94621
94625
94626
94627
94705

SAN JOSE
(02)
94035
94040
94041
94043
94086
94087
94089
94301
94303
94304
94305
94306
94536
94538
94560
95002

BAKERSFIELD
(03)
93220
93241
93304
93304
93305
93306
93307
93308
93309
93311
93312
93313

FRESNO
(04)
93612
93616
93625
93648
93650
93652
93654
93657
93662
93701
93702
93703
93704
93705
93706
93710
93711
93721
93722
93725
93726
93727
93728

LOS ANGELES
(05)
90003
90008
90016
90018
90019
90024
90025
90034
90035
90036
90037
90043
90044
90045
90047
90048
90049
90056
90061
90062
90064
90066
90067
90069
90077
90094
90210
90211
90212
90230
90232
90245
90247
90248
90249
90250
90254
90260
90266
90272
90274
90277
90278
90291
90292
90293
90301
90302
90303
90304
90305
90401
90402
90403
90404
90405
90501
90502
90503
90504
90505
90717

LONG BEACH
(06)
90021
90220
90222
90262
90620
90623
90630
90680
90701
90706
90712
90713
90715
90716
90720
90723
90740
90745
90746
90747
90804
90805
90806
90807
90808
90810
90814
90815
90840
92640
92641
92643
92644
92645
92647
92655
92683
92804

ONTARIO
CENTRAL
(07)
91701
91709
91710
91730
91739
91743
91761
91762
91763
91764
91786
92335
92336

ONTARIO
EAST
(08)
91719
91720
91752
91760
92316
92318
92324
92346
92354
92369
92373
92374
92376
92401
92404
92405
92407
92408
92409
92410
92411
92501
92503
92504
92505
92506
92507
92508
92509

ONTARIO
WEST
(09)
91010
91702
91706
91711
91722
91723
91724
91740
91744
91748
91750
91765
91766
91767
91768
91773
91789
91790
91791
91792

SACRAMENTO
(10)
95605
95608
95610
95621
95626
95628
95655
95660
95662
95670
95673
95691
95814
95815
95816
95817
95818
95819
95820
95821
95822
95823
95824
95825
95826
95827
95828
95829
95830
95831
95832
95833
95834
95835
95836
95837
95838
95841
95842
95852
95864

SAN DIEGO
(11)
92010
92011
92032
92037
92050
92101
92102
92103
92104
92105
92106
92107
92108
92109
92110
92111
92113
92114
92115
92116
92117
92118
92120
92122
92123
92133
92135
92139
92140
92154
92155

Quality 'B' Zip codes for PM2.5 from Visibility Regressions by Raoul Burchette

ALAMEDA
(01)
94530
94541
94542
94544
94545
94577
94578
94579
94580
94702
94703
94704
94706
94707
94708
94709
94710
94720
94805

SAN JOSE
(02)
94002
94005
94010
94022
94025
94030
94061
94063
94065
94066
94070
94080
94304
94401
94402
94403
94404
94539
95008
95014
95035
95050
95051
95054
95070
95110
95111
95112
95113
95116
95117
95118
95121
95122
95123
95124
95125
95126
95128
95129
95130
95131
95133
95134
95136

FRESNO
(04)
93609
93615
93618
93631
93637
93638
93642
93646
93647
93666

LOS ANGELES
(05)
90001
90002
90004
90005
90006
90007
90010
90011
90012
90013
90014
90015
90017
90020
90021
90021
90026
90027
90028
90029
90038
90039
90046
90057
90058
90059
90068
90071
90255

LONG BEACH
(06)
90242
90650
90710
90731
90732
90744
90802
90803
90813
90822
92626
92627
92666
92668
92680
92701
92703
92704
92705
92706
92707
92708
92710
92801
92802
92805

ONTARIO EAST
(08)
92360
92370
92387
92388
92518
92621
92686
92808

ONTARIO WEST
(09)
90601
90602
90603
90604
90605
90631
91006
91016
91024
91731
91732
91733
91745
91746
91770
91775
91776
91780

SACRAMENTO
(10)
95630
95678

SAN DIEGO
(11)
92002
92007
92008
92009
92013
92014
92019
92020
92021
92024
92041
92045
92054
92056
92067
92068
92071
92075
92077
92078
92119
92121
92124
92126
92127
92128
92129
92130
92131
92145

Table 4. Number of AHSMOG Study Subjects Completing 1977 Questionnaire According to Residence Within Airport Areas.

EXCLUDED:	
2,639	subjects lived outside of airport areas for total time of the post 1966 residence histories.
446	subjects lived part of the time outside of airport areas and part of the time within a quality code "A" zip code area near one or more of the airports.
275	subjects lived part of the time outside of airport areas and part of the time within a quality code "B" zip code near one or more of the airports.
96	subjects lived part of the time outside of airport areas and part of the time within a quality code "A" zip code area and part of the time within a quality code "B" zip code area.
3,456	TOTAL EXCLUDED
INCLUDED:	
2,508	subjects lived only within a quality code "A" zip code area for the total time from 1966.
1,153	subjects lived within a quality code "B" zip code area for the total time from 1966 only.
207	subjects lived within both a quality code "A" zip code area and a "B" zip code area for the total time from 1966 only.
3,868	TOTAL INCLUDED
7,324	TOTAL AHSMOG SUBJECTS

# of addresses/subject	frequency
1	5,355
2	1,301
3	441
4	140
5	53
6	34
	7,324

Table 5. Airport Cancer Incidence Subcohort Exclusions.

AHSMOG Study subjects who completed 1977 questionnaire and were in initial fine particulate analysis file from Table 2.		
A.	2,508	subjects who lived totally in a quality "A" airport zip code area from 1966 to 1977.
B.	1,153	subjects who lived totally in a quality "B" airport zip code area from 1966.
C.	207	subjects who lived totally in both a quality "A" and "B" airport zip code area from 1966.
	3,868	**TOTAL**
Using Jim Peter's (AHSMOG Cohort) exclusion criteria, the following individuals were removed from the 3,868.		
EXCLUDED SUBJECTS:		
1.	495	subjects who were not "incidence population"
2.	12	additional subjects who were not SDA (baptized or nonbaptized).
3.	53	additional subjects who were current smokers.
	560	**TOTAL EXCLUDED SUBJECTS**
The remaining subjects for the Airport Cancer Incidence Subcohort are:		
A.	2,182	subjects who lived totally in a quality "A" airport zip code area from 1966.
B.	955	subjects who lived totally in a quality "B" airport zip code area from 1966.
C.	171	subjects who lived totally in both a quality "A" and "B" airport zip code area from 1966.
	3,308	**TOTAL**

Regression Sets for Airports and Stations for PM2.5 and PM10

Set 1 PM2.5 Alameda NA Airport = 8

Stations: Station 1, 9000304, San Francisco-23rd St.
Years '79, 80, 81. Distance = 7.564 km., 80 observations used.

Set 2 PM2.5 Bakersfield Airport = 10

Stations: Station 3, 1500203 Bakersfield-Chester Street
Years '80, 81, 82, 83, 84. Distance = 8.137 km., 172 observations used.

Set 3 PM2.5 Fresno Airport = 9

Stations: Station 2, 1000234, Fresno - Olive
Years '82, 83. Distance = 4.774 km., 45 observations used.

Set 4 PM2.5 LAX Airport = 7

Stations: Station 1, 7000086, West LA-Robertson
Years '79, 80, 81 through July. Distance = 12.108 km., 101 observations used.

Station 3, 7000595, Long Beach-San Antonio Drive (possibly if regressions similar to Station 1) Distance = 23.478 km., 59 observations used.
Years '81, August through December, 82, January through October

Station 4, 7000072, North Long Beach (possibly if regression similar to Station 1) Distance = 23.826 km., 155 observations used.
Years '82, November, December. '83, 84.

Total number of observations for Set 4 = 315.

Regression Sets for Airports and Stations for PM2.5 and PM10

Set 5 PM2.5 Long Beach Airport = 6

Stations: Station 1, 7000595, Long Beach San Antonio Drive
Years '81, 82, through October. Distance = 4.573 km., 61 observations used.

Station 2, 7000072, North Long Beach
Years '82 November and December, '83, 84.
Distance = 4.650 km., 154 observations used.

Total number of observations for Set 5 = 215.

Set 6 PM2.5 San Jose - Moffett Airport = 5

Stations: Station 2, 4300382, San Jose 4th Street
Years '79, 80, 81. Distance = 16.852 km., 118 observations used.

Set 7 PM2.5 Ontario Central Subregion Airport = 4 Subregion = (C)entral

Stations: Station 2, 3600197, Fontana - Arrow Hwy.
Years '83, 84. Distance = 10.701 km., 59 observations used.

Set 8 PM2.5 Ontario Subregion East Airport = 4 Subregion = (E)ast

Stations: Station 3, 3300144, Riverside - Roubidoux
Years '79, 80, 81, 82, 83. Distance = 19.079 km., 154 observations used.

Station 4, 3300146, Riverside - Magnolia
Years '84. Distance = 22.904 km., 181 observations used.

Total number of observations for Set 8 = 335.

Set 9 PM2.5 Ontario Subregion West Airport = 4 Subregion = (W)est

Stations: Station 5, 7000591, Glendora - Laurel
Years '81, 82, 83, 84. Distance = 24.121 km., 252 observations used.

Regression Sets for Airports and Stations for PM2.5 and PM10

Station 6, 7000060, Azusa
Years '79, 80, 81. Distance = 30.172 km., 85 observations used.

Total number of observations for Set 9 = 337.

Set 10 PM2.5 Ontario Subregion West Check Set Airport = 4 Subregion = X

Stations: (Long Beach Airport) Station 6, 7000083, Pasadena - Walnut
Years '79, 80, 81. 27 observations used.

Note: Compare to set 9. If not similar, exclude portion of population in north end of Ontario B area from this study. See June 12, 1990 Sacramento Meeting Minutes.

Set 11 PM2.5 Sacramento Airport = 3

Stations: Station 3, 3400302, Sacramento Metro - Tower
Years '83, 84. Distance = 21.576 km., 69 observations used.

Set 12 PM2.5 San Diego Airport = 2

Stations: Station 3, 8000131, El Cajon - Redwood Avenue
Years '83. Distance = 23.403 km., 22 observations used.

Set 13 PM10 Alameda NA Airport = 8

Stations: Station 2, 9000306, San Francisco - Arkansas
Years '86. Distance = 8.059 km., 37 observations used.

Set 14 PM10 Bakersfield Airport = 10

Stations: Station 3, 1500203, Bakersfield - Chester Street
Years 82, 83, 84, 85, 86, Distance = 8.137 km.,
191 observations used.

Regression Sets for Airports and Stations for PM2.5 and PM10

Set 15 PM10 Fresno Airport = 9

Stations: Station 2, 1000234, Fresno - Olive
Years 82, 83, 84, 85, 86. Distance = 4.774 km.,
159 observations used.

Set 16 PM10 LAX Airport = 7

Stations: Station 4, 7000072, North Long Beach
Years 83, 84, 85, 86. Distance = 23.826 km., 152 observations used.

Set 17 PM10 LA Long Beach Airport = 6

Stations: Station 2, 7000072, North Long Beach
Years 83, 84, 85, 86. Distance = 4.650 km., 149 observations used.

Set 18 PM10 San Jose - Moffett Airport = 5

Stations: Station 2, 4300382, San Jose - 4th Street
Years 84, 85, 86. Distance = 16.852 km., 113 observations used.

Set 19 PM10 Ontario Central Subregion Airport = 4 Subregion = (C)entral

Stations: Station 1, 3600171, Ontario airport
Years 84, 85, 86. Distance = 2.191 km., 120 observations used.

Station 2, 3600197, Fontana - Arrow Hwy
Years 83, Distance = 10.701 km., 66 observations used.

Total observations for Set 19 = 186.

Set 20 PM10 Ontario Subregion East Airport = 4 Subregion = (E)ast

Stations: Station 3, 3300144, Riverside - Rubidoux
Years 83, 84, 85, 86. Distance = 19.079 km.,
147 observations used.

Station 4, 3300146, Riverside - Magnolia
Years 83, 84. Distance = 22.906 km., 72 observations used.

Total observations for Set 20 = 219.

Set 21 PM10 Ontario Subregion East Check Set Airport = 4 Subregion = F

Stations: Station 7, 3600194, San Bernardino - E 3rd
Years 86. Distance 31.366 km., 6 observations used.

Station 8, 3600203, San Bernardino - 4th Street
Years 86. Distance = 31.710 km., 27 observations used.

Total observations for Set 21 = 33.

Note: Use this check set to see if same as set 20. If so, consider pooling data. If not, consider a northeast subregion and a southeast subregion.

Set 22 PM10 Ontario Subregion West Airport = 4 Subregion = (W)est

Stations: Station 5, 7000591, Glendora - Laurel
Years 83, 84. Distance = 24.121 km., 96 observations used.

Station 6, 7000060, Azusa
Years 84, 85, 86. Distance = 30.172 km., 118 observations used.

Total observations for Set 22 = 214.

Set 23 PM10 Sacramento Airport = 3

Stations: Station 1, 3400277, Sacto - H.D. Stockton Blvd.
Years 86. Distance = 5.723 km., 46 observations used.

Station 2, 3400295, Sacramento- Del Paso Manor
Years 86. Distance = 15.812 km., 6 observations used.

Station 3, 3400302, Sacramento - Metro-Tower
Years from 6/83-11/85. Distance = 21.576 km., 92 observations used.

Regression Sets for Airports and Stations for PM2.5 and PM10

Station 4, 3400293, Citrus Heights - Sunrise Blvd.
Years 83, 84, 85, 86. Distance = 28.268 km., 60 observations used.

Total observations for Set 23 = 204.

Set 24 PM10 San Diego Airport = 2

Stations: Station 1, 8000123, San Diego - Overland
Years 86. Distance = 6.407 km., 19 observations used.

Set 25 PM10 San Diego Check Set 1 Airport = 2 Subregion = 1

Stations: Station 2, 8000114, Chula Vista
Years 86. Distance = 17.039 km., 47 observations used.

Note: If data is similar to set #24, consider pooling the data of set #25 with #24.

Set 26 PM10 San Diego Check Set 2 Airport = 2 Subregion = 2

Stations: Station 3, 8000131, El Cajon - Redwood Avenue.
Years 86. Distance = 23.403 km., 27 observations used.

Note: Compare this regression to set 24 and set 25 to see if separate regression needed for San Diego B quality area and to see if can use El Cajon PM2.5 data for all San Diego. See Sacramento Meeting, June 12, Notes.

Set 27 PM10 Stockton Airport = 1

Stations: Station 1, 3900252, Stockton - Hazelton
Years 84, 85, 86. Distance = 6.714 km., 86 observations used.

Note: Stockton has no PM2.5 data. If Stockton PM10 regression is similar to that for regression sets 13, 14, 15, 18, or 23, we will consider using the PM2.5 regression from one or more of these sites for Stockton.

Table 7
F Tests for Pooling of Seasons for PM2.5 and PM10 in Airport Areas

		SUMMER		WINTER			COMBINED SEASONS			
Sel	Name	Resid. Mean Square Error	df	Resid. Mean Square Error	df	2 Tailed f Value	Resid. Mean Square Error	df	Seasonal f Test	df
01	Alameda NAS PM2.5	34.38 (May-Oct)	45	67.89 (Nov-Apr)	31	1.97	58.45	78	9.43**	2/76
02	Bakersfield PM2.5	50.13 (Apr-Oct)	106	208.24 (Nov-Mar)	62	4.15*				
	Bakersfield PM2.5 Log(X) by Log(Y)	.02303 (Apr-Oct)	106	.02869 (Nov-Mar)	62	1.25	.03895	170	47.81**	2/168
03	Fresno PM2.5	29.04 (Apr-Oct)	26	152.89 (Nov-Mar)	15	5.26*				
	Fresno PM2.5 Log(X) by Log(Y)	.02940 (Apr-Oct)	26	.08610 (Nov-Mar)	15	2.93*	.09597	43	20.65**	2/41
04	LAX PM2.5	52.77 (Mar-Sep)	191	176.14 (Oct-Feb)	120	3.34*				
	LAX PM2.5 Log(X) by Log(Y)	.02350	191	.03617	120	1.54	.03904	313	59.70**	2/311
05	Long Beach PM2.5	37.86 (Mar-Sep)	134	121.01 (Oct-Feb)	77	3.20*				
	Long Beach PM2.5 Log(X) by Log(Y)	.01900 (Mar-Sep)	134	.02924 (Oct-Feb)	77	1.54	.03124	213	40.85**	2/211
06	San Jose PM2.5	40.66 (Mar-Oct)	82	140.71 (Nov-Feb)	32	3.46*				

Table 7 (continued)
F Tests for Pooling of Seasons for PM2.5 and PM10 in Airport Areas

	San Jose PM2.5 Log(X) by Log(Y)	.03450 (Mar-Oct)	82	.06026 (Nov-Feb)	32	1.75	.05612	116	21.00**	2/114
07	Ontario Central PM2.5	69.87 (Mar-Sep)	32	179.42 (Oct-Feb)	23	2.57*	153.58	57	10.34**	2/55
	Ontario Central PM2.5 Log(X) by Log(Y)	.01655 (Mar-Sep)	32	.10721 (Oct-Feb)	23	6.48*				
08	Ontario East PM2.5	218.66 (Mar-Sep)	197	433.39 (Oct-Feb)	134	1.98	322.96	333	10.46**	2/331
09	Ontario West PM2.5	175.31 (Mar-Sep)	179	145.81 (Oct-Feb)	154	1.20	169.96	335	9.59**	2/333
11	Sacramento PM2.5	20.02 (Apr-Oct)	35	14.94 (Nov-Mar)	30	1.34	20.15	67	5.72**	2/65
13	Alameda NAS PM10	63.09 (Apr-Aug)	21	55.15 (Sep-Mar)	12	1.14	70.32	35	3.94*	2/33
14	Bakersfield PM10	279.16 (Apr-Oct)	123	557.29 (Nov-Mar)	64	2.00	376.73	189	1.60	2/187
15	Fresno PM10	182.26 (Mar-Aug)	80	458.01 (Sep-Feb)	75	2.51*				
	Fresno PM10 Log(X) by Log(Y)	.02741 (Mar-Aug)	80	.02810 (Sep-Feb)	75	1.03	.03389	157	18.39**	2/155
16	LAX PM10	134.86 (Mar-Sep)	87	345.87 (Oct-Feb)	61	2.56*				
	LAX PM10 Log(X) by Log(Y)	.01146 (Mar-Sep)	87	.02272 (Oct-Feb)	61	1.98	.01881	150	13.61**	2/148
17	Long Beach PM10	100.54 (Mar-Sep)	88	270.23 (Oct-Feb)	57	2.69*				

F Tests for Pooling of Seasons for PM2.5 and PM10 in Airport Areas

	Long Beach PM10 Log(X) by Log(Y)	.00867 (Mar-Sep)	88	.01473 (Oct-Feb)	57	1.70	.01305	147	14.26**	2/145
18	San Jose PM10	125.28 (Mar-Oct)	77	551.71 (Nov-Feb)	32	4.40*				
	San Jose PM10 Log(X) by Log(Y)	.01803 (Mar-Oct)	77	.02843 (Nov-Feb)	32	1.58	.02697	111	16.50**	2/109
19	Ontario Central PM10	291.79 (Mar-Jun)	57	541.45 (Jul-Feb)	125	1.86	500.22	184	8.34**	2/182
20	Ontario East PM10	849.83 (Apr-Oct)	119	917.36 (Nov-Mar)	96	1.08	877.98	217	0.75	2/215
22	Ontario West PM10	214.50 (Dec-Apr)	85	380.25 (May-Nov)	125	1.77	364.52	212	18.38**	2/210
23	Sacramento PM10	166.52 (Apr-Oct)	133	296.95 (Nov-Mar)	67	1.78	222.69	202	6.99**	2/200
27	Stockton PM10	220.00 (Apr-Oct)	55	294.99 (Nov-Mar)	27	1.34	244.53	84	.97	2/82

Note: On the seasonal f test, one asterisk = statistical significance at the .05 level, two asterisks = statistical significance at the .01 level.

F Tests for Pooling of Airport Areas for PM2.5 and PM10

(Sets 1 through 12 are PM2.5; Sets 13 through 25 are PM10)

								Combined Sets		
Months	Set1	RMSE1	df1	Set2	RMSE2	df2	2TF	PRMSE	df	f val
Apr-Oct	01	34.76	49	06	40.74	72	1.17	37.95	2/121	0.40
Nov-Mar	01	71.25	27	06	139.10	42	1.95	111.76	2/69	0.75
Mar-Sep	04	52.77	191	05	37.86	134	1.39	50.73	2/325	15.39**
Oct-Feb	04	176.14	120	05	121.01	77	1.48	153.25	2/197	0.13
Mar-Sep	07	69.87	32	08	218.66	197	3.13*			
Mar-Sep*	07	.01655	32	08	.03189	197	1.93	.03067	2/229	4.59*
Oct-Feb	07	179.42	23	08	433.39	134	2.42*			
Oct-Feb*	07	.10721	23	08	.08155	134	1.31	.08469	2/157	0.43
10's Mos	09	150.49	203	10	165.87	25	1.10	164.35	2/228	10.20**
Apr-Oct	13	74.74	28	18	117.64	68	1.57	107.85	2/96	2.27
Nov-Mar	13	58.44	5	18	547.80	41	9.37*			
Nov-Mar*	13	.02878	5	18	.03541	41	1.23	.03550	2/46	1.56
Apr-Oct	15	261.76	101	23	166.52	133	1.57	215.15	2/234	5.27*
Nov-Mar	15	506.56	54	23	296.95	67	1.71	499.38	2/121	18.15**
Mar-Sep	16	134.85	87	17	100.54	88	1.34	116.62	2/175	0.27
Oct-Feb	16	345.87	61	17	270.23	57	1.28	311.06	2/118	1.34
Apr-Oct	19	395.78	101	20/21	756.27	142	1.91	672.84	2/243	14.42**
Nov-Mar	19	592.57	81	20/21	855.48	106	1.44	740.04	2/187	.80
21's Mos	20	889.49	140	21	301.99	31	2.95*			
21's Mos*	20	.03611	140	21	.01698	31	2.13*	.03257	2/171	1.60
Mar-Sep	24	31.33	17	25	44.60	30	1.42	38.75	2/47	0.35
Com Mos	24	26.73	10	26	41.51	11	1.55	35.18	2/21	1.30
Com Mos	25	54.70	29	26	113.56	25	2.07*	88.70	2/54	3.31*

Three Way f tests

A three way f test between Set 19, Ontario Central PM10, Set 20, Ontar East PM10 and Set 21, Ontario East Check Set PM10, was conducted.

1. Is April through October on non transformed data.
2. Is April through October on log(x) by log(y) transformed data.
3. Is November through March on non transformed data.
4. Is November through March on log(x) by log(y) transformed data.

	Set1	RMSE1	df1	Set2	RMSE2	df2	Set3	RMSE3	df3	2TF	PRMSE	df	f value
1.	19	395.78	101	20	849.83	119	21	275.47	21	3.08*			
2.	19	.01851	101	20	.03184	119	21	.00859	21	3.71*			
3.	19	592.57	81	20	917.36	96	21	229.70	8	3.99*			
4.	19	.04994	81	20	.05313	96	21	.03794	8	1.40	.05022	2/187	.42

A three way f test between Set 19, Ontario Central PM10, Set 20, Ontar East PM10 and Set 22, Ontario West PM10, was conducted.

F Tests for Pooling of Airport Areas for PM2.5 and PM10

(Sets 1 through 12 are PM2.5; Sets 13 through 25 are PM10)

1. Is April through October on non transformed data.
2. Is April through October on log(x) by log(y) transformed data.
3. Is November through March on non transformed data.
4. Is November through March on log(x) by log(y) transformed data.

	Set1	RMSE1	df1	Set2	RMSE2	df2	Set3	RMSE3	df3	2TF	PRMSE	df	f value
1.	19	395.78	101	20	849.83	119	22	368.12	124	2.31*			
2.	19	.01851	101	20	.03184	119	22	.02624	124	1.72	.02842	2/346	18.90**
3.	19	592.57	81	20	917.36	96	22	292.56	86	3.14*			
4.	19	.04994	81	20	.05313	96	22	.04083	86	1.30	.05404	2/265	18.40**

Notes:
RMSE=Residual Mean Square Error
df=degrees of freedom
2TF=two tailed f test
PRMSE=Pooled Residual Mean Square Error

F tests with an asterisk following the months in the months colu indicate that this f test was performed on data that was log(x) by log transformed due to statistically significant two tailed f values on transformed data.

An asterisk following an f value indicates statistical significance the .05 level, whereas two asterisks indicate statistical significance to .01 level.

Table 9

Final Regression Equations for Estimating PM2.5 and PM10 From Airport Visibility Data (b = extinction coefficient)

Includes Sets	Months	Name	Equation	r^2	Split Halves r^2	Mult. Reg. With TSP r^2	n
01/06	Apr-Oct	Alameda NAS PM2.5	Y = -1.70375 + 6.80418b	.568	.325	.764	51
06/01	Apr-Oct	San Jose PM2.5	Y = -1.70375 + 6.80418b	.568	.325	.764	74
01/06	Nov-Mar	Alameda NAS PM2.5	Y = 3.10500 + 8.53866b	.612	.545	.792	29
06/01	Nov-Mar	San Jose PM2.5	Y = 3.10500 + 8.53866b	.612	.545	.792	44
02	Apr-Oct	Bakersfield PM2.5	Y = 6.24020 + 6.38718b	.354	.328	.385	108
02	Nov-Mar	Bakersfield PM2.5	Y = 12.11259 + 10.29508b	.756	.630	.772	64
03	Apr-Oct	Fresno PM2.5	Y = -6.46577 + 12.57324b	.544	.440	.575	28
03	Nov-Mar	Fresno PM2.5	Y = 15.73720 + 11.38760b	.557	.607	.708	17
04	Mar-Sep	LAX PM2.5	Y = 12.67801 + 4.44842b	.388	.410	.464	193
05	Mar-Sep	Long Beach PM2.5	Y = 7.57503 + 7.99007b	.657	.521	.675	136
04/05	Oct-Feb	LAX PM2.5	Y = 20.52089 + 6.94479b	.530	.526	.618	122
05/04	Oct-Feb	Long Beach PM2.5	Y = 20.52089 + 6.94479b	.530	.526	.618	79
07	Mar-Sep	Ontario Central PM2.5	Y = 8.71381 + 6.71479b	.689	.460	.738	34
07	Oct-Feb	Ontario Central PM2.5	Y = 14.44896 + 9.01624b	.717	.616	.791	25
08	Mar-Sep	Ontario East PM2.5	Y = 11.77931 + 8.52377b	.564	.619	.644	199

Table 9

Final Regression Equations for Estimating PM2.5 and PM10 From Airport Visibility Data (b = extinction coefficient) (continued...)

Includes Sets	Months	Name	Equation	r^2	Split Halves r^2	Mult. Reg. With TSP r^2	n
08	Oct-Feb	Ontario East PM2.5	Y = 20.13258 + 8.71492b	.669	.601	.753	136
09	Mar-Sep	Ontario West PM2.5	Y = 12.65940 + 4.05315b	.277	.260	.349	181
09	Oct-Feb	Ontario West PM2.5	Y = 11.30730 + 5.81165b	.701	.651	.777	156
11	Apr-Oct	Sacramento PM2.5	Y = 12.02748 + 4.30092b	.292	.138	No TSP	37
11	Nov-Mar	Sacramento PM2.5	Y = 6.52349 + 7.68613b	.832	.830	No TSP	32
12	All	San Diego PM2.5	Y = 3.08044 + 6.56111b	.677	.920	.778	22
13/18	Apr-Oct	Alameda NAS PM10	Y = -26.14988 + 21.44035b	.421	.526	.795	30
18/13	Apr-Oct	San Jose PM10	Y = -26.14988 + 21.44035b	.421	.526	.795	70
13/18	Nov-Mar	Alameda NAS PM10	Y = -22.23870 + 24.89512b	.667	.531	.851	7
18/13	Nov-Mar	San Jose PM10	Y = -22.23870 + 24.89512b	.667	.531	.851	43
14	All	Bakersfield PM10	Y = 36.03415 + 11.77970b	.468	.530	.693	191
15	Apr-Oct	Fresno PM10	Y = 9.48171 + 20.82224b	.383	.399	.855	103
15	Nov-Mar	Fresno PM10	Y = 18.16987 + 20.43038b	.619	.712	.855	56
16/17	Mar-Sep	LAX PM10	Y = 31.30217 + 9.14594b	.471	.594	.696	89
17/16	Mar-Sep	Long Beach PM10	Y = 31.30217 + 9.14594b	.471	.594	.696	90
16/17	Oct-Feb	LAX PM10	Y = 40.34656 + 11.26031b	.442	.461	.760	63

APPENDIX D

Table 9

Final Regression Equations for Estimating PM2.5 and PM10 From Airport Visibility Data (b = extinction coefficient) (continued...)

Includes Sets	Months	Name	Equation	r^2	Split Halves r^2	Mult. Reg. With TSP r^2	n
17/16	Oct-Feb	Long Beach PM10	Y = 40.34656 + 11.26031b	.442	.461	.760	59
19	Apr-Oct	Ontario Central PM10	Y = 31.40898 + 11.54600b	.638	.728	.800	103
19	Nov-Mar	Ontario Central PM10	Y = 29.43525 + 14.46146b	.696	.669	.888	83
20/21	All	Ontario East PM10 Ontario East Ck Set PM10	Y = 36.48555 + 14.03907b	.604	.626	.912	252
22	Dec-Apr	Ontario West PM10	Y = 22.54945 + 7.55229b	.557	.706	.913	87
22	May-Nov	Ontario West PM10	Y = 37.48712 + 7.73931b	.442	.501	.836	127
23	Apr-Oct	Sacramento PM10	Y = 22.98296 + 16.07002b	.309	.392	.795	135
23	Nov-Mar	Sacramento PM10	Y = 22.29653 + 10.22657b	.489	.401	.794	69
24/25	All	San Diego (Overland/ Chula Vista) PM10	Y = 12.88322 + 10.66265b	.449	.196	.832	66
24/26	All	San Diego (Overland/ El Cajon) PM10	Y = 20.16823 + 8.43344b	.233	.116	.818	46
27	All	Stockton PM10	Y = 35.93943 + 4.22015b	.451	.396	.724	86

Appendix D

Table 10.

Can Trijonis National Regression Equation be Used to Predict PM2.5 from Visibility in California?

SET	BETA VALUE	95% CI	CONSTANT	95% CI	NAME OF SET
01	8.27	(6.66, 9.89)	-2.18	(-5.95, 1.58)	Alameda NAS
02	10.71	(9.58,11.84)	2.55	(-0.69, 5.79)	Bakersfield
03	14.59	(9.82,19.37)	-2.32	(-11.21, 6.56)	Fresno
04	5.98	(5.11, 6.86)	14.83	(12.68,16.98)	LAX
05	7.72	(6.84, 8.60)	11.88	(9.77,13.99)	Long Beach
06	7.70	(6.33, 9.05)	-0.09	(-3.69, 3.51)	San Jose
07	7.27	(5.71, 8.84)	11.82	(6.24,17.40)	Ontario Centra
08	8.41	(7.69, 9.14)	15.74	(12.71,18.77)	Ontario East
09	5.13	(4.60, 5.66)	10.87	(8.69,13.04)	Ontario West
10	4.39	(2.26, 6.51)	24.40	(17.62,31.19)	Ontario West C
11	6.74	(5.51, 7.97)	8.56	(6.93,10.19)	Sacramento
12	6.56	(4.45, 8.67)	3.08	(-.99, 7.15)	San Diego

Question 1: Is 7.5 included in Beta's 95% CI?
Question 2: Is 8.4 included in the Constant's 95% CI?

SET	Q1	Q2
01	Yes	No
02	No	No
03	No	No
04	No	No
05	Yes	No
06	Yes	No
07	Yes	Yes
08	No	No
09	No	No
10	No	No
11	Yes	Yes
12	Yes	No

Conclusion:
Sites where Pechan equation could be used are Ontario Central Sacramento. However, seasonal specific regressions do better for the sites.

Trijonis national regression equation used by Pechan and Associates was:

PM2.5 = 8.4 + 7.5 b,

b = Beta extinction coefficient.

Appendix E

Indirect Estimation Methods for PM10

§ 1. Comparison of Methods

Residence and work location history was available on study participants from 1966 through March of 1987. Cumulative ambient concentrations of PM10 for this 20 year time period were estimated. Because PM10 was not monitored on a consistent state-wide basis prior to 1983 an indirect method for estimating PM10 was used. Two alternative methods were considered. The first method was to estimate PM10 from daily airport visibility data available for ten airports throughout California for the years 1966 through 1986. The second method was to estimate PM10 from total suspended particulates (TSP). TSP were monitored on a state-wide basis in a consistent manner from 1973 through 1987. Some stations began monitoring PM10 as early as 1979 and an increasing number of simultaneous observations of PM10 and TSP occurred from 1982 onwards.

To compare the two methods, 17 seasonal specific regression sets were formed for the ten airports for the purpose of estimating mean 24 hour concentrations of PM10. These regression sets contained, for the years 1979 through 1986, a combined number of 1,618 paired data points of TSP and PM10 and 1,855 paired data points of PM10 and daily beta extinction coefficients derived from visibility data. Seventeen regression equations were formed to estimate PM10 from TSP and 17 other regression equations formed to estimate PM10 from the beta extinction coefficients. The r^2 for each regression equation was computed and the average r^2 overall regression sets compared. The results are shown in Table 1. The average r^2 for PM10 as estimated from visibility data was 0.491. The average r^2 for PM10 estimated from TSP was 0.746. Because of the much higher r^2 the latter method was used.

§ 2. Forming PM10/TSP Regression Estimation Equations

In order to increase the precision of the final regression estimation equations, the geographic areas were redefined and data through 1989 was used in order to include additional simultaneous measures of PM10 and TSP. Ninety-five stations throughout California had some paired PM10/TSP data available between the years 1982 and 1989. Stations used for regression estimation purposes were limited to the 70 of the 95 stations which had at least two sets of paired data points represented from every calendar month. This was done to avoid a possible seasonal bias in the regression estimation equations. The 70 stations are listed in Table 2A along with the number of paired data points for each station and the regression coefficients of PM10 on TSP forcing the regression equation through the origin. Table 2B gives this information for the rest of the 95 stations. Table 3 shows the number of stations and paired data points by year for these 70 stations. Note that there were never 70 stations operating simultaneously.

Different methods, projects, and agencies were responsible for generating the data. Method 91 was the preferred method for TSP. It was always available and hence the only method used for the 70 stations. For PM10 the methods utilized consisted of method 55, 58, 59, 65 or 61. Where multiple methods, projects, or agencies existed for the same day, a prioritization scheme was used to select the most desirable data. For PM10 method 59 was selected as having the highest priority; then 55, 58, 65, 4, 61, followed by all others; project 11, then 22, were given highest priority followed by all others; for agency prioritization was given to agency A, then I then all others. For TSP, as mentioned above method 91 was always available for the 70 stations the priority given to project was 11, 22, 31, and 23, all others. The priority given to agency was A, I, all others.

Regression equations with intercepts were examined to see if there were significant nonzero intercepts. There were 31 of the 70 stations with significantly nonzero intercepts; however, their magnitude was small relative to the average value of an observation and hence considered unimportant. This was further confirmed by noting that there was no upward or downward drift in the residual plots of the stations when the regressions equations were forced through the origin. Regression equations through the origin were considered to be more desirable since if there were no TSP then neither should there be any PM10. The slopes and intercepts of the regression equations and their statistical significancies are shown in Table 4.

The 70 stations were grouped into 38 geographical areas called "groups" at a meeting with Dane Westerdahl, Tony Van Curen, John Moore, David Abbey, and Raoul Burchette on March 29, 1991. The purpose of grouping stations was to provide more data points for stable regressions. Topographical and meteorological considerations were employed in forming the groups so that they would be homogeneous with respect to PM10/TSP characteristics. A list of the stations in each of these groups is given in Table 5.

Some stations did not have complete data by themselves but had observations forming a continuous time sequence with other stations in close proximity (usually the same town). These observations which did not overlap in time were treated as a continuous sequence at one station and the resulting regression equation was applied to both stations. The stations which formed sequences were the two San Bernardino stations, the Simi Valley stations, the Goleta and Santa Barbara stations, the Victorville stations, the Lancaster stations (these were separated in time by about five years of no data), the Oildale stations, and the Mammoth Lakes stations. Other sets of stations which initially appeared to be sequences were really simultaneous measures, such as in Fresno, Modesto, and San Jose.

Regression equations of PM10 on TSP forced through the origin were formed for each of the above groups. Residual plots were made by calendar month for each group, plotting residuals from the regression equation. The residual plots above were inspected for seasonal variation. Where monthly patterns so indicated, two seasons of the year were formed. An example of a residual plot by month indicating two seasons is shown in Figure 1. Season specific regressions were made for each group. A transformation of $\log_{10}(PM10 + 1)$ versus $\log_{10}(TSP + 1)$ was made to stabilize variance. Then the general F-test was conducted for each of the 38 groups which had seasons to determine if the seasons detected from the data could be pooled so that, instead of season specific regressions, an overall group regression could be used. In all cases except one, the seasons could not be pooled. The one area group for which seasons could be pooled was group 11, the Piru group. The seasons for this area were pooled. The results of the general F-test for pooling seasons are given in Table 6.

In addition to examining monthly plots of residuals, we looked at annual plots of residuals by geographic area (but not season specific annual plots). These annual plots did not show annual trends except in two cases: group 3, El Centro, and group 31, San Rafael. One possible explanation for these annual trends was the beginning of the five year drought in 1987. It was felt that nothing could be done about the annual trends and that the magnitude of the trend was not large.

We also examined residual plots by station for each group. The examination of these plots did not show a clear indication for separation of any station from its group though we later did separate Hawthorne from the Los Angeles group, and subdivided group 6, Riverside and San Bernardino Counties.

We then formed season specific regressions using the nontransformed data for those areas indicating specific seasons. We examined these season specific regressions and scatterplots for outliers indicating that months had been put in the wrong season. Only one group, Bethel Island, had a season adjustment, which was to place March in the winter season.

Using the season specific regressions, we sought to pool groups which had:

A. similar seasons
B. similar regression slopes
C. geographical proximity

Regression equations of $\log_{10}(PM10 + 1)$ on $\log_{10}(TSP + 1)$ with intercepts were fitted and general F-tests conducted for residual mean squares to determine where areas could be pooled which had similar seasons. Note that the log transformation was used in order to satisfy the assumption of equal variances in groups to be pooled required by the general F-test. Only pairwise pooling tests were conducted. The criteria for considering pooling were same seasons, similar regression coefficients and geographical proximity. The results of the group pooling attempts are summarized in Tables 7A and 7B. There were only a few stations for which the general F-test indicated that pooling could be done.

Using the results of the general F-test as a guide, a meeting was held with Tony Van Curen, John Moore, Raoul Burchette, and David Abbey on May 1-2, 1991. At this meeting the general F-test on pooling, the regression coefficients, as well as meteorological and physical conditions were taken into account. It was decided not to pool any of the groups but rather to leave them all separate.

Some of the groups were subdivided, specifically, groups 6 and 19, the Riverside-San Bernardino area, and the Bakersfield-Fresno area, respectively, and some of the groups were realigned, specifically, groups 5 and 8, Orange and Los Angeles Counties, respectively. Hawthorne was separated from the Los Angeles group and put with El Toro and Los Alamitos (Orange County), and Burbank was separated from Los Angeles and North Long Beach. The new groupings are listed in Table 8A. The guiding principles for these realignments were as follows.

A. If the General F-test had its equal variance assumption met and indicated that the stations in question should not be pooled, we did not pool them.

B. If the General F-test had its equal variance assumption met and indicated that the stations in question could be pooled, we sometimes decided not to pool them on the basis of other considerations, primarily, geographical and meteorological.

C. If the General F-test did not have its equal variance assumption met and the general F-test indicated that the stations should not be pooled, the results were sometimes overruled to pool on the basis of three considerations: 1.) geographical, 2.) meteorological and/or 3.) quantity of data points.

For each of the new groups season specific regressions of PM10 on TSP were formed using the nontransformed data. The reason for using nontransformed data for final regression estimation equations was to give equal weight to the higher valued observations. (The log transformation would give less weight to these values.) Also, the scatterplots indicated a linear relationship between PM10 and TSP in the nontransformed scale. Regression estimation equations were also formed for each individual station with paired data for the purpose of estimation of PM10 from TSP for years when paired data was not available. These individual station regressions were used in preference to regressions based on grouped data for estimating PM10 at a station which had some years of paired data. A set of paired data from which a final regression estimation equation was formed will henceforth be called a <u>regression set</u>. There were 84 regression sets so formed, but one (the statewide nonseasonal set) was never used. The estimated regression coefficients and their standard errors for each regression set are shown in Tables 7A and 8A.

TSP stations were assigned to regression sets using the following general principles.

A. If a TSP station was a station with paired PM10/TSP data that met the completeness criteria and was the only such station in the city or town, that station was used solely for forming the regression estimation equation for that city or town.

B. If several TSP stations in a city or town met the completeness criterion, all were combined and one <u>pooled</u> regression equation was assigned to all these stations. This occurred in Sacramento and San Jose.

C. If it was felt that a TSP station without paired data would be most closely represented by a nearby typical area regression set, the TSP station was assigned the regression equation of that nearby regression set.

D. If a station was in an isolated area which lacked sufficient paired PM10/TSP data to form estimation, the overall statewide regression was used. If the station was in a region of California where area groupings had exhibited seasonal specific regression, the seasonal specific California regression equation was applied, using a summer season of months 3 to 9. In some cases, this season was adjusted to fit a seasonal pattern observed from available but incomplete paired PM10/TSP data. This occurred in two cases: Yosemite Valley, where a "summer season" of the months March through June was formed, and also for Lake Gregory (and, hence, Big Bear Lake), where a "summer season" of June through September was formed.

The rest of the stations which had insufficient paired data and were not part of a sequence were treated the same as stations which had no data.

A file was created which had each TSP station in the state listed and its coefficient for each calendar month and an indication of the source of the coefficient. A copy of this file was sent on diskette to CARB. Table 8B shows the assignment of TSP stations to regression sets.

Final regression coefficients were checked using the split-halves technique. A 50% random sample of each regression set was taken and regression coefficients through the origin of PM10 on TSP formed on one of the halves. These coefficients were used to predict the values of PM10 on the other half of the regression set based on the observed values of TSP. These predicted values were paired with their corresponding actual values. These pairings of predicted and observed values for the second half of each regression set were aggregated into one data set and one overall correlation was run. The overall r was 0.93, and r^2 was 0.87.

Staff at the CARB used the resultant regression estimation equations and every sixth-day data value of TSP to form indirect regression estimates of mean concentration of PM10 for every sixth day. Cumulations of indirectly estimated mean concentrations of PM10 as well as exceedance frequency and excess concentration statistics were made on a monthly basis and these were interpolated to zip code centroids using the previously outlined methods for TSP. When forming monthly cumulations, the average monthly value was imputed for missing values according to rules used for TSP cumulations (see Abbey, 1991). The same quality ratings as for ozone were erroneously assigned by CARB to zip code by month interpolations. Interpolations were made for the years 1973 through March, 1987. The resultant indirectly estimated PM10 values by zip code and months were sent to Loma Linda. After initial screening for error detection and correction of errors, cumulations of PM10 for study participants were made.

§ 3. Precision of PM10 Estimation Methods

The precision of the indirect estimation and interpolation methods for PM10 were assessed by regarding fixed-site stations in turn as receptor sites and interpolating estimated PM10 values from surrounding stations. PM10 was first indirectly estimated at the surrounding stations using the regression equations described above and the value of TSP recorded at the respective stations. This comparison was made for 39 stations with at least 20 months of paired data (a total of 841 months) over the two year period of 1986 and 1987 when both PM10 and TSP were widely monitored. Excess concentration or exceedance frequencies for estimated or actual PM10 were first cumulated over the two year period and then paired t-tests and correlation coefficients were calculated between estimated and actual values. Regression equations of actual on estimated values were also formed to assess possible bias.

Comparisons between indirectly estimated and actual values are shown in Table 9. The combined effects of indirect estimation of PM10 and interpolation of indirectly estimated PM10 are assessed by these comparisons. Although estimated and actual values were highly correlated ($r = 0.86$ for mean concentration) indirectly estimated PM10 was consistently and statistically significantly lower than actual PM10 except for the cutoff of 40 mcg/m^3. The error was small (less than 10%) for mean concentration and became progressively larger for higher cutoffs, though the correlations remained high for all but the highest cutoff. These results would suggest caution in quantitative interpretation of observed health effects associations.

To assess how much of the above error was due to error in interpolation, actual monitored PM10 was interpolated from surrounding stations instead of estimated PM10 and two year cumulations compared for the same 39 stations. Table 10 shows the results. In contrast to the comparison of indirectly estimated and interpolated PM10 with actual PM10 we find that interpolated actual PM10 is not significantly different from the actual values. For exceedance frequencies interpolated actual values still tended to underestimate actual values. However, for excess concentrations interpolated values tended to over estimate actual values slightly. Correlations are high but not much higher than the correlations between interpolated indirect estimates and actual values, thus there does not appear to be a great loss in precision using the indirect regression estimated PM10.

Tables 11 and 12 compare monthly average values rather than annual averages over the two year period for the same set of stations.

§ 4. PM10 Data Which Failed to Meet EPA Quality Criteria

Not all California PM10 measurements prior to November 1986 satisfied EPA criteria for control of temperature and humidity during the weighing of the filters. For this reason, all PM10 measurements before that date were deleted from the CARB's air quality data base after we had included them in the data used to compute regressions for estimating PM10 from TSP. Laboratory studies by CARB research personnel suggested that the lack of adherence to temperature control criteria would likely have negligible effects of the PM10 measurements.

These studies suggested that lack of adherence to humidity control criteria would change PM10 measurements by less than 5% and that the absolute changes in filter weights would be less than the EPA's absolute precision requirement of 5 mg.

To access the effects of including the suspect data prior to November 1986 in the regressions used to estimate PM10 from TSP, we excluded these data (5,883 of the 14,314 data points) and recompute the regressions. All the data for two of the 130 site and season-specific regressions were excluded. Also, for 27 of the regressions, there were no longer sufficient data to satisfy our completeness criterion of at least 20 months of data with two or more observations per month.

The absolute percentage deviations of the corresponding regression coefficients were generally small. For the 101 regressions whose data satisfied the completeness criterion, the median, mean, and maximum of these deviations were 1.3%, 2.1%, and 9.7%, respectively. The mean signed deviation was 0.2%. For the entire set of regressions, the median, mean, and maximum of absolute deviations were 1.5%, 2.6%, and 14.6%, respectively. The mean signed duration was 0.3%. These deviations were generally not large, compared to the errors of estimation and interpolation found by the assessment of validity reported in Tables 1 and 2 of Paper 15. We concluded that the paired data prior to November 1986 should be included in the estimations of PM10 from TSP to increase the representativeness of these estimations with respect to both year-to-year and geographic variation.

REFERENCES

Abbey, D.A., Moore, J., Petersen, F., Beeson, L. (1991b) Estimating cumulative ambient concentrations of air pollutants: description and precision of methods used for an epidemiological study. *Arch. Environ. Health* 1991; 46(5): 281-287.

Table 1. Comparison of PM10 Versus β r^2 and PM10 Versus TSP r^2

Code	Name	Months	PM10 with β r^2	n	PM10 with TSP r^2	n
L1 L2	Alameda NAS San Jose	Apr-Oct	.421	100	.787	96
L3 L4	Alameda NAS San Jose	Nov-Mar	.667	50	.819	47
M1	Bakersfield	All	.468	191	.506	172
N1	Fresno	Apr-Oct	.382	103	.815	92
N2	Fresno	Nov-Mar	.619	56	.797	51
O1 O2	LAX Long Beach	Mar-Sep	.471	179	.607	161
O3 O4	LAX Long Beach	Oct-Feb	.442	122	.696	116
P1	Ontario Central	Apr-Oct	.638	103	.653	99
P2	Ontario Central	Nov-Mar	.696	83	.760	76
Q1	Ontario East	All	.604	252	.891	245
R1	Ontario West	Dec-Apr	.558	87	.863	82
R2	Ontario West	May-Nov	.442	127	.823	108
S1	Sacramento	Apr-Oct	.309	135	.792	68
S2	Sacramento	Nov-Mar	.489	69	.783	27
T1	San Diego (Over-land/Chula Vista)	All	.449	66	.748	56
U1	San Diego (Over-land/Chula Vista)	All	.232	46	.796	42
V1	Stockton	All	.452	86	.545	80
Overall			.464	1855	.769	1618
			Mean = .491		Mean = .746	

PM10 versus TSP overall regression equation: PM10 = -2.048488 + .58370 x TSP

r^2 = .7688 n = 1618

Table 2A.
Table of Regression Coefficients of PM10 on TSP With Standard Errors and r^2 For Stations with PM10 and TSP Paired Data and at Least Two Observations Per Month Minus Three Points Considered Outliers (Regressions Forced Through Origin).

No.	CO	SITE	ß	SE ß	r^2	No. Pts	LOCATION
1	4	630	0.63	0.015	0.78	88	Chico-State
2	7	440	0.58	0.011	0.72	210	Concord-2975 Treat Blvd
3	7	442	0.57	0.010	0.80	214	Bethel Island Rd
4	10	234	0.55	0.009	0.69	320	Fresno-Olive
5	12	503	0.61	0.008	0.87	175	Eureka-H.D. 67th & I St.
6	13	682	0.52	0.012	0.66	113	El Centro-Broadway
7	14	697	0.39	0.012	0.93	60	Keeler
8	14	699	0.62	0.010	0.90	207	Lone Pine-501 E. Locust
9	15	203	0.51	0.009	0.62	333	Bkrsfld-Chester St
10	15	213	0.50	0.008	0.86	201	Taft-North 10th St.
11	15	243	0.58	0.009	0.74	253	Oildale-3311 Manor
12	16	701	0.46	0.012	0.72	177	Hanford
13	16	715	0.47	0.009	0.85	164	Corcoran-Van Dorsten Ave
14	21	451	0.59	0.011	0.64	194	San Rafael
15	24	521	0.54	0.011	0.69	166	Merced
16	26	776	0.57	0.014	0.91	87	Mammoth Lakes -Fire Sta
17	26	785	0.62	0.011	0.90	140	Mammoth Lakes-Gateway HC
18	27	544	0.50	0.007	0.76	204	Salinas II
19	28	783	0.61	0.010	0.83	208	Napa-Jefferson St
20	30	186	0.49	0.007	0.64	293	El Toro
21	30	190	0.50	0.008	0.64	271	Los Alamitos-Orangewood*
22	31	810	0.54	0.011	0.64	140	Rocklin-Sierra College
23	33	137	0.44	0.009	0.89	131	Palm Springs-Fire Sta
24	33	144	0.64	0.005	0.91	331	Riverside-Rubidoux
25	33	146	0.53	0.010	0.86	84	Riverside-Magnolia
26	33	149	0.58	0.007	0.93	117	Perris
27	33	150	0.55	0.007	0.87	287	Banning-Allesandro
28	33	157	0.56	0.006	0.92	288	Indio-Jackson
29	34	277	0.60	0.012	0.73	216	Sacto-H.D. Stockton Blvd
20	34	293	0.61	0.009	0.68	308	Citrus Hts-Sunrise Blvd
31	34	295	0.62	0.015	0.75	160	Sacramento-Del Paso Manor
32	35	823	0.51	0.015	0.58	60	Hollister2-1979 Fairview* 2
33	36	155	0.54	0.009	0.75	174	Barstow
34	36	171	0.62	0.007	0.84	285	Ontario Airport
35	36	188	0.42	0.008	0.66	237	Trona-Market St
36	36	197	0.54	0.008	0.75	302	Fontana-Arrow Hwy
37	36	199	0.47	0.010	0.66	129	Victorville-15579 8th St
38	36	203	0.58	0.008	0.89	193	San Bernardino-Fourth ST
39	39	252	0.54	0.010	0.67	252	Stockton-Hazelton St
40	40	833	0.50	0.009	0.86	83	Morrow Bay
41	41	541	0.54	0.009	0.77	214	Redwood City
42	42	356	0.42	0.006	0.72	204	Santa Maria-Library
43	42	363	0.51	0.008	0.81	153	Goleta
44	43	377	0.70	0.016	0.82	96	San Jose-Moorpark
45	43	382	0.53	0.007	0.83	304	San Jose-4th St

Continued on next page

Table 2A.
Table of Regression Coefficients of PM10 on TSP With Standard Errors and r^2 For Stations with PM10 and TSP Paired Data and at Least Two Observations Per Month Minus Three Points Considered Outliers (Regressions Forced Through Origin).

No.	CO	SITE	ß	SE ß	r^2	No. Pts	LOCATION
46	44	850	0.44	0.011	0.66	71	Santa Cruz-966 Bostwick
47	45	555	0.53	0.013	0.73	101	Redding-H.D. Roof
48	47	861	0.59	0.009	0.83	183	Yreka
49	49	885	0.53	0.019	0.62	55	Healdsburg
50	49	886	0.61	0.011	0.76	226	Cloverdale
51	49	898	0.58	0.010	0.69	234	Healdsburg 133 Matheson
52	50	558	0.59	0.012	0.64	272	Modesto-Oakdale Rd*
53	51	895	0.51	0.011	0.69	168	Yuba City-AG Bldg
54	52	901	0.54	0.010	0.76	147	Red Bluff
55	54	568	0.51	0.009	0.62	194	Visalia-Church St
56	56	419	0.53	0.011	0.68	101	El Rio-Rio Mesa School
57	56	427	0.59	0.009	0.91	127	Piru-2SW
58	56	434	0.55	0.006	0.85	256	Simi Valley 5400 Cochran
59	57	569	0.52	0.012	0.64	171	Woodland-W Main St
60	60	340	0.59	0.013	0.61	198	Livermoor-Old FST St
61	70	60	0.54	0.006	0.83	293	Azusa
62	70	69	0.62	0.008	0.74	283	Burbank
63	70	72	0.58	0.007	0.76	317	North Long Beach
64	70	82	0.55	0.012	0.77	54	Lancaster
65	70	87	0.59	0.007	0.79	259	Los Angeles-No Main
66	70	94	0.51	0.019	0.75	50	Hawthorne
67	80	114	0.57	0.008	0.77	153	Chula Vista
68	80	131	0.60	0.008	0.79	183	El Cajon-Redwood Ave
69	80	134	0.52	0.006	0.68	234	Oceanside-1701 Mission Ave.
70	90	306	0.61	0.011	0.76	199	San Francisco-10 Arkansas
--	--	---	0.54	0.001	0.81	14318	Statewide Paired Data**

* Minus an outlier
** Less five points statewide considered outliers.

Table 2B.
Table of Regression Coefficients of PM10 on TSP With Standard Errors and r^2 For Stations with PM10 and TSP Paired Data Having Less than Two Observations Per Month (Regressions Forced Through Origin).

No.	CO	SITE	ß	SE ß	r^2	No. Pts	LOCATION
71	7	433	0.68	0.022	0.81	45	Richmond-13th St
72	10	229	0.44	0.031	0.86	9	Five Points
73	10	241	0.52	0.027	0.62	15	Fresno-Cal State#2
74	13	693	0.47	0.013	0.62	94	Brawley-401 Main St
75	13	694	0.46	0.013	0.90	40	El Centro 50 9th St
76	15	207	0.48	0.019	0.66	46	Mojave
77	15	211	0.46	0.029	0.26	57	China Lake*
78	15	241	0.64	0.024	0.82	17	Oildale-Manor St
79	22	742	0.60	0.021	0.72	42	Yosemite Village*
80	23	753	0.58	0.016	0.84	53	Willits-Firehouse
81	23	764	0.50	0.026	0.89	5	Ukiah-County Library
82	31	813	0.52	0.013	0.78	73	Auburn-DeWitt-C Ave
83	36	181	0.57	0.016	0.73	50	Unnamed
84	36	190	0.37	0.040	0.63	9	Victorville-Fairgrounds
85	36	194	0.68	.		1	San Bernardino-E 3rd St
86	40	848	0.52	0.013	0.87	41	Nipomo-1230 Eucalyptus Rd
87	42	388	0.55	0.015	0.69	36	Santa Barbara-3 W Carillo
88	43	390	0.58	0.025	0.87	26	San Jose-W San Carlos St
89	50	567	0.63	0.028	0.73	49	Modesto-1100 I Street
90	56	413	0.52	0.021	0.88	19	Simi Valley
91	56	430	0.56	0.011	0.88	55	Ojai-1768 Maricopa Hwy
92	70	336	0.63	0.022	0.76	48	Fremont-Chapel Way
93	70	591	0.61	0.012	0.90	76	Glendora-Laurel
94	70	593	0.44	0.026	0.56	39	Lancaster-North Cedar Ave.
95	80	123	0.58	0.016	0.58	18	San Diego Overland

* Minus an outlier

Table 3.

Numbers of Observing Stations and Paired Observations for PM10 and TSP from 1982 to 1989 in the State of California October 20, 1992 for the 70 Stations with at Least Two Paired Observations From Each Calendar Month.

Year	Stations	Paired Observations
1982	2	6
1983	13	352
1984	28	637
1985	48	1705
1986	63	2882
1987	63	3057
1988	56	2657
1989	45	2216

Table 4A. Intercepts and Regression Coefficients of PM10 on TSP with Standard Errors and r^2 for Stations with PM10 and TSP Paired Data and at Least Two Observations Per Month Minus Three Points Considered Outliers

No	CO/SITE		α	SE α	sig.	ß	SE ß	r^2	No. Pts	LOCATION
1	4	630	-0.2	2.09	0.9289	0.63	0.036	0.78	88	CHICO-STATE
2	7	440	-2.0	1.45	0.1631	0.61	0.027	0.72	210	CONCORD-2975 TREAT BLVD
3	7	442	-4.0	1.29	0.0024	0.63	0.022	0.80	214	BETHEL ISLAND RD
4	10	234	-9.3	2.78	0.0009	0.62	0.024	0.69	320	FRESNO-OLIVE
5	12	503	-2.2	1.19	0.0619	0.64	0.019	0.87	175	EUREKA-H.D. 6TH & I ST
6	13	682	9.1	3.62	0.0135	0.45	0.031	0.66	113	EL CENTRO-BROADWAY
7	14	697	2.5	2.32	0.2919	0.38	0.013	0.93	60	KEELER
8	14	699	0.1	0.91	0.9360	0.62	0.014	0.90	207	LONE PINE-501 E LOCUST
9	15	203	10.3	2.66	0.0001	0.44	0.019	0.62	333	BKRSFLD-CHESTER ST
10	15	213	8.8	1.71	0.0001	0.45	0.013	0.86	201	TAFT-NORTH 10TH ST
11	15	243	-1.4	2.85	0.6288	0.59	0.022	0.74	253	OILDALE-3311 MANOR
12	16	701	3.4	2.71	0.2117	0.44	0.021	0.72	177	HANFORD
13	16	715	-0.5	2.43	0.8272	0.47	0.015	0.85	164	CORCORAN-VAN DORSTEN AVE
14	21	451	2.0	1.48	0.1824	0.55	0.030	0.64	194	SAN RAFAEL
15	24	521	3.2	2.39	0.1793	0.51	0.027	0.69	166	MERCED
16	26	776	5.2	3.16	0.1012	0.55	0.019	0.91	87	MAMMOTH LAKES-FIRE STA
17	26	785	-5.3	1.61	0.0014	0.67	0.019	0.90	140	MAMMOTH LAKES-GATEWAY HC
18	27	544	2.4	0.92	0.0097	0.46	0.018	0.76	204	SALINAS II
19	28	783	-6.6	1.34	0.0001	0.71	0.023	0.83	208	NAPA-JEFFERSON ST
20	30	186	7.5	1.56	0.0001	0.41	0.018	0.64	293	EL TORO
21	30	190	2.5	2.26	0.2668	0.48	0.022	0.64	271	LOS ALAMITOS-ORANGEWOOD*
22	31	810	3.1	1.83	0.0946	0.49	0.031	0.64	140	ROCKLIN-SIERRA COLLEGE
23	33	137	4.6	1.34	0.0007	0.41	0.013	0.89	131	PALM SPRINGS-FIRE STA
24	33	144	-8.3	1.82	0.0001	0.69	0.012	0.91	331	RIVERSIDE-RUBIDOUX
25	33	146	-0.6	2.83	0.8421	0.54	0.024	0.86	84	RIVERSIDE-MAGNOLIA
26	33	149	-3.0	1.71	0.0799	0.60	0.015	0.93	117	PERRIS
27	33	150	-2.8	1.28	0.0317	0.57	0.013	0.87	287	BANNING-ALLESANDRO
28	33	157	1.8	1.41	0.1950	0.45	0.008	0.92	288	INDIO-JACKSON
29	34	277	-3.2	2.00	0.1119	0.64	0.027	0.73	216	SACTO-H.D. STOCKTON BLVD
30	34	293	-1.4	1.75	0.4077	0.63	0.025	0.68	308	CITRUS HTS-SUNRISE BLVD
31	34	295	-6.8	2.22	0.0024	0.71	0.033	0.75	160	SACRAMENTO-DEL PASO MANOR
32	35	823	6.4	1.97	0.0019	0.38	0.043	0.58	60	HOLLISTER-1979 FAIRVIEW*
33	36	155	2.0	1.60	0.2160	0.51	0.023	0.75	174	BARSTOW
34	36	171	-4.8	2.19	0.0299	0.65	0.017	0.84	285	ONTARIO AIRPORT
35	36	188	8.0	2.03	0.0001	0.36	0.017	0.66	237	TRONA-MARKET ST
36	36	197	4.7	2.56	0.0679	0.52	0.017	0.75	302	FONTANA-ARROW HWY
37	36	199	7.7	2.54	0.0029	0.40	0.025	0.66	129	VICTORVILLE-15579 8TH ST
38	36	203	-7.5	2.44	0.0024	0.63	0.016	0.89	193	SAN BERNARDINO-FOURTH ST
39	39	252	-0.5	2.45	0.8441	0.54	0.024	0.67	252	STOCKTON-HAZELTON ST
40	40	833	-1.3	1.19	0.2884	0.53	0.023	0.86	83	MORRO BAY
41	41	541	-2.0	1.32	0.1251	0.57	0.022	0.77	214	REDWOOD CITY
42	42	356	8.5	1.18	0.0001	0.33	0.014	0.72	204	SANTA MARIA-LIBRARY
43	42	363	6.9	0.92	0.0001	0.40	0.016	0.81	153	GOLETA
44	43	377	-7.2	2.13	0.0010	0.82	0.039	0.82	96	SAN JOSE-MOORPARK

Table 4A. (continued...)

No	CO/SITE		α	SE α	sig.	β	SE β	r^2	No. Pts	LOCATION
45	43	382	-11.7	1.63	0.0001	0.64	0.017	0.83	304	SAN JOSE-4TH ST
46	44	850	7.3	1.93	0.0003	0.33	0.029	0.66	71	SANTA CRUZ-966 BOSTWICK
47	45	555	0.6	1.81	0.7216	0.52	0.032	0.73	101	REDDING-H.D. ROOF
48	47	861	-3.1	1.00	0.0024	0.66	0.022	0.83	183	YREKA
49	49	885	-2.4	2.95	0.4210	0.58	0.062	0.62	55	HEALDSBURG
50	49	886	-2.4	1.25	0.0596	0.65	0.025	0.76	226	CLOVERDALE
51	49	898	0.2	1.20	0.8538	0.57	0.025	0.69	234	HEALDSBURG-133 MATHESON
52	50	558	-4.9	2.54	0.0526	0.64	0.029	0.64	272	MODESTO-OAKDALE RD*
53	51	895	5.2	1.74	0.0035	0.45	0.023	0.69	168	YUBA CITY-AG BLDG
54	52	901	3.8	1.69	0.0248	0.49	0.023	0.76	147	RED BLUFF
55	54	568	12.6	2.40	0.0001	0.43	0.019	0.72	194	VISALIA-CHURCH ST
56	56	419	5.8	2.41	0.0171	0.46	0.031	0.68	101	EL RIO-RIO MESA SCHOOL
57	56	427	0.0	1.11	0.9723	0.59	0.017	0.91	127	PIRU-2SW
58	56	434	-1.6	1.22	0.1843	0.57	0.015	0.85	256	SIMI VALLEY-5400 COCHRAN
59	57	569	2.3	2.09	0.2731	0.50	0.029	0.64	171	WOODLAND-W MAIN ST
60	60	340	5.0	1.75	0.0049	0.52	0.030	0.61	198	LIVERMORE-OLD FST ST
61	70	60	2.5	1.79	0.1617	0.52	0.014	0.83	293	AZUSA
62	70	69	-2.4	2.45	0.3325	0.64	0.023	0.74	283	BURBANK
63	70	72	1.0	1.73	0.5813	0.57	0.018	0.76	317	NORTH LONG BEACH
64	70	82	2.1	3.51	0.5597	0.53	0.041	0.77	54	LANCASTER
65	70	87	-2.0	2.14	0.3518	0.61	0.019	0.79	259	LOS ANGELES-NO MAIN
66	70	94	16.3	3.07	0.0001	0.37	0.030	0.75	50	HAWTHORNE
67	80	114	-0.1	1.57	0.9411	0.57	0.026	0.77	153	CHULA VISTA
68	80	131	-2.1	1.73	0.2329	0.63	0.024	0.79	183	EL CAJON-REDWOOD AVE
69	80	134	4.0	1.40	0.0051	0.47	0.021	0.68	234	OCEANSIDE-1701 MISSION AV
70	90	306	-1.9	1.39	0.1755	0.64	0.026	0.76	199	SAN FRANCISCO-10 ARKANSAS
--	--	---				0.54	0.001	0.81	14318	STATEWIDE PAIRED DATA**

* Minus an outlier.
** Less five points statewide considered outliers.

Table 4B. Intercepts and Regression Coefficients of PM10 on TSP With Standard Errors and r^2 for Stations with PM10 and TSP Paired Data and at Least One Month with Less Than Two Observations Minus Two Points Considered Outliers

No	CO/SITE		α	SE α	sig.	B	SE B	r^2	No. Pts	LOCATION
71	7	433	-6.7	2.73	0.0181	0.81	0.060	0.81	45	RICHMOND-13TH ST
72	10	229	11.7	8.14	0.1944	0.37	0.057	0.86	9	FIVE POINTS
73	10	241	6.1	9.01	0.5117	0.45	0.100	0.62	15	FRESNO-CAL STATE#2
74	13	693	1.5	4.63	0.7455	0.46	0.038	0.62	94	BRAWLEY-401 MAIN ST
75	13	694	11.8	3.42	0.0014	0.40	0.021	0.90	40	EL CENTRO-150 9TH ST
76	15	207	-1.6	7.77	0.8354	0.49	0.053	0.66	46	MOJAVE
77	15	211	14.1	2.03	0.0001	0.20	0.045	0.26	57	CHINA LAKE*
78	15	241	-13.6	12.73	0.3035	0.73	0.089	0.82	17	OILDALE-MANOR ST
79	22	742	-1.7	4.25	0.6898	0.62	0.062	0.72	42	YOSEMITE VILLAGE*
80	23	753	1.7	2.73	0.5454	0.56	0.034	0.84	53	WILLITS-FIREHOUSE
81	23	764	-1.4	5.81	0.8217	0.53	0.106	0.89	5	UKIAH-COUNTY LIBRARY
82	31	813	1.4	2.23	0.5439	0.50	0.031	0.78	73	AUBURN-DEWITT-C AVE
83	36	181	2.3	3.56	0.5299	0.54	0.048	0.73	50	LAKE GREGORY
84	36	190	22.0	8.90	0.0426	0.23	0.065	0.63	9	VICTORVILLE-FAIRGROUNDS
85	36	194	105.0	.	.	.	.	.	1	SAN BERNARDINO-E 3RD ST
86	40	848	-2.3	2.45	0.3476	0.00	0.034	0.87	41	NIPOMO-1230 EUCALYPTUS RD
87	42	388	10.9	4.08	0.0112	0.43	0.049	0.69	36	SANTA BARBARA-3 W CARILLO
88	43	390	-16.1	4.85	0.0029	0.77	0.061	0.87	26	SAN JOSE-W SAN CARLOS ST
89	50	567	-16.1	6.82	0.0223	0.79	0.070	0.73	49	MODESTO-1100 I STREET
90	56	413	-15.4	6.00	0.0198	0.66	0.060	0.88	19	SIMI VALLEY
91	56	430	1.9	1.72	0.2841	0.53	0.028	0.88	55	OJAI-1768 MARICOPA HWY
92	60	336	-10.7	3.73	0.0063	0.80	0.066	0.76	48	FREMONT-CHAPEL WAY
93	70	591	-5.5	2.22	0.0152	0.67	0.026	0.90	76	GLENDORA-LAUREL
94	70	593	0.8	4.67	0.8569	0.43	0.063	0.56	39	LANCASTER-NORTH CEDAR AVE
95	80	123	9.2	4.77	0.0718	0.42	0.088	0.58	18	SAN DIEGO-OVERLAND

* Minus an outlier.

Table 5. Tentative (first) Grouping of All Stations With Paired PM10 and TSP Data (1982-1989) For Regional Regression Equation Calculations.

No	CO	Site	β	SEβ	r^2	No.Pts	Location
Group 1							
67	80	114	0.57	0.008	0.77	153	Chula Vista
68	80	131	0.60	0.008	0.79	183	El Cajon -Redwood Ave
95	80	123	0.58	0.016	0.58	18	San Diego-Overland
Group 2							
69	80	134	0.52	0.006	0.68	234	Oceanside-1701 Mission Ave.
Group 3							
6	13	682	0.52	0.012	0.66	113	El Centro-Broadway
74	13	693	0.47	0.013	0.62	94	Brawley-401 Main ST
75	13	694	0.46	0.013	0.90	40	El Centro-150 95th St
Group 4							
23	33	137	0.44	0.009	0.89	131	Palm Springs-Fire Sta.
28	33	157	0.46	0.006	0.92	288	Indio-Jackson
Group 5							
20	30	186	0.48	0.007	0.64	293	El Toro
21	30	190	0.50	0.008	0.64	271	Los Alamitos-Orangewood*
Group 6							
24	33	144	0.64	0.005	0.91	331	Riverside-Rubidoux
25	33	146	0.53	0.010	0.86	84	Riverside-Magnolia
26	33	149	0.58	0.007	0.93	117	Perris
27	33	150	0.55	0.007	0.87	287	Banning-Allesandro
34	36	171	0.62	0.007	0.84	285	Ontario Airport
36	36	197	0.54	0.008	0.75	302	Fontana-Arrow Hwy
38	36	203	0.58	0.008	0.89	193	San Bernardino-Fourth St
61	70	60	0.54	0.006	0.83	293	Azusa
85	36	194	0.68	..	..	1	San Bernardino-E 3rd St
93	70	591	0.61	0.012	0.90	76	Glendora-Laurel
Group 7							
83	36	181	0.57	0.016	0.73	50	Lake Gregory
Group 8							
62	70	69	0.62	0.008	0.74	283	Burbank
63	70	72	0.58	0.007	0.76	317	North Long Beach
65	70	87	0.59	0.007	0.79	259	Los Angeles-No Main
66	70	94	0.51	0.019	0.75	50	Hawthorne
Group 9							
58	56	434	0.55	0.006	0.85	256	Simi Valley-5400 Cochran
90	56	413	0.52	0.021	0.88	19	Simi Valley
Group 10							
40	40	833	0.50	0.009	0.86	83	Morro Bay
43	42	363	0.51	0.008	0.81	153	Goleta
56	56	419	0.53	0.011	0.68	101	El Rio-Rio Mesa School
87	42	388	0.55	0.015	0.69	36	Santa Barbara-3 W Carillo
Group 11							
57	56	427	0.59	0.009	0.91	127	Piru-2SW
91	56	430	0.56	0.011	0.88	55	Ojai-1768 Maricopa Hwy

Table 5. (continued...) Appendix E

Group 12							
37	36	199	0.47	0.010	0.66	129	Victorville-15579 8th St.
76	15	207	0.48	0.019	0.66	46	Mojave
84	36	190	0.37	0.040	0.63	9	Victorville-Fairgrounds
Group 13							
33	36	155	0.54	0.009	0.75	174	Barstow
64	70	82	0.55	0.012	0.77	54	Lancaster
94	70	593	0.44	0.026	0.56	39	Lancaster-North Cedar Ave
Group 14							
35	36	188	0.42	0.008	0.66	237	Trona-Market St
Group 15							
77	15	211	0.47	0.029	0.26	57	China Lake*
Group 16							
42	42	356	0.42	0.006	0.72	204	Santa Maria-Library
86	40	848	0.52	0.013	0.87	41	Nipomo-1230 Eucalyptus Rd
Group 17							
18	27	544	0.50	0.007	0.76	204	Salinas II
32	35	823	0.51	0.015	0.58	60	Hollister-1979 Fairview*
Group 18							
46	44	850	0.44	0.011	0.66	71	Santa Cruz-966 Bostwick
Group 19							
4	10	234	0.55	0.009	0.69	320	Fresno-Olive
9	15	203	0.51	0.009	0.62	333	Bkrsfld-Chester St
11	15	243	0.58	0.009	0.74	253	Oildale-3311 Manor
55	54	568	0.51	0.009	0.72	194	Visalia-Church St
73	10	241	0.52	0.027	0.62	15	Fresno-Cal State #2
78	15	241	0.64	0.024	0.82	17	Oildale-Manor St
Group 20							
10	15	213	0.50	0.008	0.86	201	Taft-North 10th St
12	16	701	0.46	0.012	0.72	177	Hanford
13	16	715	0.47	0.009	0.85	164	Corcoran-Van Dorsten Ave
72	10	229	0.44	0.031	0.86	9	Five Points
Group 21							
7	14	697	0.39	0.012	0.93	60	Keeler
Group 22							
8	14	699	0.62	0.010	0.90	207	Lone Pine-501 E Locust
Group 23							
52	50	558	0.59	0.012	0.64	272	Modesto-Oakdale Rd*
89	50	567	0.63	0.028	0.73	49	Modesto-1100 I Street
Group 24							
15	24	521	0.54	0.011	0.69	166	Merced
Group 25							
79	22	742	0.60	0.021	0.72	42	Yosemite Village*
Group 26							
39	39	252	0.54	0.010	0.67	252	Stockton-Hazelton St

Table 5. (continued...)							Appendix E
Group 27							
2	7	440	0.58	0.011	0.72	210	Concord-2975 Treat Blvd
19	28	783	0.61	0.010	0.83	208	Napa-Jefferson St
60	60	340	0.59	0.013	0.61	198	Livermore-Old Fst ST
Group 28							
41	41	541	0.54	0.009	0.77	214	Redwood City
44	43	377	0.70	0.016	0.82	96	San Jose-Moorepark
45	43	382	0.53	0.007	0.83	304	San Jose-4th ST
71	7	433	0.68	0.022	0.81	45	Richmond-13th St
88	43	390	0.58	0.025	0.87	26	San Jose-W San Carlos St
92	60	336	0.63	0.022	0.76	48	Fremont-Chapel Way
Group 29							
3	7	442	0.57	0.010	0.80	214	Bethel Island Rd
Group 30							
70	90	306	0.61	0.011	0.76	199	San Francisco-10 Arkansas
Group 31							
14	21	451	0.59	0.011	0.64	194	San Rafael
Group 32							
22	31	810	0.54	0.011	0.64	140	Rocklin-Sierra College
29	34	277	0.60	0.012	0.73	216	Sacto-H.D. Stockton Blvd
30	34	293	0.61	0.009	0.68	308	Citrus Hts-Sunrise Blvd
31	34	295	0.62	0.015	0.75	160	Sacramento-Del Paso Manor
82	31	813	0.52	0.013	0.78	73	Auburn-Dewitt-C Ave
Group 33							
1	4	630	0.63	0.015	0.78	88	Chico-State
53	51	895	0.51	0.011	0.69	168	Yuba City-AG Bldg
59	57	569	0.52	0.012	0.64	171	Woodland-W Main St
Group 34							
49	49	885	0.53	0.019	0.62	55	Healdsburg
50	49	886	0.61	0.011	0.76	226	Cloverdale
51	49	898	0.58	0.010	0.69	234	Healdsburg-133 Matheson
80	23	753	0.58	0.016	0.84	53	Willits-Firehouse
81	23	764	0.50	0.026	0.89	5	Ukiah-County Library
Group 35							
47	45	555	0.53	0.013	0.73	101	Redding-H.D. Roof
54	52	901	0.54	0.010	0.76	147	Red Bluff
Group 36							
48	47	861	0.59	0.009	0.83	183	Yreka
Group 37							
16	26	776	0.57	0.014	0.91	87	Mammoth Lakes-Fire Sta
17	26	785	0.62	0.011	0.90	140	Mammoth Lakes-Gateway HC
Group 38							
5	12	503	0.61	0.008	0.87	175	Eureka-H.D. 6th & I St

Table 6. **PM10 on TSP (1982-89 Data)** Seasonal F-Tests for Groups Which Seemed to have Seasons.

Group	MSE - Winter	WI - DF	MSE - Summer	SU - DF	2-Tailed F-Value	MSE - Combined	CO - DF	Season F-Test	DF1/DF2	
1	0.00520	162	0.00366	170	1.42	0.00501	334	23.41	2	332
2	no season									
3	no season									
4	no season									
5	no season									
6	no season									
7	no season									
8	no season									
9	0.00727	125	0.00329	127	2.21	0.00576	254	13.08	2	252
10	0.00665	168	0.00510	164	1.30	0.00647	334	17.67	2	332
11	0.01092	88	0.00406	35	2.69	0.00902	125	1.35*	2	123
12	no season									
13	no season									
14	no season									
15	no season									
16	0.00623	75	0.00764	125	1.23	0.00728	202	3.30	2	200
17	0.00764	137	0.00485	123	1.57	0.00782	262	32.07	2	260
18	0.00320	29	0.00374	37	1.17	0.00658	68	30.89	2	66
19	0.01390	543	0.00884	553	1.44	0.01397	1098	132.24	2	1096
20	0.01449	208	0.01620	330	1.15	0.01813	540	32.74	2	538
21										
22										
23	0.01441	115	0.01105	154	1.30	0.02186	271	102.76	2	269
24	0.01201	84	0.00533	78	2.25	0.01215	164	32.31	2	162
25										
26	0.01146	106	0.00816	142	1.40	0.01455	250	65.98	2	248
27	0.00851	126	0.01012	486	1.19	0.01180	614	64.03	2	612
28	0.01409	242	0.00685	368		0.01117	612	46.48	2	610
29	0.01694	88	0.00482	122	3.52	0.01177	212	21.08	2	210
30	0.01955	104	0.00660	91	2.96	0.01450	197	8.28	2	195
31	0.01117	87	0.00664	103	1.68	0.01068	192	22.70	2	190
32	0.00945	277	0.00812	403	1.16	0.01398	682	210.08	2	680
33	0.01363	188	0.00865	234	1.59	0.01290	424	40.68	2	422
34	0.00746	239	0.00950	272	1.27	0.01419	513	170.11	2	511
35	0.02038	134	0.00865	110	2.36	0.01650	246	12.45	2	244
36	0.00703	65	0.00488	113	1.44	0.00752	180	30.45	2	178
37	0.01022	52	0.00972	171	1.05	0.01060	225	9.77	2	223
38	no season									

* Not significant at 0.05 level, the rest are significant.

TABLE 7A. PM10 on TSP (1982-1989 Data) General F-Tests for Combining Groups which are Similar in Season, Regression Coefficients, and Location.

Season	Group 1	Group 1 DF	Group 1 MSE	Group 2	Group 2 DF	Group 2 MSE	Two-tailed F	Combined DF	Combined MSE	Combined F	DF1	DF2
0	2	232	0.00542	5	562	0.01091	2.01	796	0.00932	1.55	2	794
0	2	232	0.00542	16	202	0.00728	1.34	436	0.00725	34.6	2	434
0	3	111	0.01515	4	417	0.01317	1.15	530	0.01434	15.6	2	528
0	4	417	0.01317	6	1889	0.00936	1.41	2308	0.01096	104.7	2	2306
0	6	1889	0.00936	8	907	0.00746	1.26	2798	0.00889	23.9	2	2796
0	8	907	0.00746	11	125	0.00902	1.21	1034	0.00773	6.61	2	1032
0	11	125	0.00902	16	202	0.00728	1.24	329	0.00957	34.8	2	327
0	12	127	0.01221	13	226	0.01576	1.29	355	0.01540	12.3	2	353
0	12	127	0.01221	14	235	0.01831	1.50	364	0.01672	7.18	2	362
0	13	226	0.01576	14	235	0.01831	1.16	463	0.01928	31.2	2	461
1	1	162	0.00520	9	125	0.00727	1.40	289	0.00642	8.57	2	287
2	1	170	0.00366	9	127	0.00329	1.11	299	0.00366	7.66	2	297
1	9	125	0.00727	17	137	0.00764	1.05	264	0.00800	10.5	2	262
2	9	127	0.00329	17	123	0.00485	1.48	252	0.00444	13.0	2	250
1	17	137	0.00764	19	543	0.01390	1.82	682	0.01280	5.17	2	680
2	17	123	0.00485	19	553	0.00884	1.82	678	0.00812	1.23	2	676
1	17	137	0.00764	33	188	0.01363	1.78	327	0.01167	8.32	2	325
2	17	123	0.00485	33	234	0.00865	1.78	359	0.00751	5.06	2	357
1	33	188	0.01363	35	134	0.02038	1.50	324	0.01664	2.97	2	322
2	33	234	0.00865	35	110	0.00865	1.00	346	0.00864	0.72	2	344
1	16	75	0.00623	20	208	0.01449	2.32	285	0.01326	12.1	2	283
2	16	125	0.00764	20	330	0.01620	2.12	457	0.01394	3.01	2	455
1	16	75	0.00623	23	115	0.01441	2.31	192	0.01754	55.6	2	190

TABLE 7A. (continued) PM10 on TSP (1982-1989 Data) General F-Tests for Combining Groups which are Similar in Season, Regression Coefficients, and Location.

Season	Group 1	Group 1 DF	Group 1 MSE	Group 2	Group 2 DF	Group 2 MSE	Two-tailed F	Combined DF	Combined MSE	Combined F	DF1	DF2
1	16	75	0.00623	28	242	0.01409	2.26	319	0.01338	15.9	2	317
2	16	125	0.00764	28	368	0.00685	1.12	495	0.00728	8.99	2	493
1	16	75	0.00623	32	277	0.00945	1.52	354	0.01343	95.1	2	352
2	16	125	0.00764	32	403	0.00812	1.06	530	0.00846	15.9	2	528
1	16	75	0.00623	34	239	0.00746	1.20	316	0.01030	70.1	2	314
2	16	125	0.00764	34	272	0.00950	1.24	399	0.00910	5.09	2	397
1	20	208	0.01449	23	115	0.01014	1.43	325	0.01686	50.3	2	323
2	20	330	0.01620	23	154	0.01105	1.47	486	0.01453	0.49	2	484
1	20	208	0.01449	26	106	0.01146	1.26	316	0.01448	12.9	2	314
2	20	330	0.01620	26	142	0.00816	1.99	474	0.01376	0.61	2	472
1	23	115	0.01441	26	106	0.01146	1.26	223	0.01327	3.39	2	221
2	23	154	0.01105	26	142	0.00816	1.35	298	0.00972	1.92	2	296
1	26	106	0.01146	28	242	0.01409	1.23	350	0.01389	8.86	2	348
2	26	142	0.00816	28	368	0.00685	1.19	512	0.00722	1.02	2	510
1	26	106	0.01146	32	277	0.00945	1.21	385	0.01037	7.87	2	383
2	26	142	0.00816	32	403	0.00812	1.00	547	0.00819	3.06	2	545
1	28	242	0.01409	32	277	0.00945	1.49	521	0.01379	49.7	2	519
2	28	368	0.00685	32	403	0.00812	1.19	773	0.00755	2.73	2	771
1	28	242	0.01409	34	239	0.00746	1.89	483	0.01197	27.2	2	481
2	28	368	0.00685	34	272	0.00950	1.39	642	0.00815	7.85	2	640
1	32	277	0.00945	34	239	0.00746	1.27	518	0.00858	2.38	2	516
2	32	403	0.00812	34	272	0.00950	1.17	677	0.00899	13.0	2	675
1	24	84	0.01201	31	87	0.01117	1.08	173	0.01146	0.09	2	171
2	24	78	0.00533	31	103	0.06407	1.24	183	0.00605	0.66	2	181

Ta 7B

Results of Several F-Tests for Group Poolings by Season of Regression Equations of PM10 on TSP

Groups[1] with No Season	Groups with Season 2 of Months 3-9
2,5* 3,4 4,6 6,8 8,111 11,16 12,13 12,14 13,14 21 22 38	16,20* 16,23* 16,26 16,27 16,32 16,34 20,23* 20,26* 23,26* 26,28* 26,32* 28,32* 28,34 32,34*
Groups with Season 2 of Months 4-9	
1,9 9,17 17,19 17,33 29 33,35	
Groups with Season 2 of Months 4-11 18	Groups with Season 2 of Months 30
Groups with Season 2 of Months 24,31	Groups with Season 2 of Months 36

* General F-Test indicated pooling could be done but pooling was not done.

[1] Group numbers refer to tentative groups of Table 5.

Table 8A. Appendix E List of the Final Regression Sets (84) - Based on 1982-1989 PM10 and TSP paired Data Whose Coefficients were used to Estimate PM10 from TSP Data Monitoring Stations.

June 25, 1991

SET[a]	IDX[b]	MO[c]	SN[d]	β*	SEβ*	r²*	PTS	LOCATION
1-0	1G		1	6275	82	8333	164	CHULA VISTA
	1G	4-9	2	5446	60	7788	172	CHULA VISTA
1-1	67I		1	6133	116	8483	72	CHULA VISTA
	67I	4-9	2	5294	90	7419	81	CHULA VISTA
1-2	68I		1	6352	113	8115	92	EL CAJON
	68I	4-9	2	5549	79	7982	91	EL CAJON
2	69I		0	5218	64	6804	234	OCEANSIDE
3	3C		0	4962	96	7624	153	EL CENTRO
4-0	4G		0	4589	47	9223	419	PALM SPRINGS
4-1	23I		0	4391	91	8903	131	PALM SPRINGS
4-2	28I		0	4619	55	9238	288	INDIO
5-0	5G		0	4953	52	6570	614	EL TORO
5-1	20I		0	4901	73	6443	293	EL TORO
5-2	21I		0	4974	80	6426	271	LOS ALAMITOS
5-3	66I		0	5063	185	7514	50	HAWTHORNE
6-1	24I		0	6447	52	9131	331	RUBIDOUX
6-2	25I		0	5347	96	8620	84	MAGNOLIA
6-3	26I		0	5780	67	9349	117	PERRIS
6-4	27I		0	5473	70	8666	287	BANNING
6-5	52C		0	5847	76	8880	194	SAN BERNARDINO
6-6	34I		0	6177	70	8415	285	ONTARIO
6-7	50P		0	5495	47	8117	594	FONTANA-AZUSA
6-8	36I		0	5577	72	8012	301	FONTANA
6-9	61I		0	5383	58	8262	293	AZUSA
7-1	62I		0	6230	77	7432	283	BURBANK
7-2	8P		0	5855	47	8848	576	LA/NLB
7-3	63I		0	5822	66	7626	317	N LONG BEACH
7-4	65I		0	5885	67	7910	259	LOS ANGELES
8	9C		1	6015	96	9041	134	SIMI VALLEY
	9C	4-9	2	5151	54	8317	129	SIMI VALLEY
9-0	10G		1	5694	77	7770	194	MORRO BAY
	10G	4-8	2	4928	58	8614	178	MORRO BAY

* These values are times 10^{-4}.

Table 8A (continued)

SET[a]	IDX[b]	MO[c]	SN[d]	β*	SEβ*	r^2*	PTS	LOCATION
9-1	53C		1	5671	102	7369	106	GOLETA-SB
	53C	4-8	2	5114	81	8663	82	GOLETA-SB
9-2	40I		1	5149	162	7893	35	MORRO BAY
	40I	4-8	2	4955	98	9167	48	MORRO BAY
9-3	56I		1	5926	150	7844	53	EL RIO
	56I	4-8	2	4728	117	7774	48	EL RIO
10	57I		0	5738	96	8923	127	PIRU
11	12C		0	4573	98	6308	138	VICTORVILLE
12-0	13G		0	5293	77	7211	267	BARSTOW
12-1	54C		0	5147	136	6720	93	LANCASTER
12-2	33I		0	5396	92	7522	174	BARSTOW
13	35I		0	4166	84	6595	237	TRONA
14	42I		1	4520	100	7939	77	SANTA MARIA
	42I	3-9	2	4061	80	6921	127	SANTA MARIA
15-0	17G		1	5487	92	7200	139	SALINAS
	17G	4-9	2	4614	60	8569	125	SALINAS
15-1	18I		1	5406	101	7268	109	SALINAS
	18I	4-9	2	4645	66	8807	95	SALINAS
15-2	32I		1	5873	214	7189	30	HOLLISTER
	32I	4-9	2	4501	135	7003	30	HOLLISTER
16	46I		1	4869	153	7443	31	SANTA CRUZ
	46I	4-11	2	3900	91	8641	39	SANTA CRUZ
17-1	4I		1	6164	130	7209	164	FRESNO
	4I	4-9	2	4631	74	7880	156	FRESNO
17-2	9I		1	5930	141	6523	153	BAKERSFIELD
	9I	4-9	2	4478	78	7498	180	BAKERSFIELD
17-3	55C		1	6409	124	7885	131	OILDALE
	55C	4-9	2	5415	92	7572	139	OILDALE
17-4	55I		1	6151	143	7900	97	VISALIA
	55I	4-9	2	4531	79	8537	97	VISALIA
18-0	20G		1	5672	93	8380	210	TAFT
	20G	3-9	2	4314	54	8706	332	TAFT
18-1	10I		1	5696	133	8404	88	TAFT
	10I	3-9	2	4667	89	9019	113	TAFT
18-2	12I		1	5603	193	7954	66	HANFORD
	12I	3-9	2	4101	119	7623	111	HANFORD
18-3	13I		1	5704	172	8776	56	CORCORAN
	13I	3-9	2	4175	63	9400	108	CORCORAN
19	37C		1	6071	170	8956	54	MAMMOTH LAKES
	37C	3-11	2	5325	106	7595	173	MAMMOTH LAKES
20	7I		0	3901	118	9324	60	KEELER

* These values are times 10^{-4}.

SET[a]	IDX[b]	MO[c]	SN[d]	β*	SEβ*	r^2*	PTS	LOCATION
21	8I		0	6224	104	9042	207	LONE PINE
22	52I		1	7218	170	7377	117	MODESTO
	52I	3-9	2	4228	94	6153	156	MODESTO
23	15I		1	5945	169	7200	86	MERCED
	15I	3-8	2	4681	78	7981	80	MERCED
24	39I		1	6501	136	8050	108	STOCKTON
	39I	3-9	2	4514	76	7371	144	STOCKTON
25-0	27G		1	7298	111	8726	128	CONCORD
	27G	3-11	2	5419	65	6754	488	CONCORD
25-1	2I		1	7283	199	8501	43	CONCORD
	2I	3-11	2	5116	104	6994	167	CONCORD
25-2	19I		1	7279	168	9086	43	NAPA
	19I	3-11	2	5569	89	7953	165	NAPA
25-3	60I		1	7343	218	8458	42	LIVERMORE
	60I	3-11	2	5525	136	5636	156	LIVERMORE
26-0	28G		1	5967	89	8012	244	REDWOOD CITY
	28G	3-9	2	4710	49	7862	370	REDWOOD CITY
26-1	41I		1	5869	164	7371	80	REDWOOD CITY
	41I	3-9	2	4917	85	7700	134	REDWOOD CITY
26-2	56P		1	5990	108	7964	164	SAN JOSE
	56P	3-9	2	4653	60	7580	236	SAN JOSE
27	3I		1	6519	165	8322	90	BETHEL ISLAND
	3I	4-9	2	5234	89	8539	124	BETHEL ISLAND
28	70I		1	6549	162	7565	106	SAN FRANCISCO
	70I	3-7	2	5304	104	8148	93	SAN FRANCISCO
29	14I		1	6344	170	6117	89	SAN RAFAEL
	14I	3-8	2	5235	97	6696	105	SAN RAFAEL
30-0	32G		1	7271	81	8412	324	ROCKLIN
	32G	3-9	2	4976	47	7516	500	ROCKLIN
30-1	57P		1	7338	87	8373	279	SACRAMENTO
	57P	3-9	2	4978	51	7620	405	SACRAMENTO
30-2	22I		1	6441	164	8086	45	ROCKLIN
	22I	3-9	2	4968	111	6791	95	ROCKLIN
31-0	33G		1	6242	125	7539	190	CHICO
	33G	4-9	2	4900	66	7476	236	CHICO
31-1	1I		1	7257	202	8683	50	CHICO
	1I	4-9	2	5380	102	8923	38	CHICO
31-2	53I		1	5683	203	7037	74	YUBA CITY
	53I	4-9	2	4924	116	6656	94	YUBA CITY
31-3	59I		1	6246	200	7559	66	WOODLAND
	59I	4-9	2	4734	88	7911	104	WOODLAND

* These values are times 10^{-4}.

Table 8A (continued)

SET[a]	IDX[b]	MO[c]	SN[d]	β*	SEβ*	r^2*	PTS	LOCATION
32-0	34G		1	6939	81	8308	241	HEALDSBURG
	34G	3-9	2	4832	58	8098	274	HEALDSBURG
32-1	58C		1	6666	120	7344	134	HEALDSBURG
	58C	3-9	2	4834	155	8302	155	HEALDSBURG
32-2	50I		1	7194	102	8936	107	CLOVERDALE
	50I	3-9	2	4830	97	7880	119	CLOVERDALE
33-0	35G		1	6062	124	7692	136	REDDING
	35G	4-9	2	4924	84	8458	112	REDDING
33-1	47I		1	5878	193	7841	56	REDDING
	47I	4-9	2	4830	97	7880	119	REDDING
33-2	54I		1	6176	161	7575	80	RED BLUFF
	54I	4-9	2	4982	108	8345	67	RED BLUFF
34	48I		1	6645	116	8970	67	YREKA
	48I	3-10	2	5146	77	8243	115	YREKA
35	5I		0	6087	80	8691	175	EUREKA
36-1			0	5415	18	8254	5497	STATEWIDE**
36-2			1	6219	27	8269	3321	STATEWIDE
		3-9	2	4714	16	8403	4393	STATEWIDE
36-3		3-6	1	6219	27	8269	3321	STATEWIDE
36-4		6-9	1	6219	27	8269	3321	STATEWIDE

* These values are times 10^{-4}.
** This regression set, statewide, nonseasonal, was never used.

TABLE 8A (continued)

[a] Regression set numbers are similar to the original groups. The first number indicates a geographic region, the second is an extension indicating a finer division. If the extension is a zero (-0) then all the data from the other following regression sets with the same first number are included in it.

[b] The index numbers show the connection to older representations. The code is as follows, based upon the final letter:

G This is a group coefficient from the final group list whose number precedes. E.g., 4G means this is the group coefficient from group 4.

C This is a sequence of stations from the final group list whose number precedes. E.g., 3C means this is the sequence composed of members of group 3.

P This is a pool of stations from the final group list whose number precedes. E.g., 50P means this is the pool of stations composed of members of group 50. The name of the location shows the names of the stations included if these were also used separately and their individual members follow; specifically 50P and 8P.

I This is an individual stations numbered according to our original station numbers (1-70 and 71-95). E.g., 67I is station number 67 on the original group list (ordered by county and site).

[c] These are the numbers of the months that constitute the second season.

[d] This column indicates seasons: 0 means non-season, 1 is the non-summer season, 2 is the summer season indicated in the month column.

Table 8B. Assignments of Stations to Regression Sets Based on PM10 Or TSP Data for 1982-1989 from Stations that had Pairs of Observations on the Same Day. June 25, 1991

Set[a]	Name	Co/Site	Basin	Coefficient[b]
1-0	Chula Vista	\	6	6275/5446 (4-9)
	San Diego	80/120	6	
	San Diego	80/123	6	
1-1	Chula Vista	80/114	6	6133/5294 (4-9)
1-2	El Cajon	80/131	6	6352/5549 (4-9)
	El Cajon	80/104	6	
2	Oceanside	80/134	6	5218
	Oceanside	80/121	6	
3	El Centro		10	4962
	El Centro	13/682	11	
	El Centro	13/694	11	
	Calexico	13/681	11	
	Brawley	13/682	11	
	Brawley	13/683	11	
4-1	Palm Springs	33/137	11	4391
4-2	Indio	33/157	11	4619
	Indio	33/139	11	
5-0	El Toro		5	4953
	Anaheim	30/176	5	
	La Habra	30/177	5	
	Costa Mesa	30/185	5	
	Santa Ana	30/187	5	
	San Juan Capistrano	30/188	5	
	Laguna Beach	30/189	5	
	Santa Ana	30/191	5	

Set[a]	Name	Co/Site	Basin	Coefficient[b]
	Costa Mesa	30/192	5	
	West Los Angeles	70/71	5	
	Lennox	70/76	5	
	West Los Angeles	70/86	5	
	West Los Angeles	70/91	5	
5-1	El Toro	30/186	5	4901
5-2	Los Alamitos	30/190	5	4974
5-3	Hawthorne	70/94	5	5063
6-1	Rubidoux	33/144	5	6447
6-2	Magnolia	33/146	5	5347
	Riverside	33/142	5	
6-3	Perris	33/149	5	5780
6-4	Banning	33/150	11	5473
	Banning	33/133	11	
6-5	San Bernardino		5	5847
	San Bernardino	36/203	5	
	San Bernardino	36/194	5	
	San Bernardino	36/151	5	
	Redlands	36/165	5	
	Redlands	36/192	5	
6-6	Ontario	36/171	5	6177
	Ontario	36/185	5	
	Chino	36/173	5	
	Chino	36/198	5	
6-7	Fontana-Azusa		5	5495
	Rialto	36/166	5	
	Upland	36/174	5	
	Upland	36/175	5	

Set[a]	Name	Co/Site	Basin	Coefficient[b]
	Upland	36/189	5	
	Glendora	70/591	5	
6-8	Fontana	36/197	5	5577
	Fontana	36/170	5	
	Fontana	36/176	5	
6-9	Azusa	70/60	5	5383
7-1	Burbank	70/69	5	6230
	Reseda	70/74	5	
	Pasadena	70/83	5	
	Pasadena	70/88	5	
	Temple City	70/580	5	
	Mount Lee	70/581	5	
7-2	LA/Long Beach		5	5855
	Lynwood	70/84	5	
	Pico Rivera	70/85	5	
7-3	North Long Beach	70/72	5	5822
7-4	Los Angeles	70/87	5	5885
	Los Angeles	70/1	5	
8	Simi Valley		4	6015/5151 (4-9)
	Simi Valley	56/434	4	
	Simi Valley	56/413	4	
	Moorpark College	56/411	4	
	Thousand Oaks	45/415	4	
9-0	Morro Bay		4	5694/4928 (4-8)
	Carpenteria	42/359	4	
	El Capitan Beach	42/370	4	
	Ventura	56/401	4	
	Santa Paula	56/404	4	

Set[a]	Name	Co/Site	Basin	Coefficient[b]
	Oxnard	56/405	4	
	Camarillo	56/408	4	
	Point Mugu	56/409	4	
	Oxnard	56/410	4	
	Port Hueneme	56/412	4	
	Ventura	46/414	4	
	Camarillo	45/416	4	
	Ventura	56/420	4	
	Ventura	56/421	4	
	Ventura	56/429	4	
9-1	Goleta-Santa Barbara		4	5671/5114 (4-8)
	Goleta	42/363	4	
	Santa Barbara	42/388	4	
	Santa Barbara	42/354	4	
	Santa Barbara	42/355	4	
	Santa Barbara	42/378	4	
9-2	Morro Bay	40/833	4	5149/4955 (4-8)
9-3	El Rio	56/419	4	5926/4728 (4-8)
10	Piru	56/427	4	5738
	Piru	56/418	4	
	Ojai	56/402	4	
	Ojai	56/430	4	
11	Victorville		11	4573
	Victorville	36/199	11	
	Victorville	36/190	11	
	Victorville	36/168	11	
	Mojave	15/207	11	
	Hesparia	36/201	11	

Table 8B (continued)

Set[a]	Name	Co/Site	Basin	Coefficient[b]
12-0	Barstow		11	5293
	Boron	15/209	11	
	Twentynine Palms	36/191	11	
12-1	Lancaster		11	5147
	Lancaster	70/82	11	
	Lancaster	70/593	11	
12-2	Barstow	36/155	11	5396
13	Trona	36/188	11	4166
	China Lake	15/211	11	
14	Santa Maria	42/536	4	4520/4061 (3-9)
	Santa Maria	42/366	4	
	San Luis Obispo	40/831	4	
	San Luis Obispo	40/835	4	
	San Luis Obispo	40/845	4	
	Nipomo	40/834	4	
	Nipomo	40/848	4	
	Lompoc	42/360	4	
	Lompoc	42/365	4	
	Lompoc	42/381	4	
	Santa Ynez	42/369	4	
15-0	Salinas		3	5487/4614 (4-9)
	Gonzales	27/537	3	
15-1	Salinas	27/544	3	5406/4645 (4-9)
	Salinas	27/531	3	
	Salinas	27/536	3	
15-2	Hollister	35/823	3	5873/4501 (4-9)
	Hollister	35/821	3	
16	Santa Cruz	44/850	3	4869/3900 (4-11)

Set[a]	Name	Co/Site	Basin	Coefficient[b]
	Santa Cruz	44/841	3	
	Santa Cruz	44/849	3	
	Monterey	27/538	3	
	Aptos	44/845	3	
17-1	Fresno	10/234	9	6164/4631 (4-9)
	Fresno	10/226	9	
	Fresno	10/232	9	
	Fresno	10/241	9	
17-2	Bakersfield	15/203	9	5930/4478 (4-9)
	Bakersfield	15/202	9	
17-3	Oildale		9	6409/5415 (4-9)
	Oildale	15/243	9	
	Oildale	15/241	9	
	Oildale	15/230	9	
17-4	Visalia	54/568	9	6151/4531 (4-9)
	Visalia	54/561	9	
18-0	Taft		9	5672/4314 (3-9)
	Five Points	10/229	9	
	Parlier	10/230	9	
	Coalinga	10/231	9	
	Kern Refuge	15/213	9	
	McKittrick	15/234	9	
	McKittrick	15/240	9	
	Kettleman City	16/707	9	
18-1	Taft	15/213	9	5696/4667 (3-9)
	Taft	15/204	9	
18-2	Hanford	16/701	9	5603/4101 (3-9)
18-3	Corcoran	16/715	9	5704/4175 (3-9)

Table 8B (continued)

Set[a]	Name	Co/Site	Basin	Coefficient[b]
	Corcoran	16/708	9	
19	Mammoth Lakes		10	6071/5325 (3-11)
	Mammoth Lakes	26/776	10	
20	Keeler	14/697	10	3901
21	Lone Pine	14/699	10	6224
	Bishop	14/695	10	
	Bishop	14/702	10	
22	Modesto	50/558	9	7218/4228 (3-9)
	Modesto	50/557	9	
	Modesto	50/567		
	Turlock	50/561	9	
23	Merced	24/521	9	5945/4681 (3-8)
	Merced	24/524	9	
	Los Banos	24/522	9	
24	Stockton	39/252	9	6501/4514 (3-9)
	Stockton	39/264	9	
	Stockton	39/265	9	
	Lodi	39/260	9	
	Union Island	39/261	9	
25-0	Concord		2,8	7298/5419 (3-11)
	Pittsburg	7/430	2	
	Vallejo	48/874	2	
	Vallejo	48/879	2	
	Vacaville	48/878	8	
	Vacaville	48/881	8	
25-1	Concord	7/440	2	7283/5116 (3-11)
	Concord	7/436	2	
25-2	Napa	28/783	2	7279/5569 (3-11)

Ser[a]	Name	Co/Site	Basin	Coefficient[b]
25-3	Livermore	60/340	2	7343/5525 (3-11)
	Livermore	60/335	2	
26-0	Redwood City		2	5967/4710 (3-9)
	North Richmond	7/431	2	
	Burlingame	41/545	2	
	Sunnyvale	43/384	2	
	Saratoga	43/388	2	
	Berkeley	50/326	2	
	Oakland	60/327	2	
	Oakland	60/334	2	
	Fremont	60/336	2	
26-1	Redwood City	41/541	2	5869/4917 (3-9)
26-2	San Jose		4	5990/4653 (3-9)
	San Jose	43/377	4	
	San Jose	43/382	4	
	San Jose	43/390	4	
27	Bethel Island	7/442	2	6519/5234 (4-9)
	Rio Vista	48/877	8	
	Rio Vista	48/880	8	
28	San Francisco	90/306	2	6549/5304 (3-7)
	San Francisco	90/303	2	
	San Francisco	90/304	2	
29	San Rafael	21/451	2	6344/5235 (3-8)
30-0	Rocklin		8,12	7271/4976 (3-9)
	Auburn	31/809	12	
	Auburn	31/813	12	
	Rancho Cordova	34/284	8	
	North Highlands	34/294	8	

Table 8B (continued)

Set[a]	Name	Co/Site	Basin	Coefficient[b]
	West Sacramento	57/570	8	
30-1	Sacramento		8	7338/4978 (3-9)
	Sacramento	34/277	8	
	Citrus Heights	34/293	8	
	Sacramento	34/295	8	
	Sacramento	34/282	8	
	Sacramento	34/283	8	
30-2	Rocklin	31/810	12	6441/4968 (3-9)
31-0	Chico		8	6242/4900 (4-9)
	Oroville	4/626	8	
	Gridley	4/627	8	
	Willows	11/674	8	
	Live Oak	51/892	8	
	Pleasant Grove	51/893	8	
	Sutter City	51/894	8	
	Sutter City	51/896	8	
	Dunnigan	57/571	8	
	Marysville	58/931	8	
	Wheatland	58/932	8	
	Wheatland	58/934	8	
31-1	Chico	4/630	8	7257/5380 (4-9)
	Chico	4/621	8	
	Chico	4/628	8	
31-2	Yuba City	51/895	8	5683/4924 (4-9)
	Yuba City	51/891	8	
31-3	Woodland	57/569	8	6246/4734 (4-9)
32-0	Healdsburg		1,2	6939/4832 (3-9)
	Calpella	23/752	1	

Set[a]	Name	Co/Site	Basin	Coefficient[b]
	Willits	23/753	1	
	Ukiah	23/760	1	
	Ukiah	23/764	1	
	Guerneville	49/895	1	
	Santa Rosa	49/884	2	
	Santa Rosa	49/893	2	
32-1	Healdsburg		1	6666/4832 (3-9)
	Healdsburg	49/885	1	
	Healdsburg	49/898	1	
32-2	Cloverdale	49/886	1	7194/4830 (3-9)
33-0	Redding		8	6062/4924 (4-9)
	Anderson	45/554	8	
	Anderson	45/558	8	
	Buckeye	45/559	8	
	Corning	52/907	8	
	Los Molinos	52/908	8	
33-1	Redding	45/555	8	5878/4733 (4-9)
	Redding	45/553	8	
33-2	Red Bluff	52/901	8	6176/4982 (4-9)
	Red Bluff	52/903	8	
	Red Bluff	52/905	8	
	Red Bluff	52/906	8	
34	Yreka	47/861	7	6645/5146 (3-10)
	Weed	47/865	7	
	Fort Jones	47/868	7	
35	Eureka	12/503	1	6087
	Eureka	12/506	1	
	Arcata	12/504	1	

Table 8B (continued)

Set[a]	Name	Co/Site	Basin	Coefficient[b]
	Fort Bragg	23/756	1	
	Fort Bragg	23/760	1	
36-2	STATE			6192/4713 (3-9)
	Carmel Valley	27/550	3	6192/4713 (3-9)
	Paso Robles	40/832	4	
	Lockwood Valley	56/417	4	
	Escondido	80/115	6	
	Alpine	80/128	6	
	Alturas	25/761	7	
	Alturas	25/763	7	
	Burney	45/556	8	
	Burney	45/565	8	
	Smartville	58/933	8	
	Lakeport	17/711	13	
	Kelseyville	17/712	13	
	Lakeport	17/713	13	
	Kelseyville	17/714	13	
	Upper Lake	17/715	13	
36-3	Yosemite	22/742	12	6192/4713 (3-6)
36-4	Lake Gregory	36/181	5	6192/4713 (6-9)
	Big Bear Lake	36/184	5	

a Set numbers. If these end with -0 then these are group coefficients composed of more than one station and hence have no Co/Site designation. Otherwise, the regression set season is indicated and the stations receiving that coefficient follow.

b Format is: one number means there are no seasons and this is the regression coefficient (the number is times 10^{-4}); two numbers separated by a slash indicate regression coefficients for separate seasons with numbers in parenthesis indicating months for second (summer) season.

Example: 5218, no season, coefficient is 0.5218.

Example: 6275/5446 (4-9); two seasons, 0.6275 for (non-summer) months, 0.5446 for the months 4 thu 9.

Table 9. Comparison of Average Annual Actual vs Interpolated Indirectly Estimated Exceedance Frequencies[1], Mean Concentrations and Excess Concentrations at 39 Monitoring Stations with at Least 20 Months of Paired Data for 1986 and 1987.
(n = 39)

Statistic Hours >	Actual Mean	Interpolated Estimated Mean	Paired t	p	Correl-ation	Regression coefficients[2] b	SEb	a	SEa[2]
Exceedance and Mean Concentration Frequencies (n=39)									
>40 mcg/m³	4057	4087	-.15	.88	.84	.75	.08	996.6	379.5
>50 mcg/m³	2897	2755	.79	.43	.84	.82	.09	614.6	294.5
>60 mcg/m³	2053	1786	2.03	.05	.86	.96	.09	344.4	211.4
>80 mcg/m³	983	669	3.39	.002	.86	1.34	.13	90.3	122.3
>100 mcg/m³	484	313	2.22	.03	.67	1.23	.22	99.2	103.9
Mean Conc.	45.7	42.8	2.33	.03	.86	.98	.10	48.1	51.4
Excess Concentrations									
>40 mcg/m³	9909	8046	2.54	.02	.84	1.11	.12	953	1199
>50 mcg/m³	6991	5222	2.79	.01	.83	1.24	.14	518	939
>60 mcg/m³	4912	3346	2.85	.01	.81	1.36	.16	355	758
>80 mcg/m³	2419	1467	2.38	.02	.70	1.46	.24	279	528
>100 mcg/m³	1247	699	2.00	.05	.53	1.26	.33	363	359

[1] Two year cumulations were divided by 2 so as to be equivalent to annual cumulations.
[2] Regression Coefficients and their standard errors for slope b and intercept a of regression of actual on interpolated values.

Table 10. Comparison of Average Annual Actual vs Interpolated Actual Exceedance Frequencies[1], Mean Concentrations and Excess Concentrations at 39 Monitoring Stations with at Least 20 Months of Paired Data for 1986 and 1987.

(n = 39)

Statistic	Actual Annual Mean	Interpolated Actual Mean	Paired t	p	Correl-ation	Regression coefficients[2] b	SEb	a	SEa
				Exceedance Frequency and Mean Concentration					
>Hours >40 mcg/m³	4057	3950	.58	.56	.85	.84	.09	733.7	382.4
>50 mcg/m³	2897	2882	.10	.92	.88	.87	.08	382.4	268.7
>60 mcg/m³	2053	2036	.13	.90	.87	.91	.08	206.2	211.2
>80 mcg/m³	983	983	.001	.999	.87	.99	.09	8.8	125.5
>100 mcg/m³	484	483	.01	.99	.79	1.00	.13	1.0	87.9
Mean Conc.	45.7	45.4	.22	.83	.88	.90	.08	4.6	3.8
				Excess Concentrations					
Hour x Mcg/m³ > >40 mcg/m³	9909	9958	-.08	.94	.89	.94	.08	508	1030
>50 mcg/m³	6991	7099	-.21	.84	.88	.97	.09	130	800
>60 mcg/m³	4912	5031	-.27	.79	.87	.98	.09	-37	639
>80 mcg/m³	2419	2525	-.38	.71	.85	1.01	.10	-127	384
>100 mcg/m³	1247	1343	-.54	.60	.83	.99	.11	-80	234

[1] Two year cumulations were divided by 2 so as to be equivalent to annual cumulations.
[2] Regression Coefficients and their standard errors for slope b and intercept a of regression of actual on interpolated actual values.

Table 11. Comparison of Actual vs Interpolated Indirectly Estimated Exceedance Frequencies, Mean Concentration and Excess Concentrations at 39 Monitoring Stations with at Least 20 Months of Paired PM10 and TSP Data for 1986 and 1987.

(n = 841 Station Months)

Statistic	Actual Mean	Interpolated Estimated Mean	Paired t	p	Correl-ation	Regression coefficients[1]			
						b	SEb	a	SEa
Exceedance Frequencies and Mean Concentrations									
> Hours >40 mcg/m^3	341.4	344.7	-.52	.60	.75	.75	.02	81.3	10.0
>50 mcg/m^3	245.4	233.3	2.01	.04	.74	.78	.02	63.2	8.1
>60 mcg/m^3	174.5	151.5	4.37	.0001	.72	.83	.03	49.5	6.6
>80 mcg/m^3	84.3	56.9	7.04	.0001	.67	.93	.04	31.3	4.4
>100 mcg/m^3	41.7	26.4	5.31	.0001	.56	.80	.04	20.7	3.0
Mean Conc	46.1	43.0	6.76	.0001	.81	.95	.02	5.1	1.1
Excess Concentrations									
Hour x mcg/m^3 > >40 mcg/m^3	10144	8172	7.08	.0001	.77	.99	.03	2014	363
>50 mcg/m^3	7188	5309	7.63	.0001	.74	1.01	.03	1833	297
>60 mcg/m^3	5072	3401	7.75	.0001	.71	1.01	.03	1650	245
>80 mcg/m^3	2518	1484	6.43	.0001	.63	.95	.04	1112	171
>100 mcg/m^3	1308	701	5.08	.0001	.54	.87	.05	700	123

[1] Regression Coefficients and their standard errors for slope b and intercept a of regression of actual on interpolated values.

Table 12. Comparison of Actual vs Interpolated Actual Exceedance Frequencies, Mean Concentration, and Excess Concentrations at 841 Monitoring Stations with at Least 20 Months of Paired PM10 and TSP Data for 1986 and 1987.

(n = 841 Station Months)

Statistic	Actual Mean	Interpolated Actual Mean	Paired t	p	Correl-ation	Regression coefficients[1]			
						b	SEb	a	SEa
Exceedance Frequencies and Mean Concentration									
Hours >									
>40 mcg/m³	341.4	333.2	1.41	.16	.79	.81	.02	72.7	9.2
>50 mcg/m³	245.4	244.3	.20	.84	.80	.82	.02	46.0	7.2
>60 mcg/m³	174.5	173.3	.26	.80	.78	.81	.02	34.3	6.0
>80 mcg/m³	84.3	84.0	.07	.94	.73	.79	.03	17.8	4.2
>100 mcg/m³	41.7	41.2	.19	.85	.66	.71	.03	12.3	2.8
Mean Conc.	46.1	45.8	.65	.52	.86	.87	.02	6.2	0.9
Excess Concentrations									
Hr x mcg/m³ >									
>40 mcg/m³	10144	10167	-.09	.93	.83	.86	.02	1447	313
>50 mcg/m³	7188	7266	-.35	.73	.82	.84	.02	1080	261
>60 mcg/m³	5072	5158	-.44	.66	.79	.82	.02	835	217
>80 mcg/m³	2518	2596	-.55	.58	.76	.77	.02	506	148
>100 mcg/m³	1308	1384	-.74	.46	.74	.74	.02	278	101

[1] Regression Coefficients and their standard errors for slope b and intercept a of regression of actual on interpolated values.

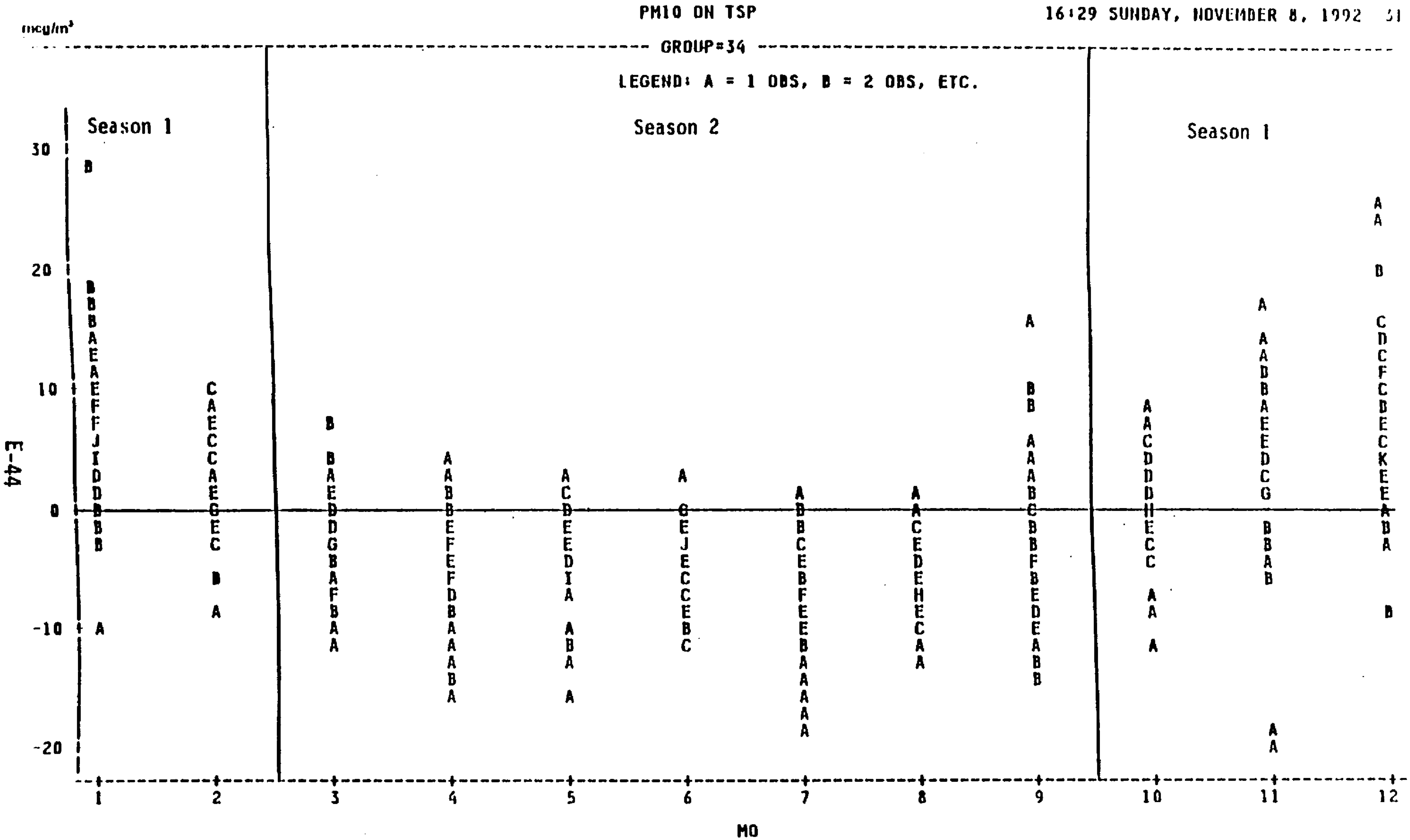


Figure 1. Plot by Month of Residuals from Regression of PM10 on TSP.

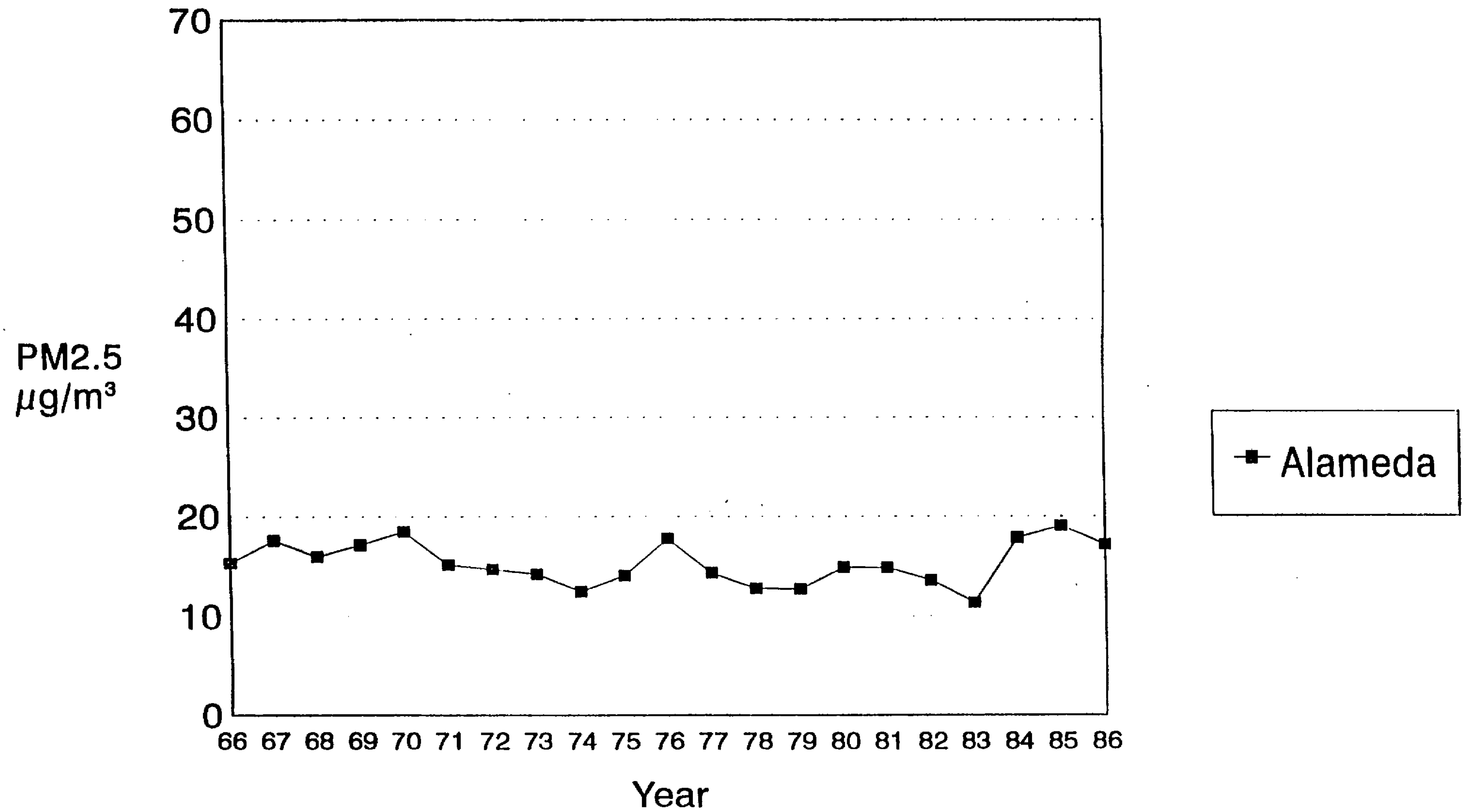
APPENDIX F

Appendix F

Table 1. Comparison of Actual Versus Interpolated Average Annual Cumulative Exceedance Frequencies (hrs/yr) and Mean Concentration at Monitoring Stations for SO_2 (1985 through 1986)

(n = 57 stations)

Statistic	Actual Mean	Interp. Mean	Paired t	p	Corre-lation	Regression Coefficients			
						ß	SEß	α	SEα
Exceedance Frequency	Hrs/yr								
> 2 pphm	33.1	33.3	-0.01	.99	0.36	0.46	0.16	18.0	9.8
> 4 pphm	0.99	0.76	0.72	0.48	0.12	0.17	0.19	0.86	0.32
> 8 pphm	0.18	0.13	0.60	0.55	0.18	0.24	0.17	0.15	0.08
>14 pphm	0.025	0.005	0.90	0.37	-0.034	-0.12	0.47	0.03	0.02
Mean Conc. (pphm)	0.103	0.105	-0.17	0.87	0.64	0.67	0.11	0.033	0.015



Appendix F

Figure 1. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of Alameda Airport.

"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.

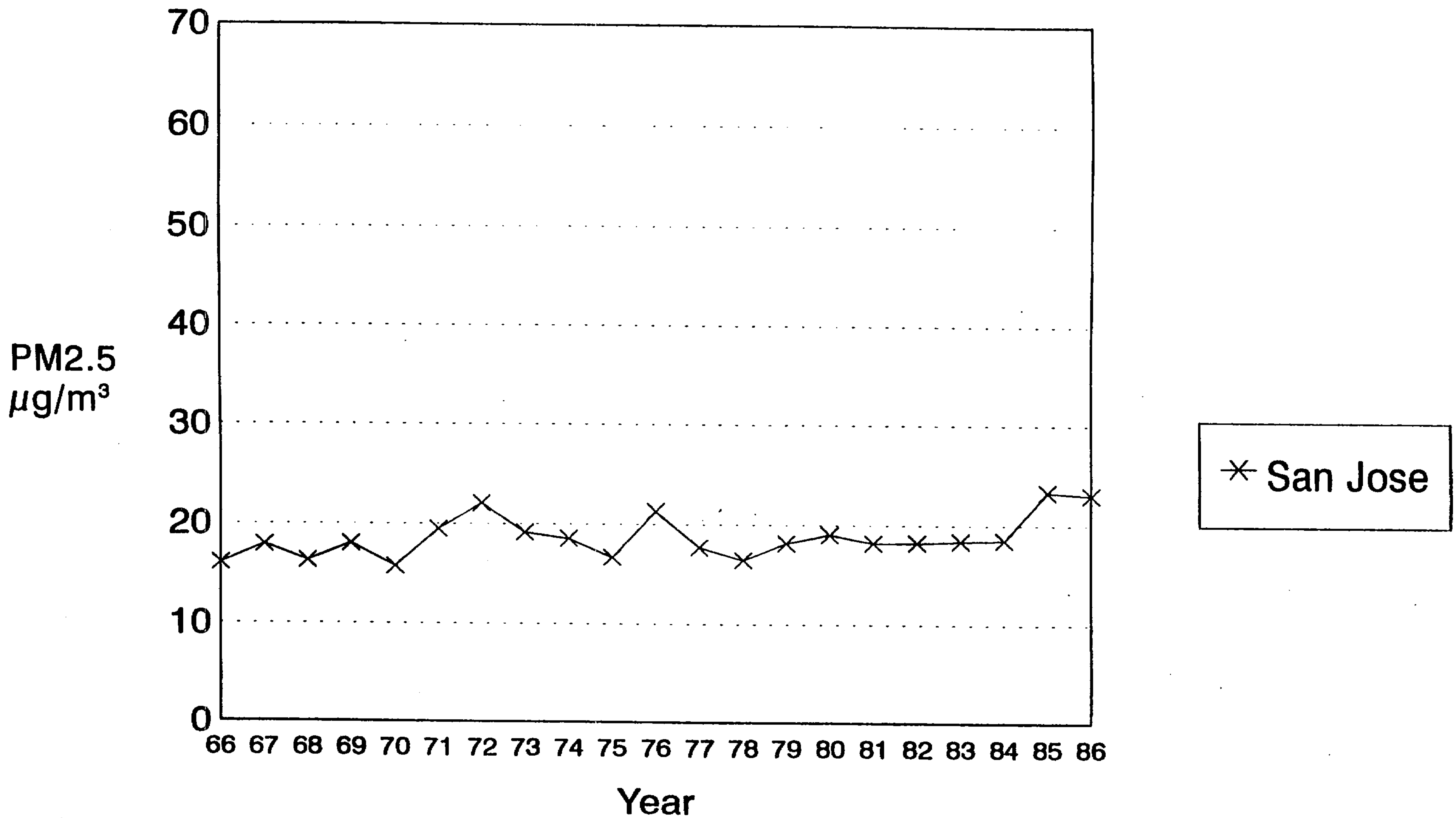
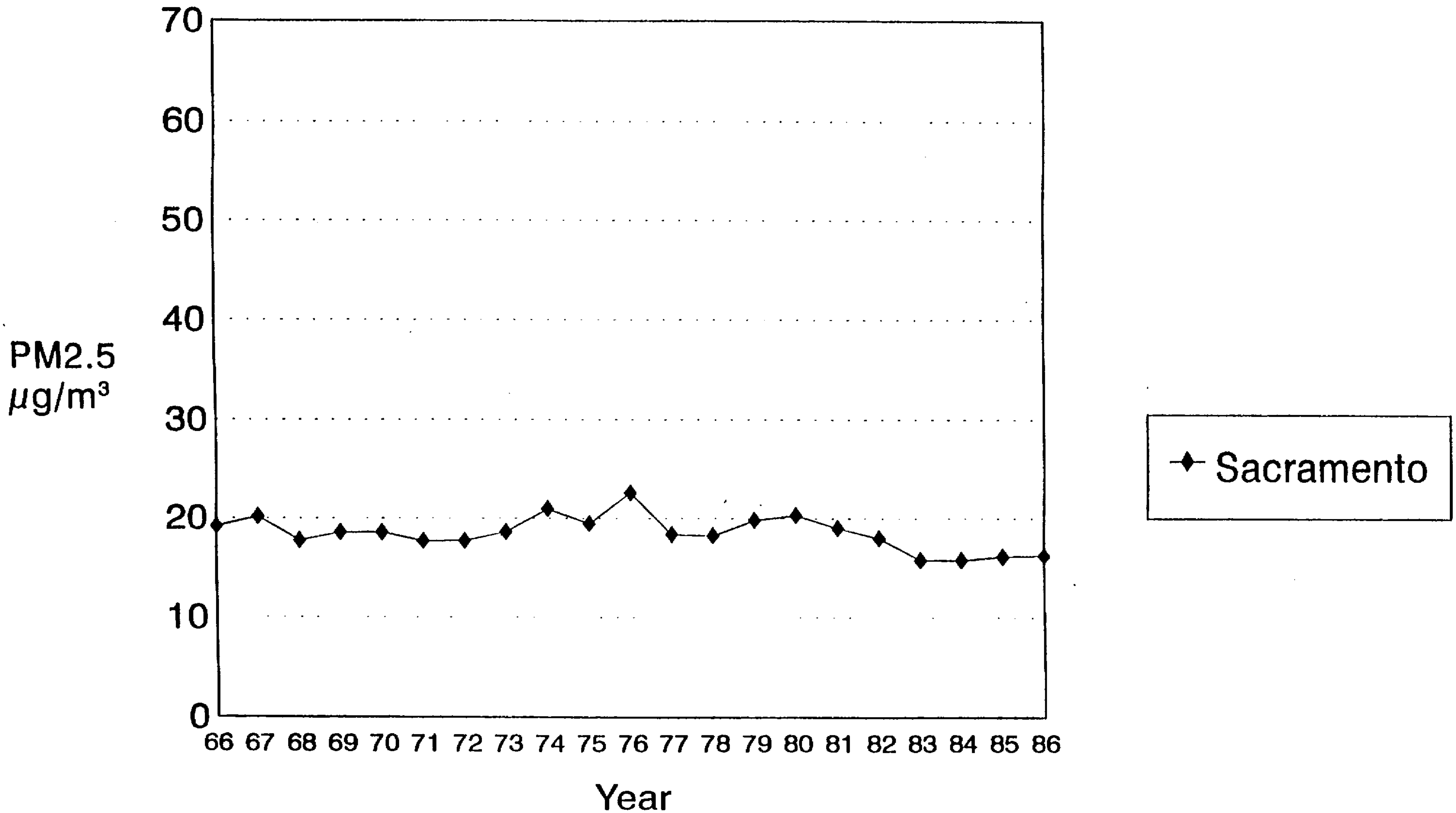


Figure 2. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of San Jose Airport.

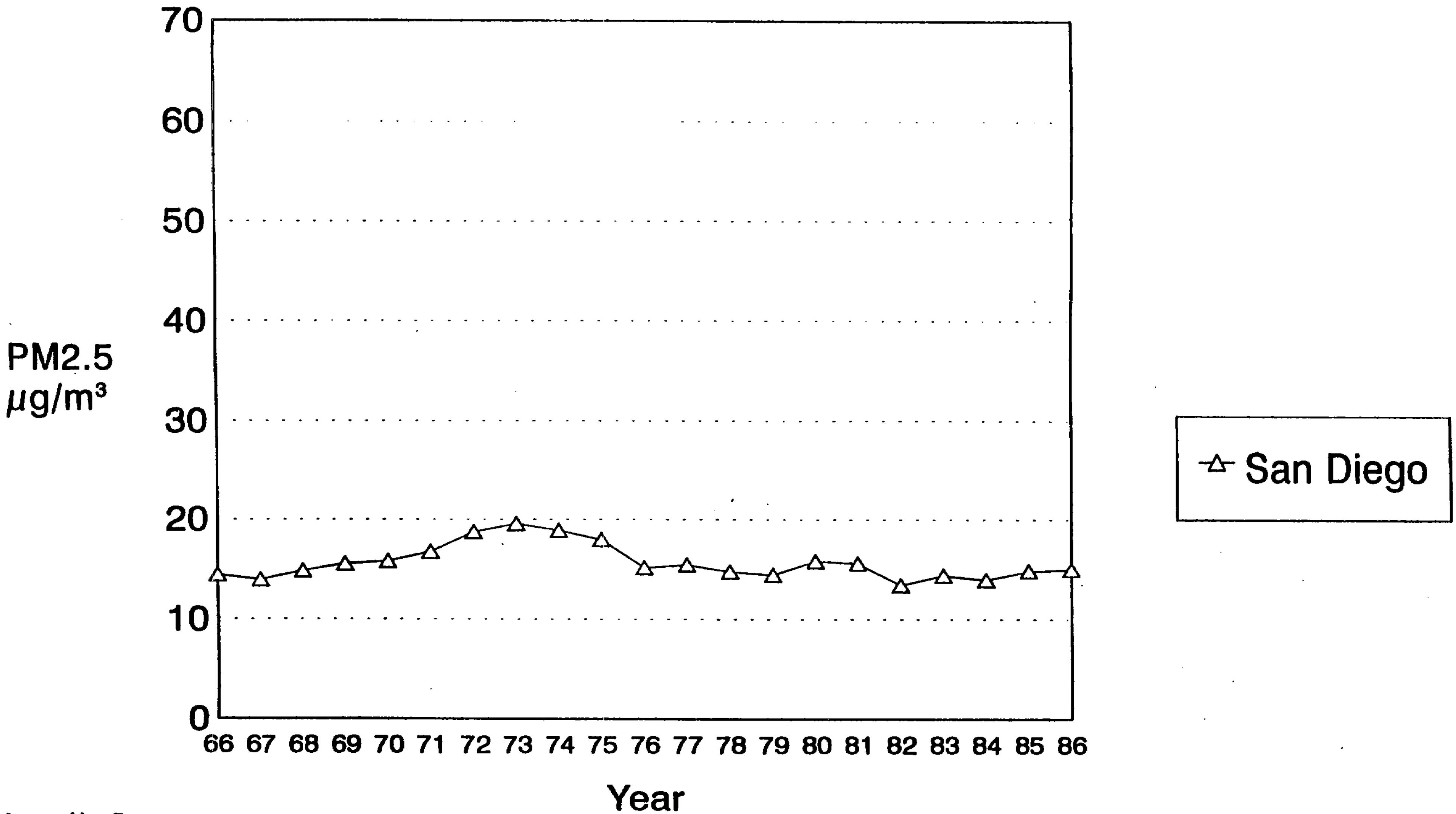
"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.



Appendix F

Figure 3. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of Sacramento Airport.

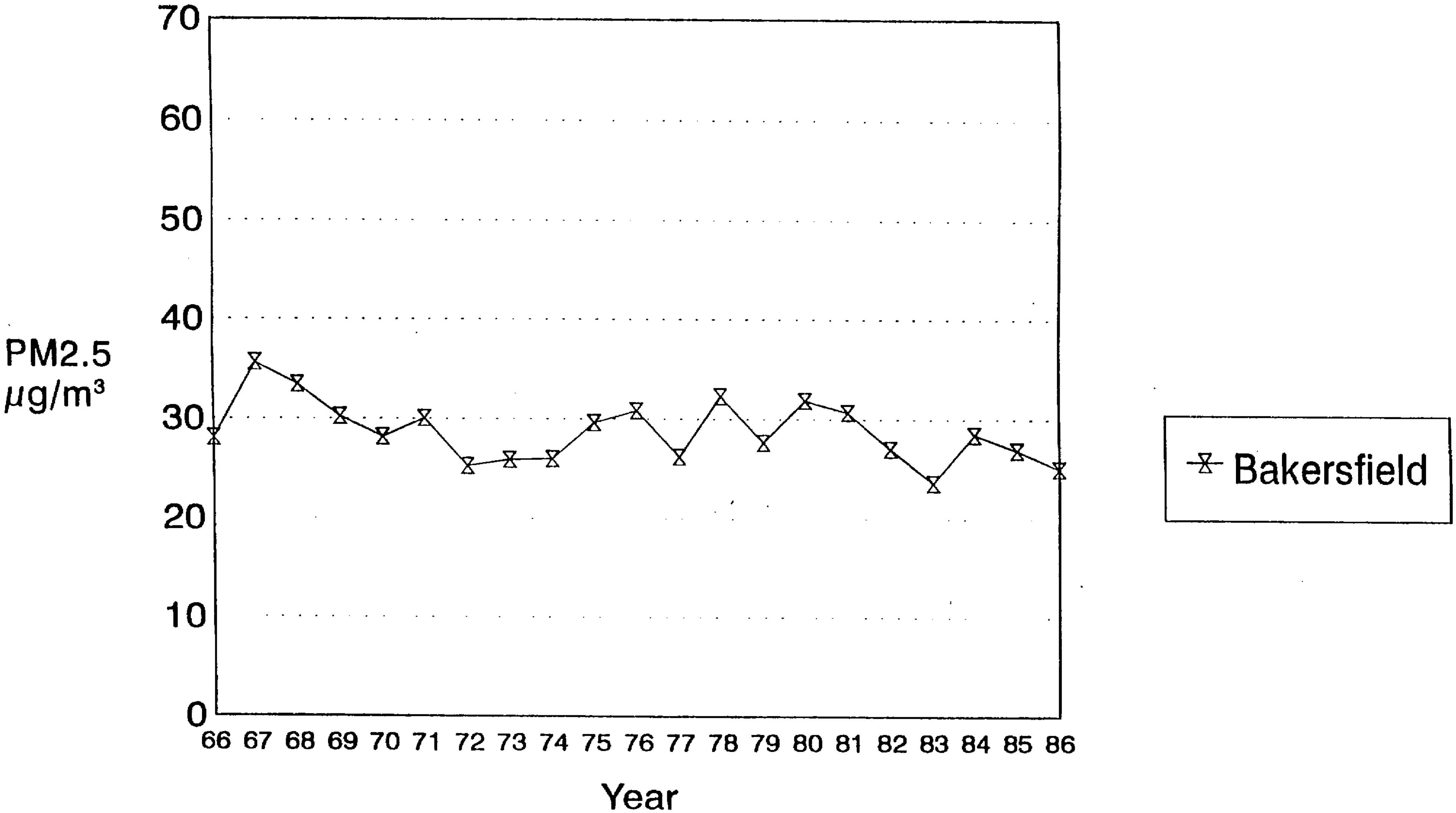
"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.



Appendix F

Figure 4. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of San Diego Airport.

"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.



Appendix F

Figure 5. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of Bakersfield Airport.

"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.

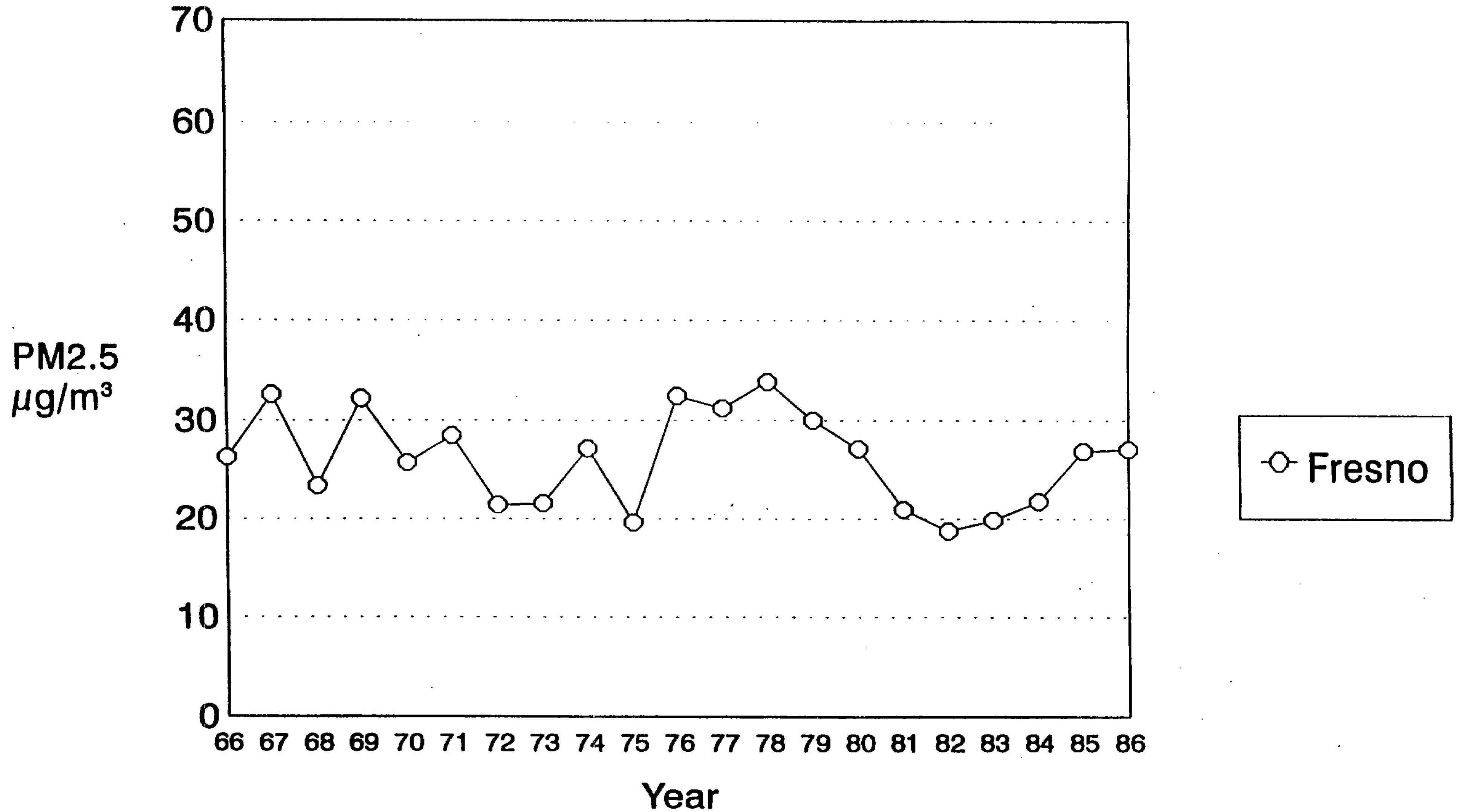
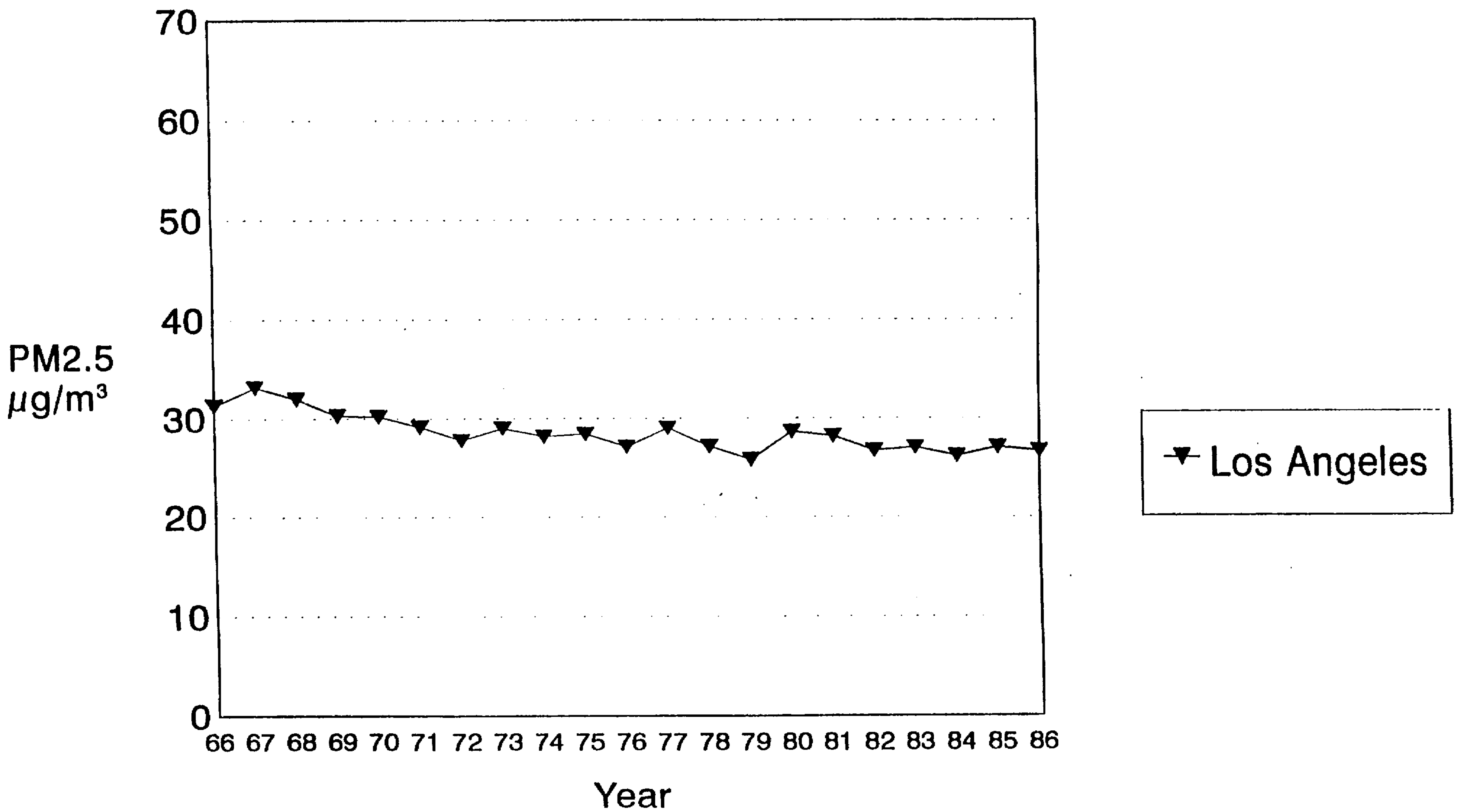


Figure 6. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of Fresno Airport.

"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.



Appendix F

Figure 7. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of Los Angeles International Airport.

"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.

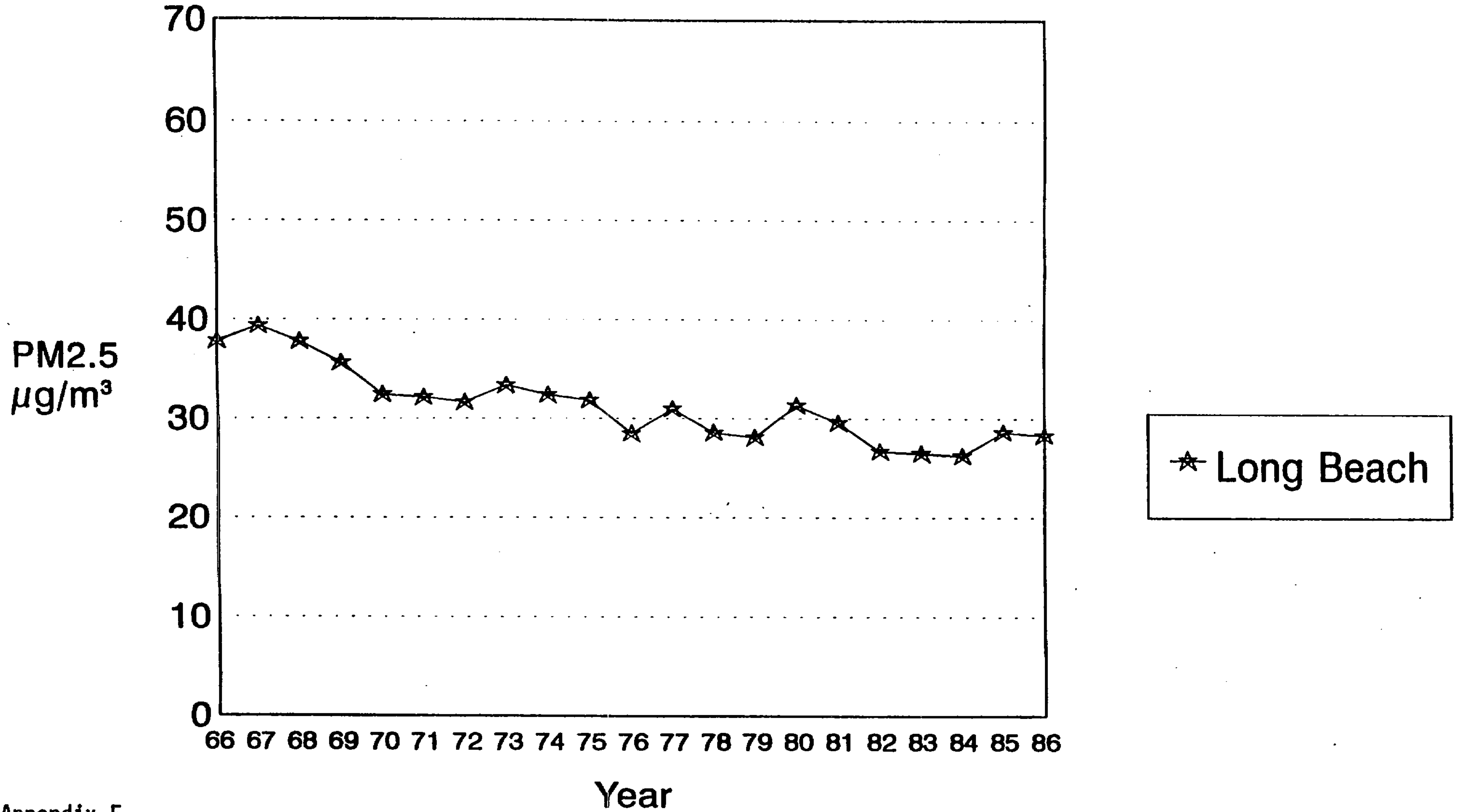


Figure 8. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of Long Beach-Dougherty Airport.

"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.

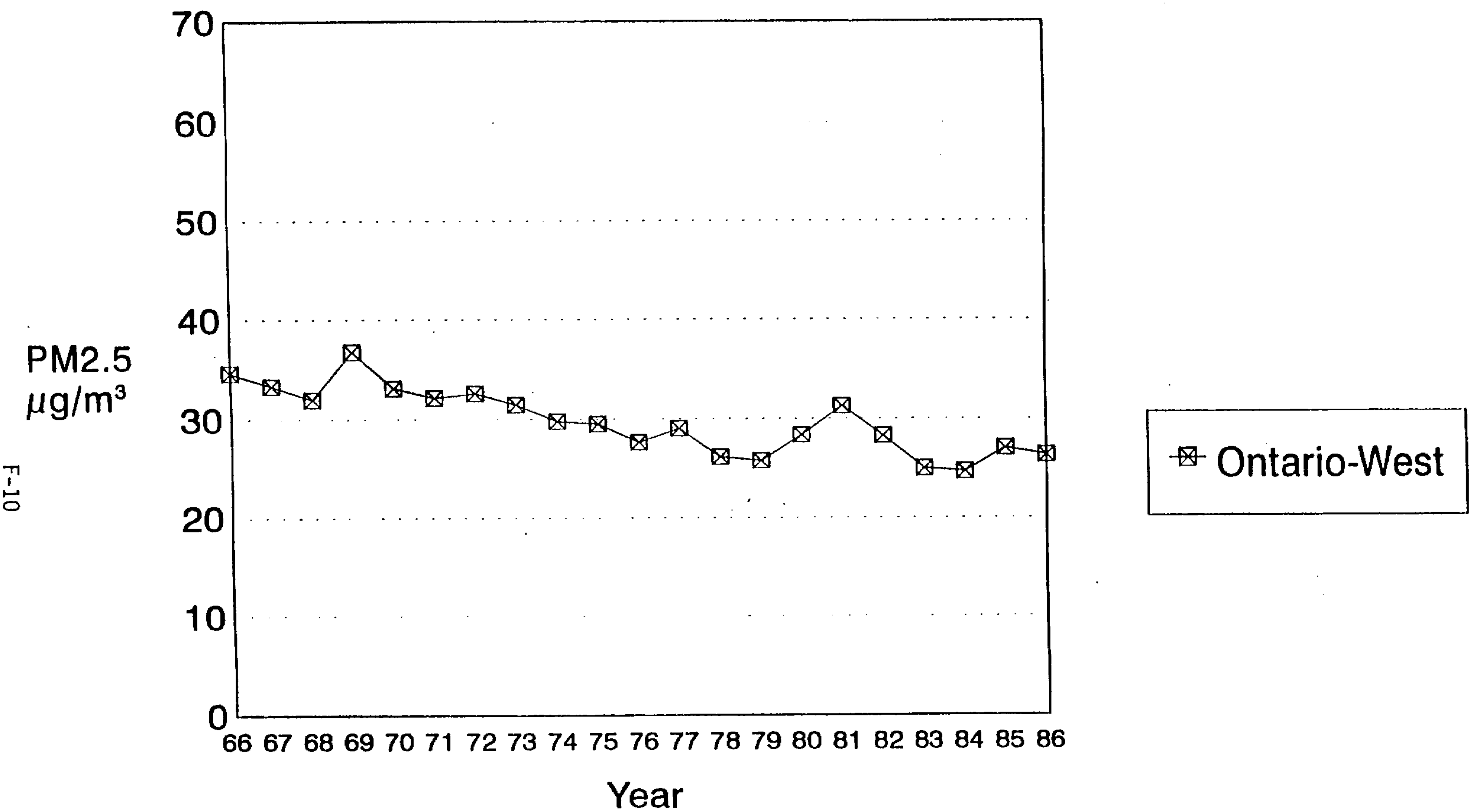


Figure 9. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of Ontario Airport - West Subregion.

"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.

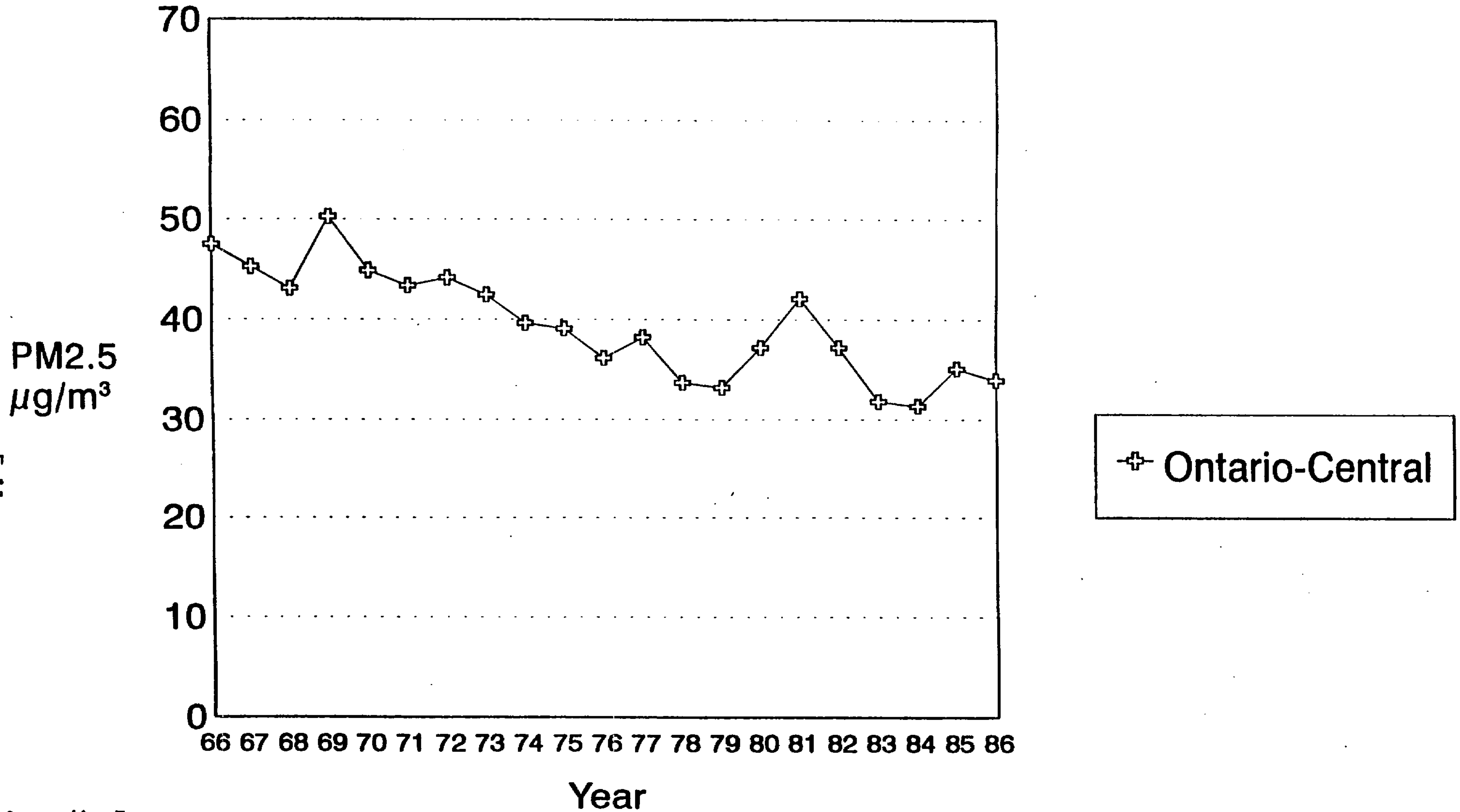


Figure 10. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of Ontario Airport - Central Subregion.

"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.

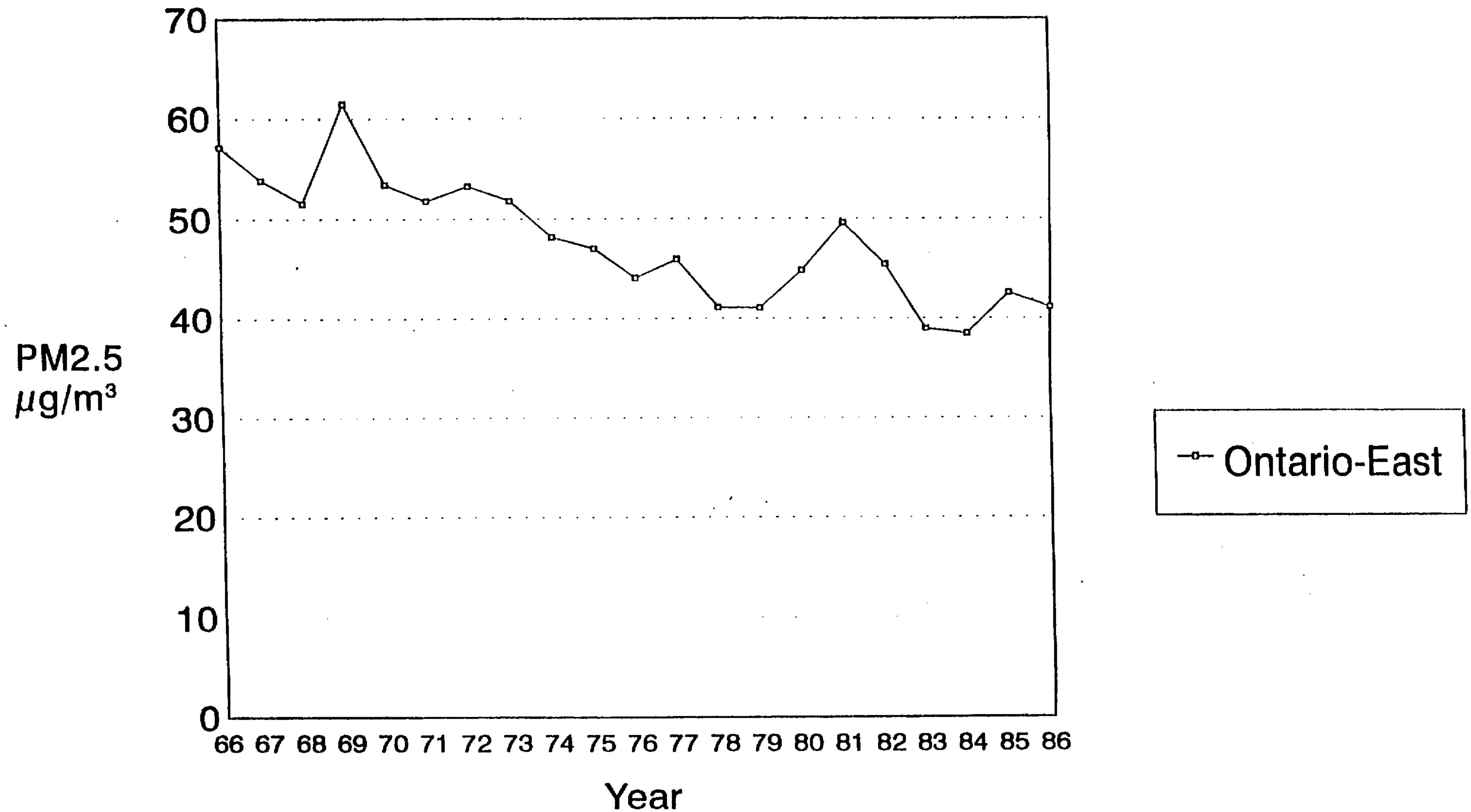
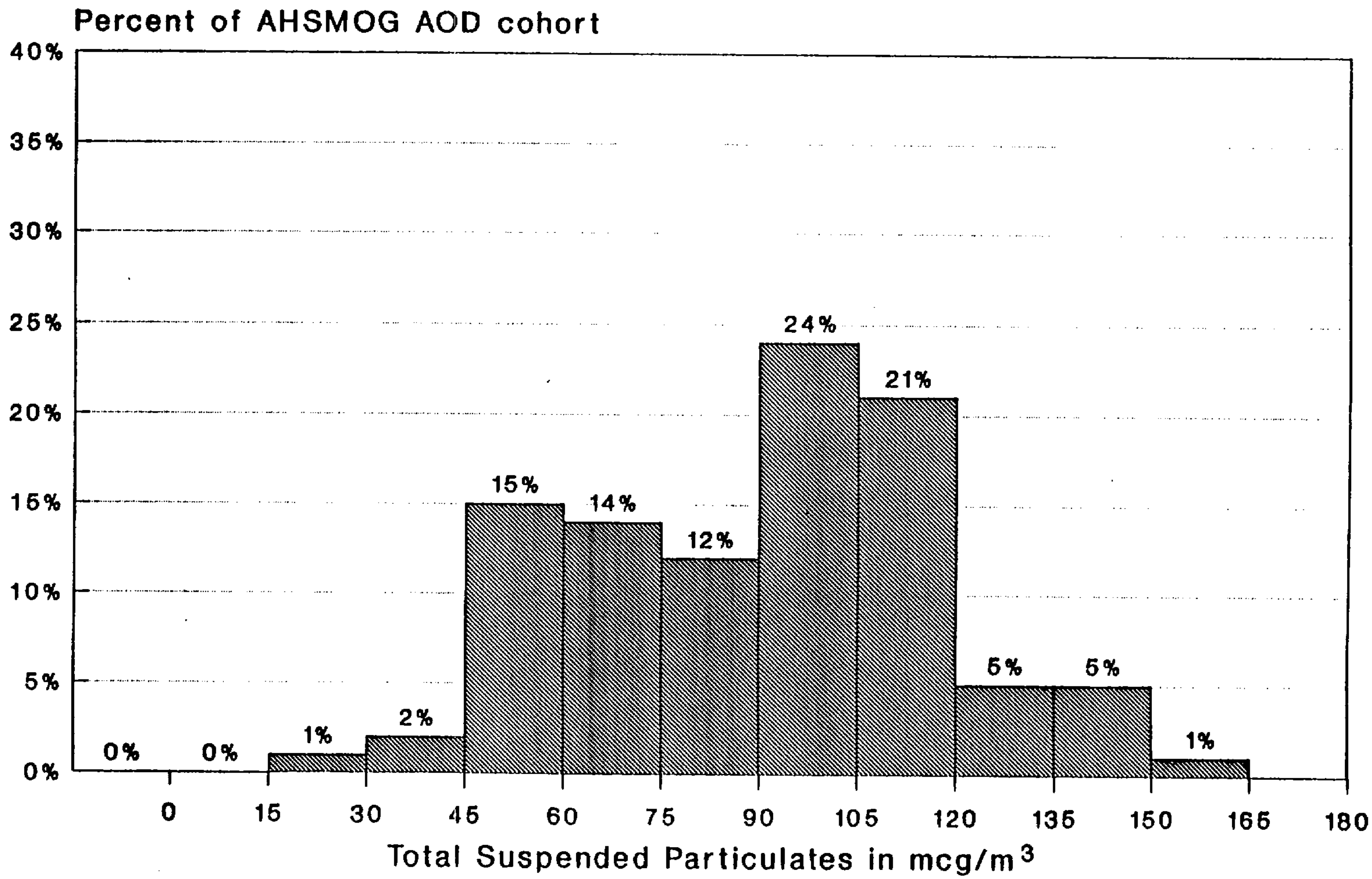


Figure 11. Plot of Annual Ambient Mean Concentration of Estimated PM2.5 in Vicinity of Ontario Airport - East Subregion.

"Low" Airports: Alameda, San Jose, Sacramento, and San Diego.
"Medium" Airports: Bakersfield, Fresno, Los Angeles, Long Beach, Ontario-Central and Ontario-West.
"High" Airport: Ontario-East.

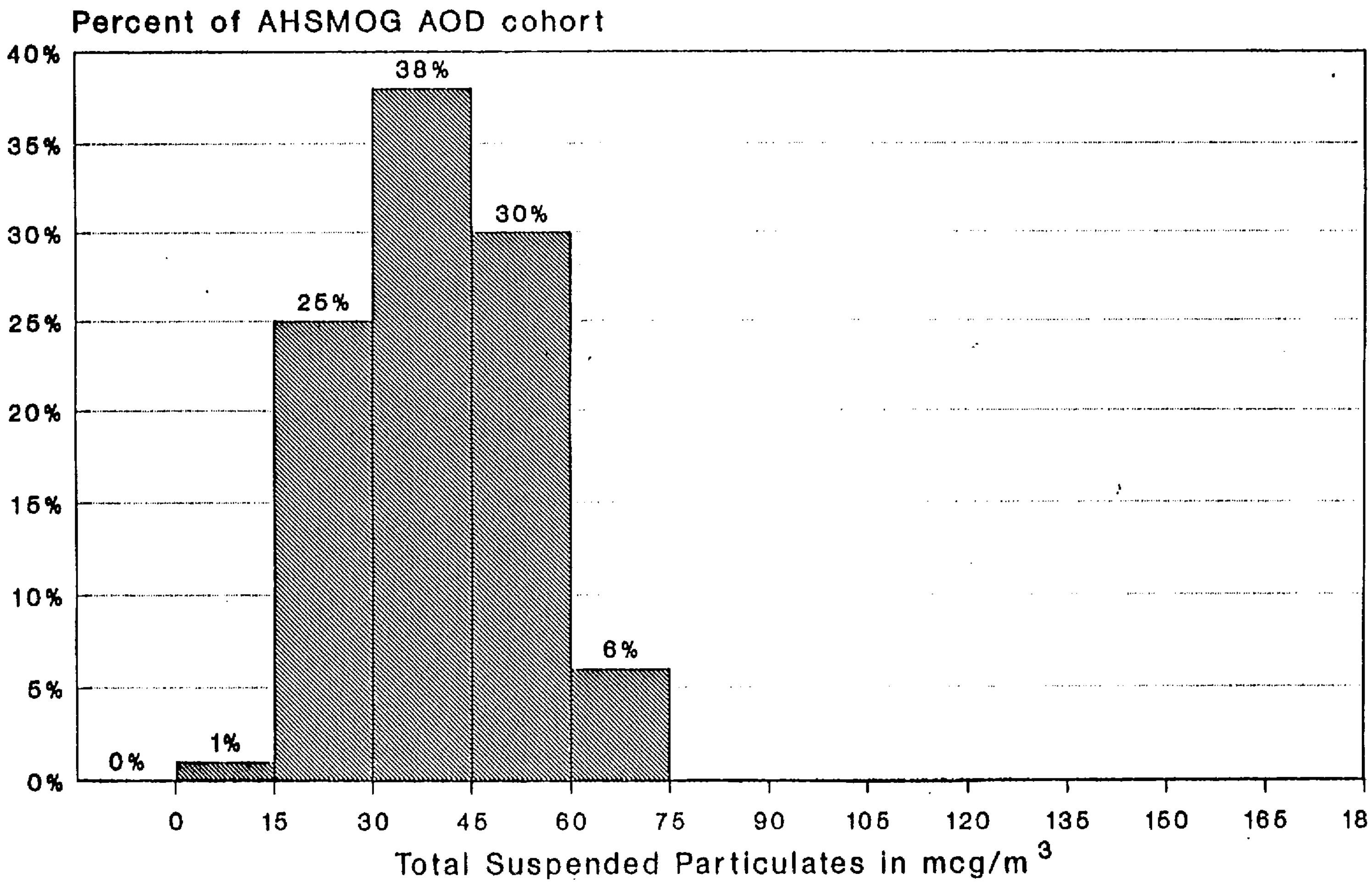
Appendix F. Figure 12. Mean Concentration of TSP in mcg/m^3
AHSMOG AOD cohort 1973-87
(n=3755)



Appendix F. Figure 13.

Adjusted Mean Concentration of TSP in mcg/m^3

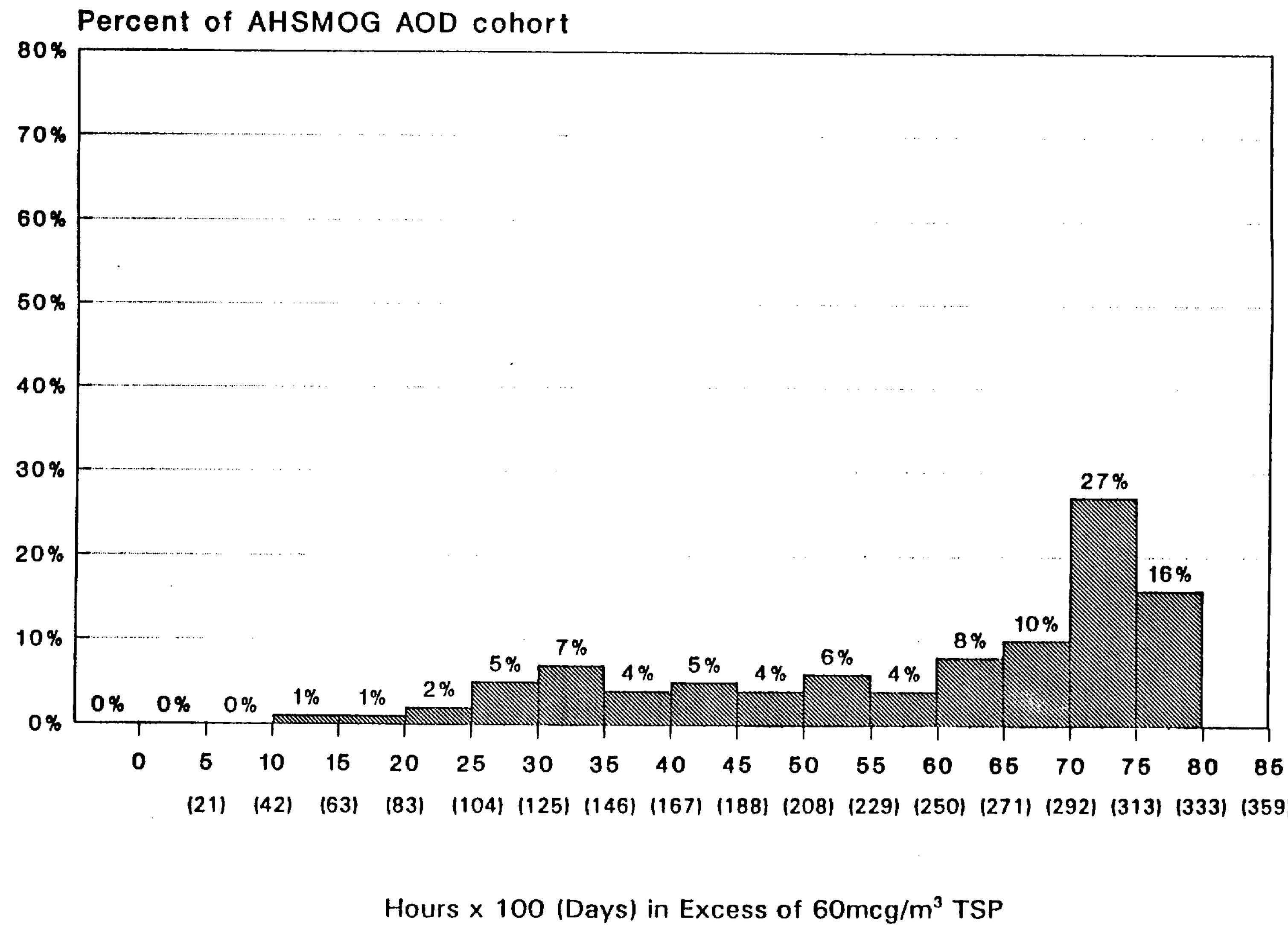
AHSMOG AOD cohort 1973-87
(n=3755)



Appendix F. Figure 14. Average Annual Hours of TSP > 60 mcg/m^3

AHSMOG AOD cohort 1973-87

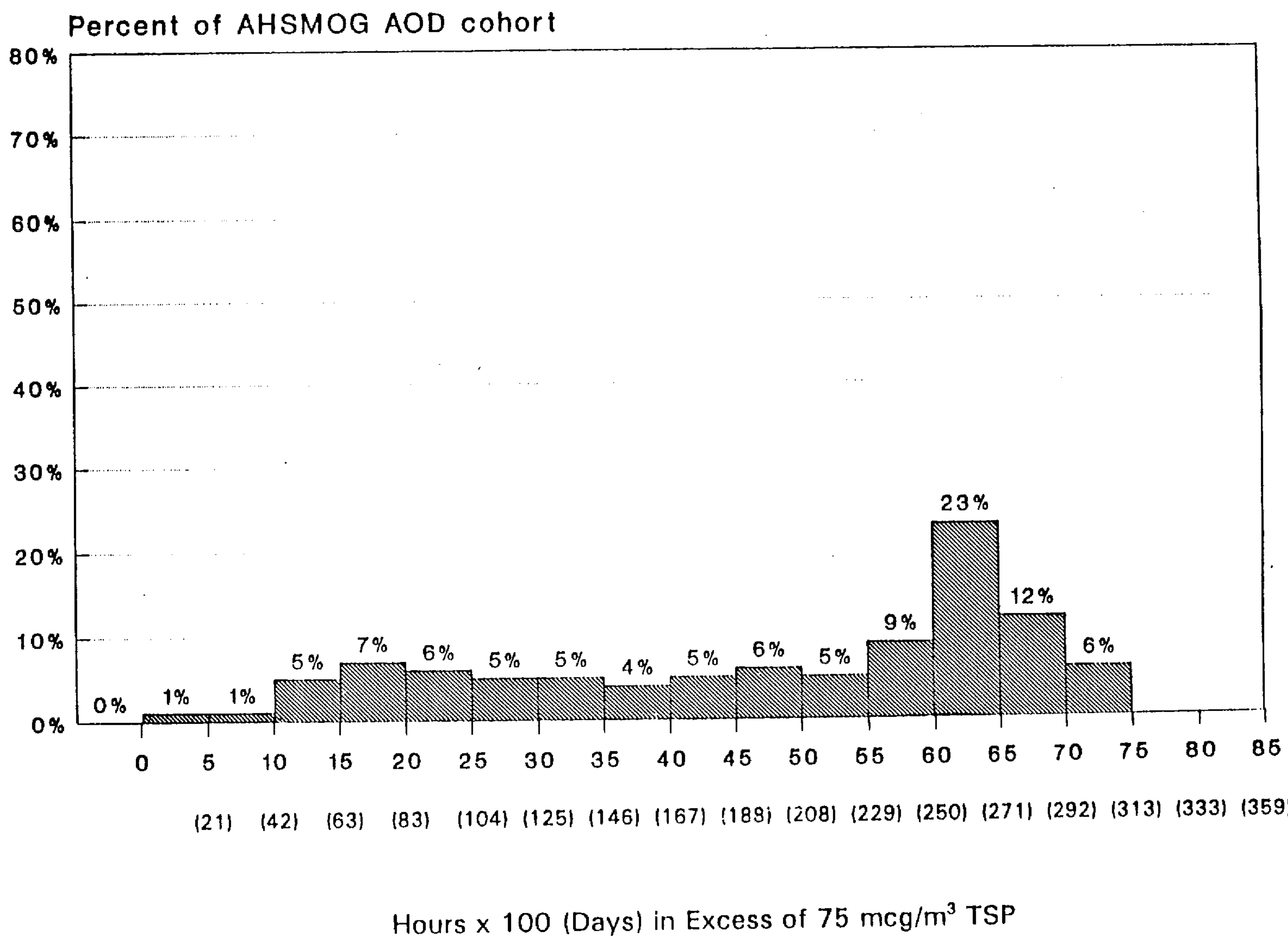
(n=3755)



Appendix F. Figure 15. Average Annual Hours of TSP > 75 mcg/m^3

AHSMOG AOD cohort 1973-87

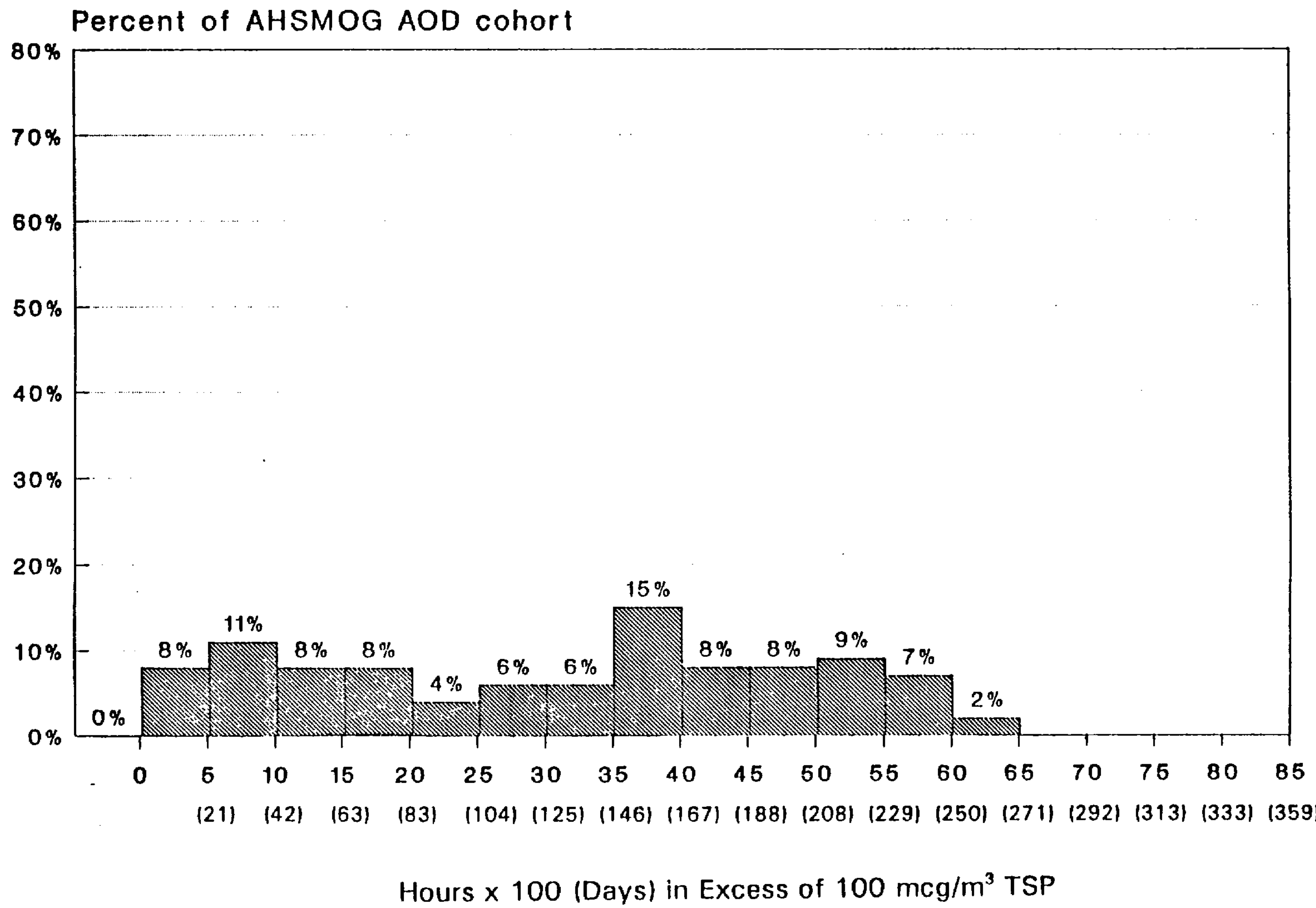
(n=3755)



Appendix F. Figure 16. Average Annual Hours of TSP > 100 mcg/m^3

AHSMOG AOD cohort 1973-87

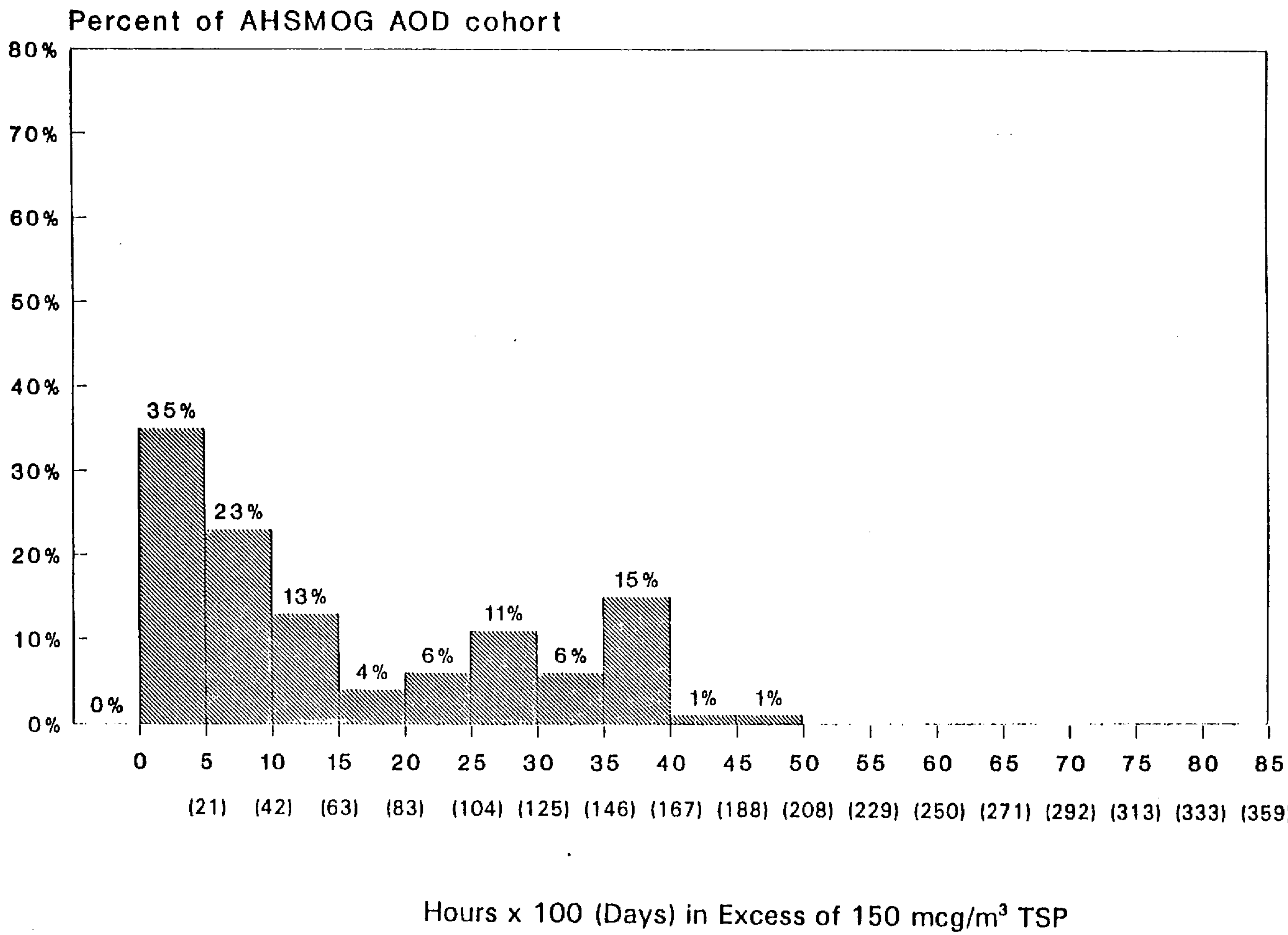
(n=3755)



Appendix F. Figure 17. Average Annual Hours of TSP > 150 mcg/m^3

AHSMOG AOD cohort 1973-87

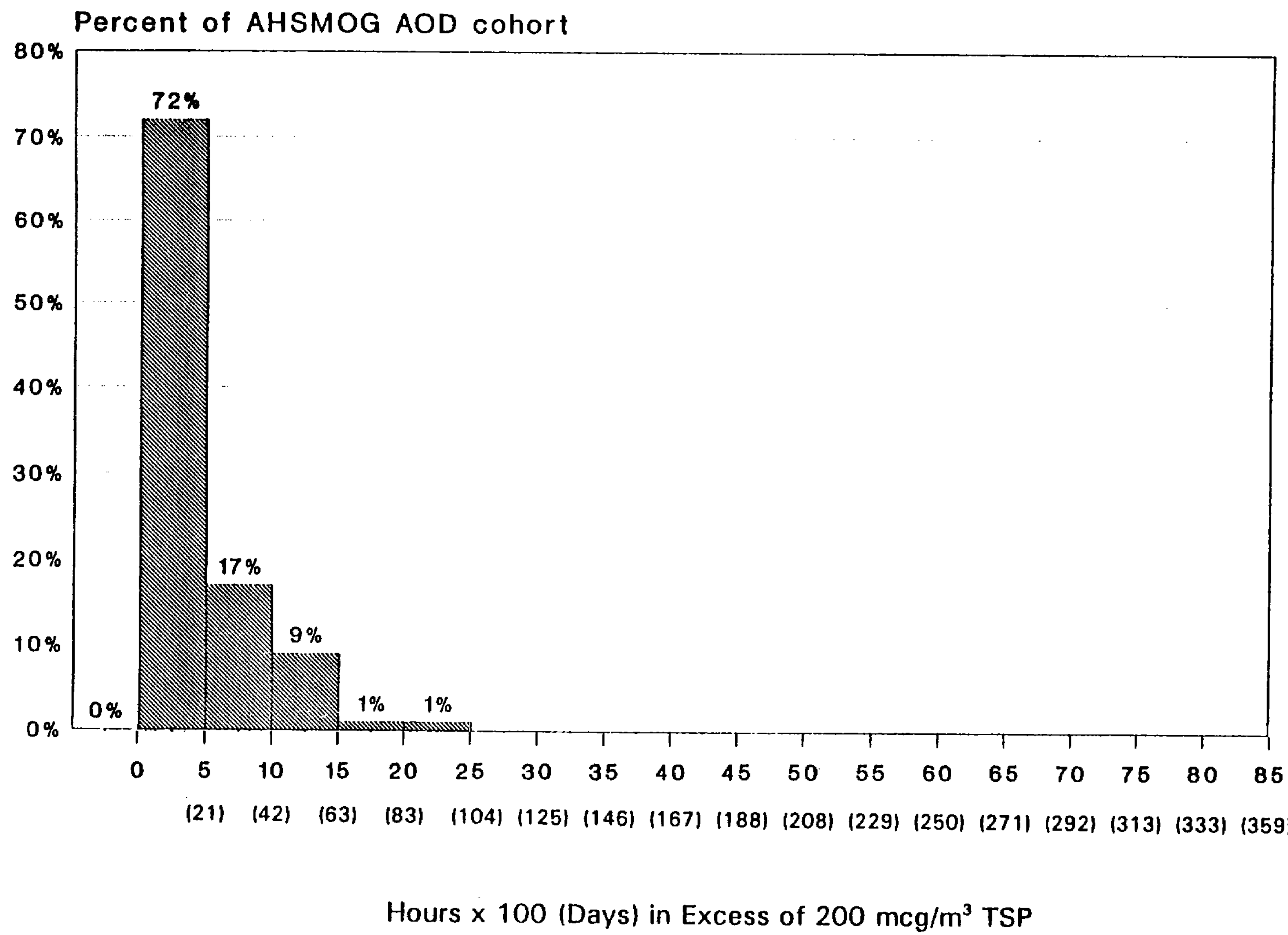
(n=3755)



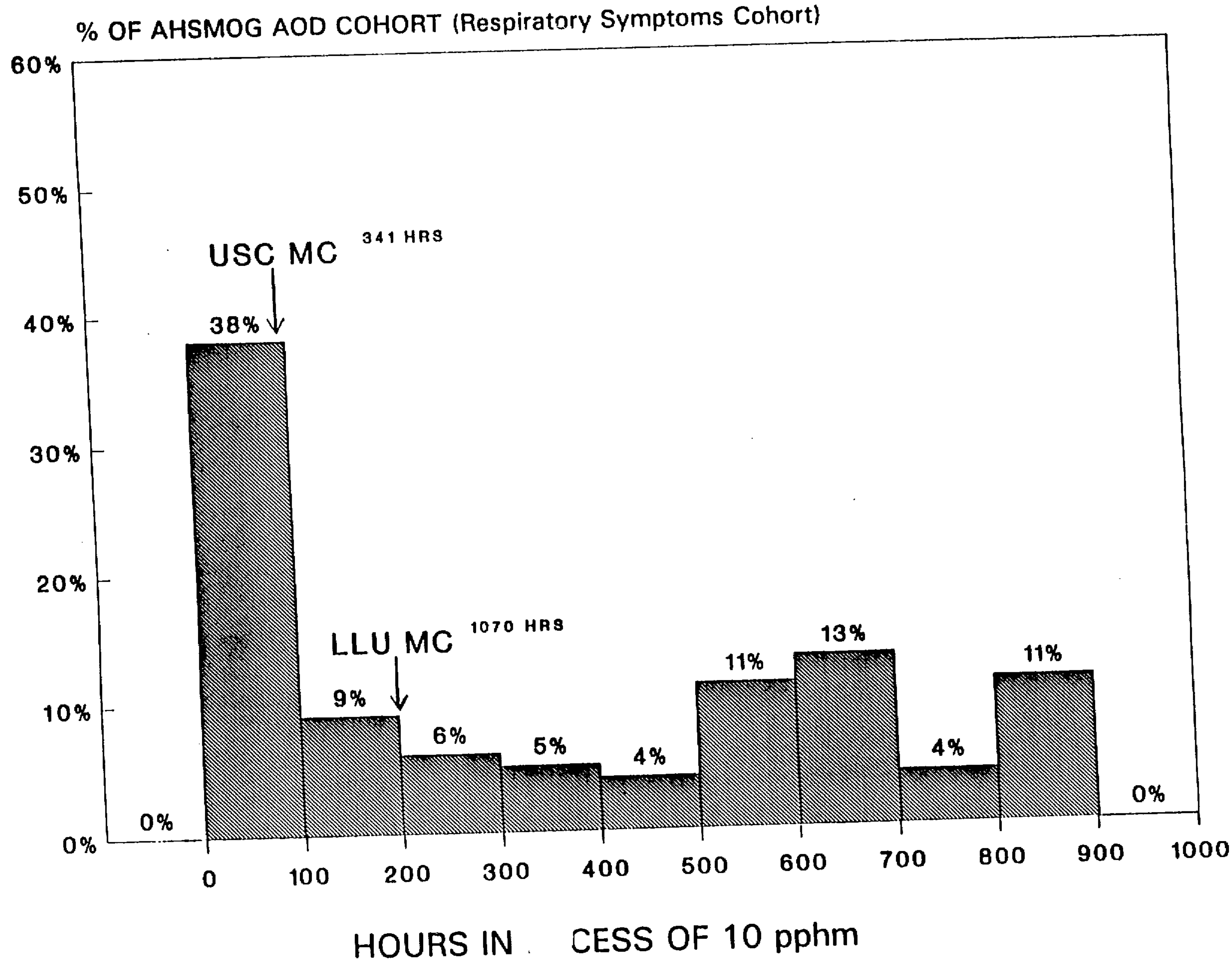
Appendix F. Figure 18. Average Annual Hours of TSP > 200 mcg/m^3

AHSMOG AOD cohort 1973-87

(n=3755)

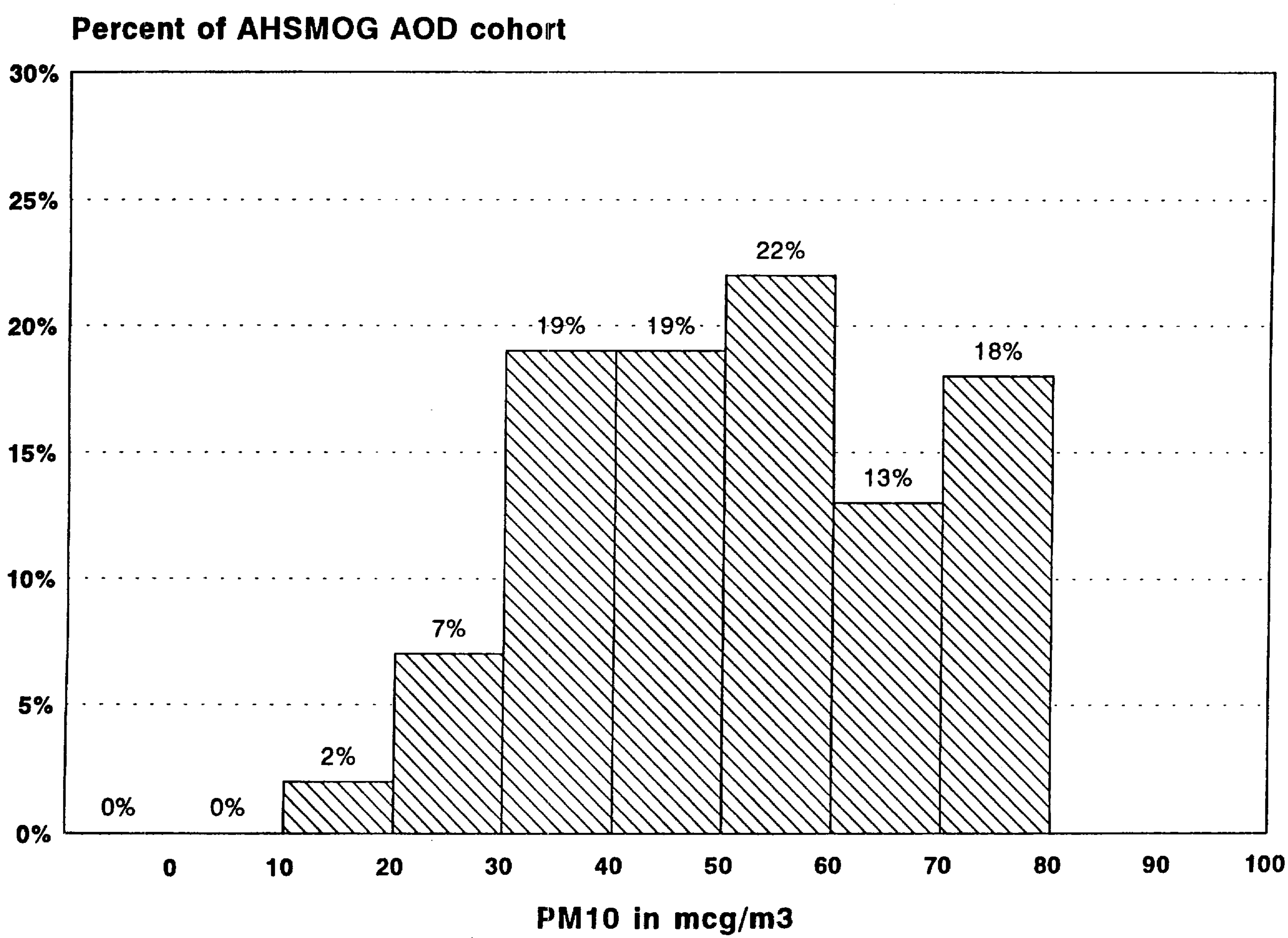


Appendix F. Figure 19. Distribution of Average Annual Hours (1973-1987) in Excess of 10 pphm Ozone for Members of the AHSMOG AOD Cohort (n=3707) Showing University of Southern California Medical Center (USC MC) and Loma Linda University Medical Center (LLUMC).



Appendix F. Fig. 20 Mean Concentration of PM10 in mcg/m3

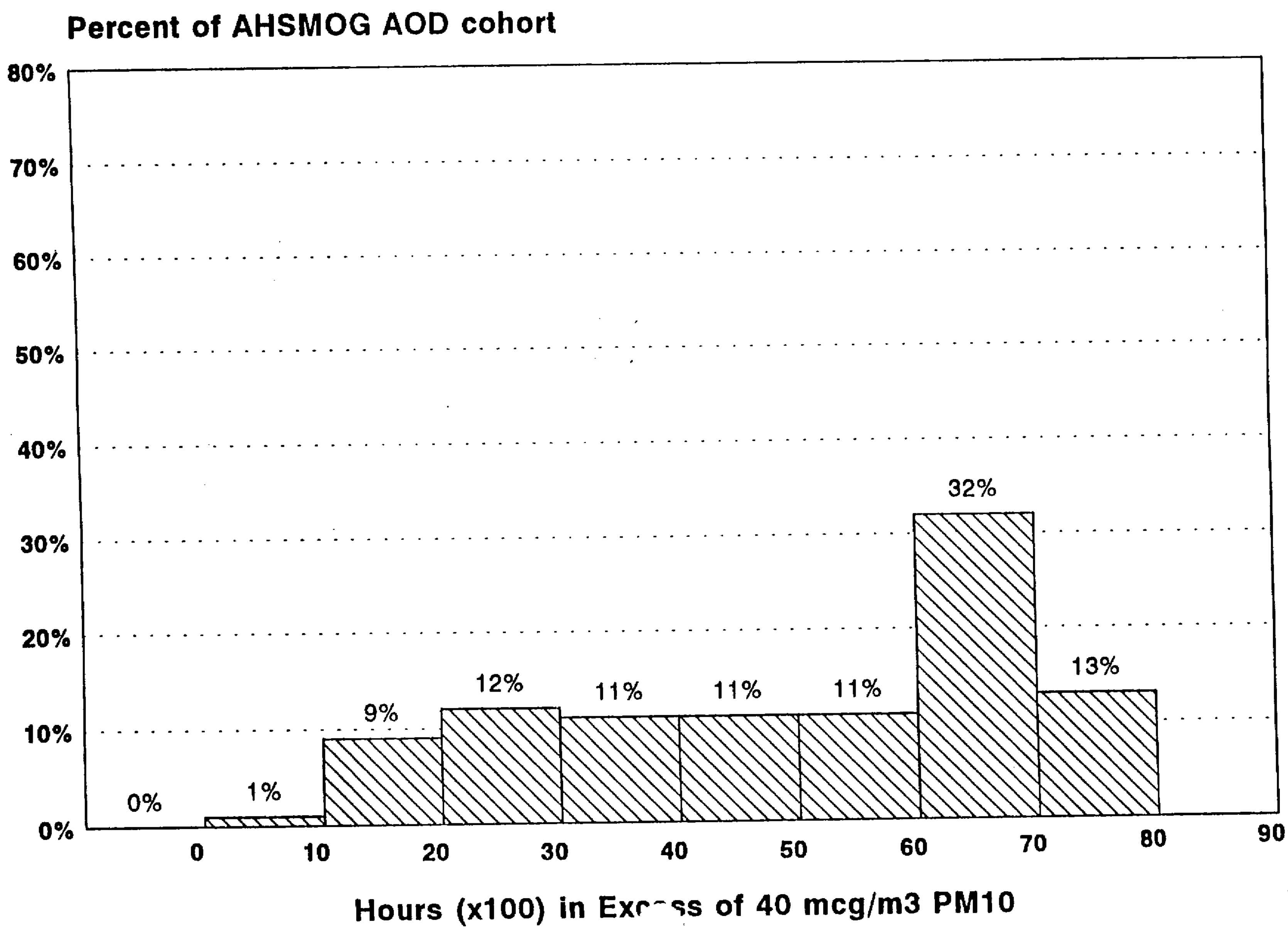
AHSMOG AOD cohort 1973-87
(n=3769)



Appendix F. Fig 21. Average Annual Hours of PM10 > 40 mcg/m3

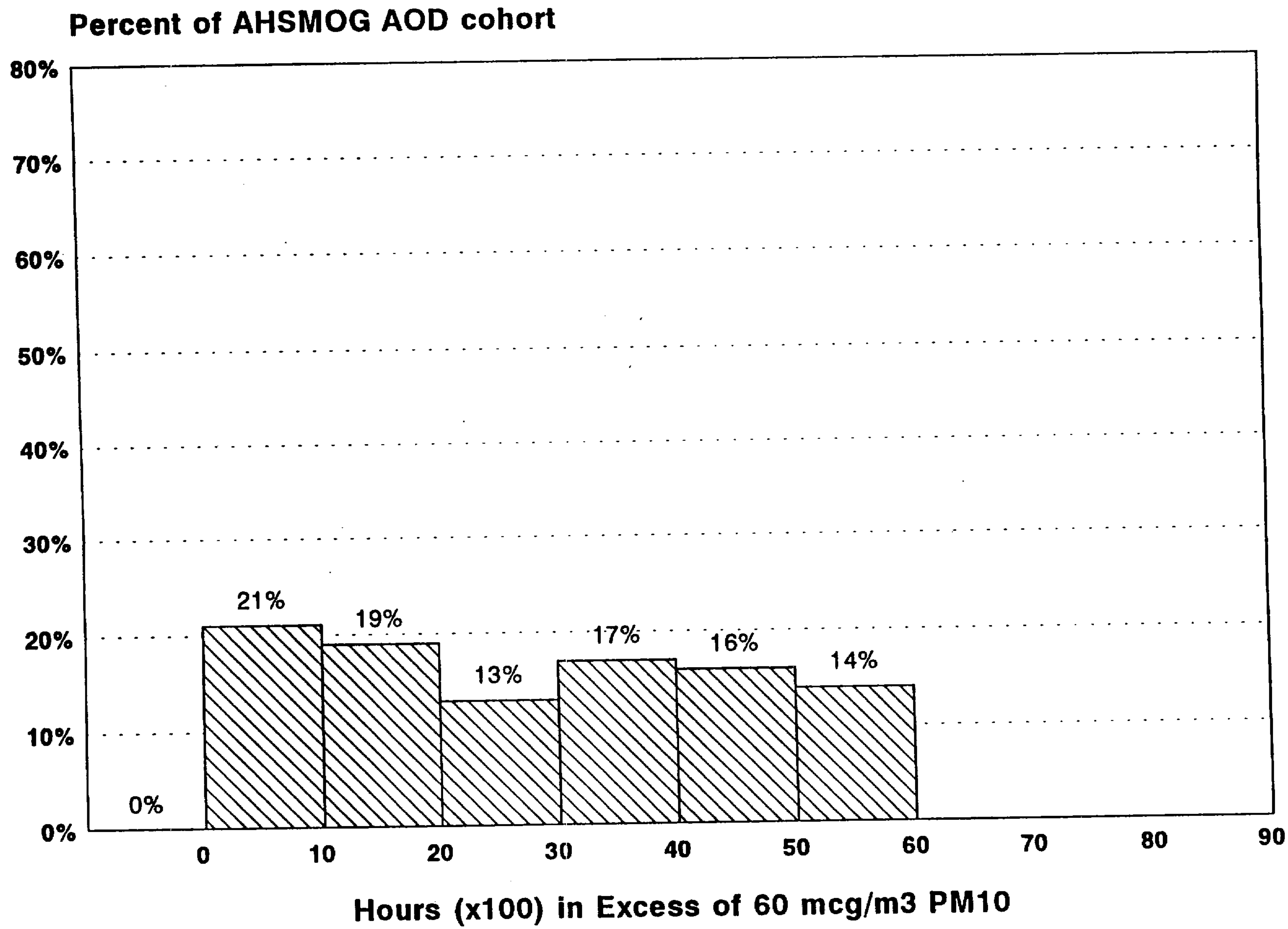
AHSMOG AOD cohort 1973-87

(n=3769)



Appendix F. Fig 22. Average Annual Hours of PM10 > 60 mcg/m3

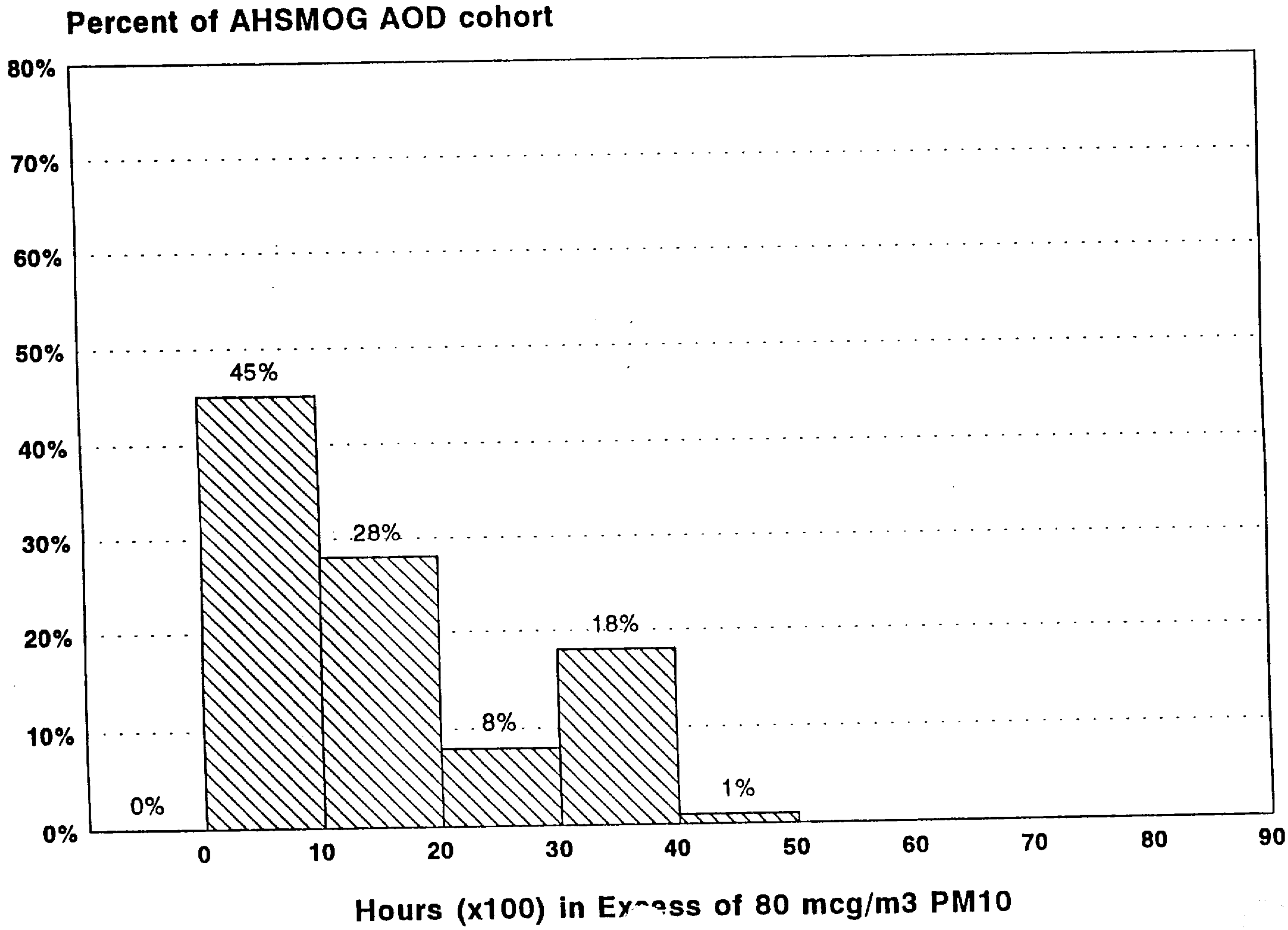
AHSMOG AOD cohort 1973-87
(n=3769)



Appendix F. Fig 23. Average Annual Hours of PM10 > 80 mcg/m3

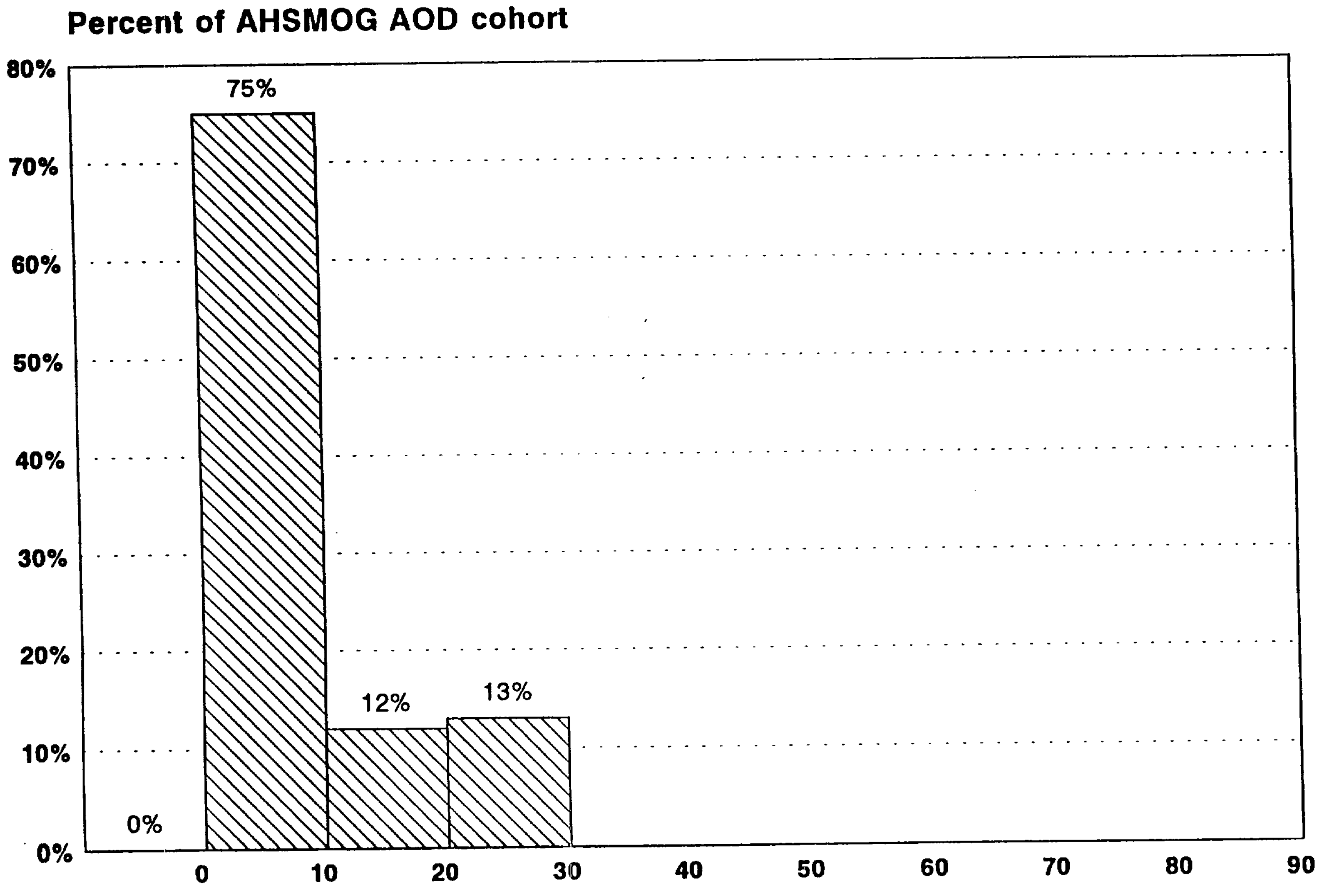
AHSMOG AOD cohort 1973-87

(n=3769)



Appendix F. Fig 24. Average Annual Hours of PM10 > 100 mcg/m

AHSMOG AOD cohort 1973-87
(n=3769)



APPENDIX G

Appendix G

Details of Ambient Air Pollution Calculations

§ Introduction

Calculation of ambient concentration estimates consisted of five tasks:

A. Calculating ambient concentration statistics at monitoring stations;
B. Interpolating ambient concentration statistics to zip code centroids;
C. Checking the consistency of the old (previously used, papers 1-4) and new (current, papers 5+) methods for calculating ambient concentration statistics;
D. Checking the validity of the interpolation method;
E. Cumulating the ambient concentration statistics of individual study members.

The first two tasks were performed by California Air Resources Board CARB staff; the latter three tasks were performed by Loma Linda University as soon as the data were received. The methodology used for each task is described below in sections A-E. First the time periods for which calculations were made are discussed.

The first phase of the AHSMOG study, referred to as the "old study", is described in section 2 of Chapter 1 of this report. The old study had calculated ambient concentrations for the time period 1966-1976.

It was necessary to calculate updated estimates of the study members' ambient concentrations of air pollutants for the period January 1977-March 1987. Ambient concentration estimates for 1973-76 were also recalculated because the data for those years were subsequently adjusted by CARB and because the previous and present ambient concentration calculations use slightly different methods.

There was no need to recalculate ambient concentration estimates for 1966-72, the earlier years of the old study, since the old study found that 1973-76 ambient concentrations were a satisfactory surrogate for ambient concentrations during the entire ten year period of the first study. This was established by re-running the final statistical models for each of the cancer, heart disease, mortality and respiratory symptoms outcomes using cumulations for the entire time period 1966-1976 instead of just 1973-1976 if the later time period was selected for the model by the stepwise procedures. In no case was the statistical significance of associations between ambient air pollutants and health outcomes changed. (Note in section E below a description of the alternate time periods which were allowed to be selected for the final model by stepwise procedures are described. The actual baseline time period used in more recent analyses was 1973 - March 1977, but for the old study (Papers 1 through 4) cumulations only went through 1976.) The good performance of the 1973-76 estimates as surrogates for 1966-76 ambient concentrations is not surprising, inasmuch as most study participants reported that they had not moved out of their neighborhood during that ten-year period.

§ A. Calculating Ambient Concentration Statistics at Monitoring Stations

The calculation of ambient concentration statistics at monitoring stations is described in the following subsections:

1. Selecting the set of monitoring stations to be used;
2. Extracting the best data for each combination of station, pollutant, and month from the CARB's air quality database;
3. Computation of ambient concentration indices at monitoring stations.

1. Selecting the set of monitoring stations to be used

There are a large number of monitoring stations (about 850) in the CARB data base. However, many of these stations have been operated only intermittently or have been terminated after a few years. The station or stations monitoring pollutants in a given locality are fairly frequently relocated from one site in the locality to another nearby site. The relocated stations are regarded as new stations and are assigned new identification numbers.

In the interest of consistency, an initial decision was made to include only stations which have at least 3 full years of data for at least one pollutant. A "full year" of data is defined as "enough data" (see section A.3 below for the criteria of "enough data") in at least ten months of the year to permit calculation of monthly exceedance statistics. It was later found necessary to relax this criterion, as explained below.

It should also be noted that only a small proportion of the monitoring stations monitored all of the five pollutants whose effects were investigated in this study.

Three or more full years of monitoring in some localities was performed by a sequence of stations, some of which were not in operation long enough to meet the three full years criterion. All the stations in such sequences were included in the set of stations.

Because monitoring stations were sparsely distributed in many parts of the state where members of the cohort had resided, CARB staff decided to include a number of stations which did not satisfy the three full years criterion and were not part of a sequence of stations that did. For the most part these stations had data for high pollution seasons for a number of years.

The set of stations selected according to these guidelines consisted of 348 of the approximately 850 stations with data in the CARB database. 126 of these stations were located in the three air basins with the most intensive monitoring -- the San Francisco, South Coast, and San Diego Air Basins, and 222 were in the remainder of the state. All the ambient concentrations required for the present study were estimated from the set of 348 stations selected according to the above criteria. We will refer to these set of 348 stations as the "1987 set."

Inspection of a tabulation of the data available from stations that were not used shows that either (1) the omitted stations have only temporary or limited-term monitoring or (2) the omitted stations are located more than 50 km distant from all members of the study population. The inspection also showed that a very large proportion of the available data for each pollutant was accounted for by the stations included in the set.

Geographic locations of the selected stations not previously located for the first study were verified and recorded on topographic maps, using the maps and descriptions of station locations in the CARB's records of California air quality monitoring stations.

Ambient concentrations in the San Francisco, South Coast, and San Diego Air Basins for 1973-76 previously estimated for the first study were re-estimated using the 1987 set of stations. The 1987 set and the previously used set contain almost the same stations in these three air basins. The only difference between the sets is the exclusion from the 1987 set of a few stations which had data for only a few months and were not part of a sequence of stations providing data for a longer period.

To simplify the generation of interpolated values, the stations to be used in interpolations of all pollutants to a given zip code were selected from the 1987 set and arranged in order of increasing distance, up to 50 km, from the zip code centroid. Hence, available data for all pollutants from the selected stations were considered for inclusion in the interpolations, although the station may not have satisfied the three full years of data criterion for some of the monitored pollutants.

2. <u>Extracting the best data for a station, pollutant, and month</u>

The first step in computing the exceedance statistics was to extract from a month's data for a station and pollutant the best set of concentrations for the month. Although one would expect there to be only one value for each monitoring period (hours for gaseous pollutants, days for particulate pollutants), there may in fact have been multiple monitors for the same pollutant at a site for all or part of the month.

Utilizing values from all the different monitors for a pollutant at a site may have resulted in a more complete set of data for the month than could have been obtained from any single monitor. In any event, given multiple values, it was necessary to choose the best among them.

These multiple values occurred for several reasons: (1) monitoring by more than one method at a given site, generally to compare the values obtained by old and by new and presumably improved methods; (2) duplicate monitoring by the same method at the same site to investigate variation between monitors; and (3) simultaneous short-term and regular monitoring at the same site -- an example was daily monitoring of TSP and SO_4 in Los Angeles during the summer together with the regular every-sixth-day monitoring of these pollutants. An example of monitoring producing multiple values for the same monitoring period was the monitoring of oxidant and ozone. Prior to 1976, total oxidants were measured at most CARB monitoring stations. Between 1976 and 1980 a transition was made to measuring ozone. During this transition period, as stations changed one by one, some stations monitored both pollutants for a short time. The current computation algorithm selected ozone concentrations where they were available and only used total oxidants concentrations if ozone concentrations were not available. Oxidant and ozone data are discussed further in section C.1 below.

All values in the air quality database are characterized by the project--such as routine monitoring, special purpose monitoring, etc.--and by the monitoring method that produced them. The project and method combinations for each pollutant were ranked in order of merit by CARB air quality monitoring staff. Generally, routine monitoring was ranked above special purpose monitoring and the newer and more accurate monitoring methods were ranked above the older methods. Where multiple values occurred for a monitoring period, the value associated with the highest-ranking combination of project and method for which values were available was selected as the best value.

3. Computation of ambient concentration indices at monitoring stations

The ambient concentration to air pollution whose effects are being studied are ambient concentrations to concentrations exceeding various cutpoints. Most of the cutpoint values were used in the previous study; some new ones have been added. The same cutpoints are used for total oxidants and ozone.

Ambient concentration indices are computed for a pollutant at a station for a month only if sufficient data existed. If there were sufficient data, but the data were not complete, the computed indices were adjusted to account for the missing data. These two topics are discussed in later parts of this section.

Two types of indices of cumulative monthly ambient concentrations exceeding cutpoints were calculated for both gaseous and particulate pollutants:

1. "exceedance frequency", the number of hours during which the concentration exceeded the cutpoint;

2. "excess concentration", the integrated excess of concentrations above the cutpoint.

"Excess concentration" is best defined by an example for a gaseous pollutant. Gaseous pollutants are monitored hourly. Suppose that ozone is monitored on a day during which concentrations are less than or equal to the cutpoint of 10 pphm for all but three hours and the values for those three hours are 12, 16, and 14 pphm, respectively. Then the excess concentration of ozone above the 10 pphm cutpoint for this day is

$$(12-10) + (16-10) + (14-10) = 12 \text{ pphm hours.}$$

The units of excess concentration are units of the pollutant times hours.

Two other indices of ambient concentrations are also calculated: 1) the total concentration, the sum of all concentrations monitored during the month, and 2) the average concentration for the month.

We now describe the computation of these two types of indices for particulate pollutants, which are monitored for 24-hour periods, usually on every sixth day. Both types of indices for particulate pollutants are expressed in units of hours for consistency with the indices for gaseous pollutants. Exceedance frequencies are computed by multiplying the percentage of observations for which a cutpoint was exceeded by the total number of hours in the month. Excess concentrations are computed in terms of pollutant units times days, by analogy with the preceding example, and then converted to pollutant units times hours by multiplying by 24.

Work location versus home location

Since both the work locations and residence locations of employed subjects are known, ambient concentrations for subjects for working hours and non-working hours can be utilized in the statistical models. Such ambient concentrations can be computed for ozone, NO_2, and SO_2, which are monitored hourly, but cannot be computed for TSP and SO_4, for which only daily values are available. The ambient concentration statistics at stations for oxidants, NO_2, and SO_2 are computed separately for working and non-working hours. For gaseous pollutants working hours are defined as 8 AM to 6 PM, Monday through Friday, with the exception of holidays established by Congress. Standard time was used throughout the year. For particulate pollutants working hours were defined as 9/24 of a day. Note: Papers published from this study state that working hours were defined as 8 AM to 5 PM rather than 6 PM. The discrepancy between what the CARB computer cumulation programmers did and what the authors thought they did, was not discovered until shortly before writing this report.

It is realized that applying fixed working hours and holidays to all working members of the study population involves a crude approximation. However, this is the best that can be done as individual data regarding specific hours for working, and which days were taken as holidays, were not obtained. Even if such data had been obtained it would not have been feasible to perform the extensive computations, allowing for each individual to have different working hours and holidays.

Determining if sufficient data exist for a pollutant and month

Ambient concentration indices were calculated for a combination of station, pollutant, and month only if sufficient data were available for the indices to be representative. For ozone, NO_2, and SO_2, which are monitored hourly, data must be available for 75% of the hours in the month. TSP and SO_4 are always monitored for 24-hour periods. Routine monitoring of these pollutants is performed every sixth day. Four 24-hour values (three for February) must be available from the month for particulate data to be considered representative.

Adjusting for missing data

Almost always, data are not available for all the hours or days of a month. Monitoring stations for particulate pollutants are very seldom operated for all the days of a month. Monitoring stations for gaseous pollutants must be briefly shut down every day for calibration and maintenance. The computed exceedance frequencies, excess concentrations, and total concentrations were adjusted to account for missing data by multiplication by the appropriate ratio -- the ratio of total hours in the month to the number of hours in the month with data.

§ B. Interpolating Ambient Concentration Statistics to Zip Code Centroids

Interpolation of ambient concentration statistics to zip codes is described in the following subsections:

1. Determination of zip code centroids;
2. Selection of stations for interpolation for each zip code;
3. Interpolation to zip code centroids;
4. Assignment of ambient concentration for zip codes with no nearby stations.

1. Determination of zip code centroids

Two sources of zip code centroids were available: centroids determined by CARB staff for the first project ("CARB centroids") and geographic centroids corresponding to January, 1987, zip code boundaries supplied by Geographic Data Technology, Inc. ("GDT centroids"). Information was also obtained from 1977 and 1986 zip code maps published by Western Economic Research Co.

CARB centroids are approximate population centroids determined by eye from 1977 zip code boundary maps in conjunction with population density maps from the 1970 census.

Many zip code areas have been changed since 1977. Boundaries have changed, many new zip codes have been created, and a few have been deleted. Hence, it was necessary to verify that CARB centroids were also sufficiently accurate centroids for the years of the current study. Boundary changes were detected by comparison of the 1977 and 1986 zip code maps and probable centroid changes judged large enough to affect interpolated values significantly were noted, as were zip code areas created since 1977.

The existence of both CARB and GDT centroids for zip codes in the first study facilitated checking CARB centroids for implausible locations. All zip codes with deviations between CARB and GDT centroids of greater than four miles were visually checked on the maps and appropriate corrections were made.

After all the above verifications had been carried out, coordinates were assigned to zip code centroids by two sets of rules, one for the 1973-76 period and another for the 1977-87 period. These rules are depicted by tree diagrams in Figures 1 and 2. These rules give preference to CARB coordinates if they exist and are deemed to be accurate, since CARB centroids are presumably closer to the actual population centroids.

Not all zip codes for which interpolations from monitoring stations are made represent geographic areas. However some zips are assigned by the post office to post office boxes, business firms, etc. Almost all of these zip codes were assigned the coordinates of the geographic zip code areas within which they are enclosed. The only exceptions to this rule were those few enclosed zip codes for which valid CARB coordinates different from those of the enclosing zip code had been determined.

2. Selection of stations used for interpolation for each zip code

Each zip code in which a study participant had resided or worked for one month or more was assigned a set of stations to provide data for the interpolation of ambient concentration statistics to the zip code. This set of stations was intended to be large enough that all the data that would be used in all the interpolations--for every combination of pollutant and month for all 15 years--could be obtained from the stations in the set. We will refer to this set of enough "relevant" stations for the zip code as the "superset" for the zip code, to distinguish it from the sets of no more than three stations whose data were actually used in individual interpolations.

The selection of the superset is described in the first major subsection of section 2 and summarized in Figure 3. The selection of stations from the superset to provide data for an individual interpolation is described in the second major subsection and summarized in Figure 4.

Selection of the superset

The stations in the superset for a zip code satisfy the following conditions: 1) they are within 50 km of the zip code centroid; and 2) they are not on the opposite side of a significant topographic barrier to airflow. Concentrations at stations on the opposite side of a significant barrier are considered not to be relevant to concentrations in the zip code. Supersets contain no more than 12 of the nearest stations satisfying these two conditions.

Exclusion of stations on the opposite sides of barriers to airflow

We next discuss the procedures used to determine whether zip codes and stations are close enough together and are located on opposite sides of a significant barrier to airflow. The first step was a computerized check of the stations and zip codes against barriers previously used in air quality modeling computations in the San Francisco Bay area, the South Coast Air Basin, and some adjacent areas. These barriers were determined for the purposes of air quality modeling on a regional scale, rather than for approximation of local air quality, and hence are fairly crude approximations to the actual physical barriers. Each of these approximate barriers is a straight line or several connected straight lines whose end points were digitized.

The computer program which located the stations within 50 km and performed the checking against the barriers was instructed to output only the nearest 12 stations satisfying the two conditions imposed on stations in the superset. Admittedly, the number of potential members of the superset having data for at least one of the five pollutants for at least one month during the 15 years could be quite large, but the interpolations were to use data from no more than three stations. Choosing to output no more than 12 stations was an attempt to balance between the conflicting goals of: a) not including enough stations to supply all the data that would be used in any of the interpolations, and b) limiting the time and effort required to check the very large number of long station lists.

The second step was checking the zip codes in the rest of the state to determine if they were separated from stations in their lists by barriers to airflow. Significant barriers to air flow in regions other than the San Francisco or South Coast Air Basins were determined with the assistance of Thomas P. Hayes (1984), CARB meteorologist and the primary author of reference 1, who sketched the barriers on a 1:500,000 topographical map of California. The locations of zip codes where sample members had resided had previously been plotted on the map so that the meteorologist could take into account all the locations for which significant barriers had to be determined. These barriers have not been computerized so the stations to be excluded for zip codes in these regions had to be determined by manual comparison of the location of the zip code, the locations of stations within 50 km, and the sketched barriers to air flow.

All of the computer generated station lists for zip codes in areas where computerized barriers were used were also manually checked. In a few instances the approximate straight line computerized barriers inappropriately separated zip codes from stations; in other instances they failed to separate irrelevant stations from zip codes. Lists were made including all stations within 50 km of such zip codes, irrespective of barriers, and these lists were manually edited using topographic maps to more accurately define the true location of the air flow barriers.

After applying these exclusion criteria, the resultant list of up to 12 stations for each zip code was ordered from closest to farthest from the zip code centroid. The stations in this superset of stations for each zip code will be referred to as "relevant" stations. Only one set of relevant stations was determined for the entire time period, 1973-March, 1987. Stations other than relevant stations were never used for interpolation. If none of the relevant stations had data for a particular pollutant for a given month, a missing value code was assigned to the zip code for that pollutant and month.

Selection of the best one to three stations for each zip code, pollutant, and month

There are up to 12 relevant stations for each zip code, but no more than three of the nearest ones with available data for the pollutant and month are used for any interpolation, so that the influence of the closer stations is not diluted by data from more distant stations. How many of the three nearest stations with data are used in each interpolation is determined by a rule based upon the distances from monitoring stations over which the EPA (3) considers monitored concentrations of each pollutant to be of good ("A") quality, fair ("B") quality or marginally representative ("C") quality. These distances are given in Table 2.2 of the main report. The distances for the primary pollutants NO_2, SO_2, and TSP are smaller than the corresponding distances for OX and SO_4, which are formed by chemical processes in the atmosphere during transport of the primary pollutants from which they are derived. The quality rating for OX is also used for ozone. Note that the quality rating applies to the quality of the data in the context of the interpolation and not to the quality of the monitored data itself.

The rule specifies that only stations with data of "A" and "B" quality for the zip code are to be used, if data are available at one or more stations with data of those qualities. Consequently, data of "C" quality are to be used only if no data of "A" or "B" quality are available. A flow chart for this rule is given in Figure 4.

3. Interpolation to zip code centroids

The values of monthly ambient concentrations statistics at zip code centroids are computed by spatial interpolation of values at the three or fewer monitoring stations selected by the procedure indicated above. The values at each monitoring station are weighted by $1/R^2$, where R is the distance between the zip code centroid and the monitoring station. The ambient concentration statistics for each pollutant for each zip code and month are assigned a quality index reflecting the quality of the data used in the interpolations. A quality rating of "A" is assigned if the nearest station was within the "A" quality interpolation distance, a quality rating of "B" if the nearest station was within the "B" range, and a quality rating of "C" if the nearest station was within the "C" range. Note that the assignment of quality ratings allows for sensitivity analyses to be conducted to determine the effects on results of including data with lower quality ratings in estimates of individual ambient concentration.

4. Assignment of ambient concentrations for zip codes with no relevant stations

For some zip codes, no relevant stations were selected by the procedure outlined above. For many California zip codes, there are no monitoring stations located on the same side of an airflow barrier and less than 50 km away. Some of these zip codes are at a great distance from any significant sources of emissions or are up-wind from all significant sources. For example, small communities on the coast are up-wind from any significant sources. These zip codes are judged to have no significant concentrations of any pollutant, and "background" levels of zero for all pollutants are assigned to them. Other zip codes that are distant from significant sources of most pollutants have a monitoring station within 50 km, but that monitoring station does not measure concentrations of all pollutants. Usually, these stations measure only TSP. For these zip codes, background levels were assigned for all pollutants except those measured. All interpolated exceedance statistics set equal to background values are assigned the quality code I.

Other zip codes are more than 50 km from relevant monitoring stations, but in the judgment of CARB staff, pollution from urban areas is transported to them. These zip codes were assigned the average concentrations of pollutants measured at an "assigned set" of stations selected by the CARB staff in the respective up-wind urban areas. All interpolated exceedance statistics set equal to the average values at an assigned set of stations are assigned the quality code H.

§ C. Checking the Consistency of the Old and New Methods for Calculating Ambient Concentration Statistics

1. Comparison of the ambient concentration calculations for the two studies

The first study calculated ambient concentrations only for zip codes in three air basins -- San Francisco, South Coast, and San Diego -- in which about 85% of the study members were residing. These air basins, which will frequently be referred to as "the three air basins" for brevity, have the most intensive air quality monitoring and, presumably, mostly higher pollution levels. The first study assumed that the 862 study members residing outside the three air basins had negligible levels of ambient air pollution and were assigned a value of zero for all pollutants.

In contrast, the current ambient concentration calculation effort calculated ambient concentration estimates for all zip codes in which study members resided for the period January 1973 to March 1987, which will be referred to as "1973-1987" for brevity. Calculating ambient concentration for all zip codes is necessary because some study members have moved outside the three air basins since 1977. Ambient concentration outside the three air basins for 1973-1976 are also being calculated for consistency. Estimates of ambient concentration of SO_4 were not calculated for 1976 and earlier years because insufficient SO_4 monitoring data were available before 1977.

Newly calculated values of NO_2 ambient concentration statistics for 1973-76 may differ significantly from the values calculated for the previous study. After the values for the previous study had been calculated, the CARB adjusted the monitored NO_2 concentrations reported by various air quality districts to compensate for differences in monitoring and calibration methods. The adjustment multiplied the reported concentrations by 0.88.

Values of the same ambient concentration statistics at the same monitoring stations calculated for the previous study and the present study may differ slightly even if the air monitoring data involved have not been adjusted. The possible explanations for these discrepancies are: the interpolation algorithms used in the two calculations differ slightly, the sets of stations used may be different, and the sets of concentrations from which the monthly statistics are calculated may be different. A complete description of the previous calculation is no longer available. However, comparisons of some data sets for which the largest differences between old and new values occur suggest that the old calculations did not make an effort to use all the available data. It follows that the interpolated values calculated from these ambient concentration statistics differ slightly. Comparisons of corresponding ambient concentration statistics calculated for the two studies were made, as discussed in the next section.

Monitored values of TSP and SO_4 before 1982 are not consistent with values for 1982 and after because of a change in the alkalinity of the filter material on which the particulate samples were collected. Note that this inconsistency in the TSP and SO_4 data is an inconsistency within data for the new study as well as an inconsistency with TSP data for the previous study.

During 1976, the CARB began to switch from measuring total oxidants to measuring ozone. This switch had been completed by the monitoring stations by 1979. As this change was implemented at different monitoring stations, measurement of total oxidants was generally discontinued, though some stations continued to measure both for a short time. In the case where both total oxidants and ozone were measured, the present calculations used ozone data whenever they were available. Total oxidants were used when ozone was not available. The previous calculations for 1966-1972 and 1973-1976 used total oxidants; thus this is another source of inconsistency in the data. When cumulative ambient concentration to ozone is added over the years 1973 to 1987, it will consist primarily of total oxidant data prior to 1979 and primarily ozone data after that date. The CARB has conducted some studies of how hourly values of total oxidants agree with ozone. As a general rule, either pollutant can be substituted for the other but there are sometimes large differences between the two measures. No constant adjustment factor has been estimated which can be used to convert total oxidant measures to ozone measures. Thus, for computing long term cumulative ambient concentrations for this study for time periods extending back past 1979, we will have to add total oxidants and ozone together.

2. Checking the consistency of the old and new ambient concentration statistics

The old method for generating ambient concentration statistics developed by CARB programmer, Paul Allen, and used for the analyses of the 1977 data was compared to the new computer generated statistics for the time period 1973-1976 for all stations for which new data were generated in the South Coast, San Diego, and San Francisco air basins. The results of these comparisons will first be summarized by specific air basin for oxidants and TSP. A summary of comparisons for each pollutant will be made and the conclusions stated.

COMPARISON OF OXIDANTS

South Coast Air Basin

There were 1,961 station month observations for the time period 1973-1976. Mean values over these 1,961 observations were compared for the two methods using a paired t-test. The results are displayed in Table 1. Old means are always larger than new means, but never more than 12% larger. Differences between the old and new methods are highly statistically significant with p-values generally being less than .001 and always less than .01. This is due to the very large sample size and small standard deviations of differences between paired values. The statistical significance of these paired t-tests here and in the remainder of this section have little practical importance. It is the magnitude of the difference in the means which is important. Correlation coefficients were computed between the old and new methods. All correlation coefficients were in excess of 0.995. Scattergrams were made plotting individual values, old versus new, against each other. The resulting data ellipses were very tight about a straight line with no outlying values.

All individual relative differences between paired old and new statistics with an absolute value larger than 20% of the new value were manually inspected. For the South Coast Air Basin, only two observations of total ambient concentration had relative differences greater than 20%. For one the old value was 23% larger than the new value, for the other the old value was 32% smaller. For hours in excess of 10 pphm, there were 431 observations which differed by more than 20%. None of these differences were consequential, since all occurred for very small values. For example, the largest difference was 267%, but this occurred for very small values, 11 hours for the old value and 3 hours for the new value. For most discrepancies the old value was larger than the new value. For excess dosage above 10 pphm, there were 84 observations with discrepancies greater than 20%. Here again none of the differences were consequential since all occurred for small values. The largest differed by 160%. This again was for two relatively small values, one with a value of 5 pphm hours for the new method and 13 pphm hours for the old method. Although the majority of differences were positive, there were a number of negative differences indicating that the old method was smaller than the new method for some cases.

All individual discrepancies which differed by more than 20% for the other ambient concentration statistics for oxidants/ozone were also manually inspected. Percentage differences were not larger than those described above and the larger differences always occurred where the new values were relatively small. Thus, the practical consequence of such differences is not large. The new method used ozone values when they were available, whereas the old method used oxidant values. Towards the latter part of 1976, ozone began to be monitored. This could explain some of the discrepancies.

San Diego Air Basin

There were 257 station months for San Diego. Again, paired t-tests were conducted comparing old versus new averages. Although most old means were larger than new means, this was not always the case. The means agreed closely in value with the largest difference occurring for hours in excess of 10 pphm where the old mean was 2% larger than the new mean. Some means differed significantly and some did not. See Table 1. All correlations were in excess of 0.990. Scattergrams were plotted for all ambient concentration statistics of the old versus new methods, with the exception of ambient concentration statistics in excess of 25 pphm where there were only four non-zero values. Very tight ellipses about a straight line were obtained with no outliers. All individual differences greater than 20% were manually inspected. No differences occurred for total dosage that were larger than 20%. The differences between the old and new methods for the other ambient concentration statistics are comparable to those described above for the South Coast Air Basin.

San Francisco Air Basin

There were 1,059 station months for the San Francisco Air Basin. Comparison of mean values is shown in Table 1. Means of the two methods give close agreement with the largest difference occurring for hours in excess of 10 pphm, where the old method give a mean value 21% larger than the new method. Some means were statistically significantly different and some were not. Ambient concentration statistics for cutpoints higher than 15 pphm did not have sufficient non-zero values to warrant calculation of statistics. Correlations for exceedance statistics up to 15 pphm were all larger than 0.980. Scattergrams were formed, plotting the old method versus the new method for the exceedance statistics up to 15 pphm. All gave tight ellipses about a straight line with no outliers with the exception of excess concentration for 15 pphm. It had one outlying point with an old value of 7.5 pphm-hours and a new value of 19 pphm-hours. All differences greater than 20% were manually inspected, with results comparable to those described for the South Coast Air Basin.

Summary for Oxidants

Although the two methods give close agreement in an aggregate sense and correlate highly, there are some individual station months for which fairly large discrepancies occur. Nearly all of these large discrepancies occur for relatively small values of the statistics, so they are of minor consequence.

COMPARISON OF TOTAL SUSPENDED PARTICULATES

South Coast Air Basin

There were 1,292 station months for total suspended particulates (TSP) for the South Coast Air Basin. Paired t-tests were conducted to compare the average values for the two methods. The means for all ambient concentration statistics agreed very closely with differences being much less than 1%. All computed correlation coefficients had values in excess of 0.99. For only one ambient concentration statistic, excess concentration above 150 $\mu g/m^3$, was there a statistically significant difference ($p < .001$); but the magnitude of this difference was much less than 1%. Scattergrams were plotted to compare the two methods. All gave virtually straight lines with no outlying values. All individual differences more than 20% were manually inspected. There was one such case for hours in excess of 100 $\mu g/m^3$ with a value of 372 hours for the old data and 298 hours for the new data. There were two such cases for exceedance concentration above 100 $\mu g/m^3$. One case had a value of 9,486 pphm for the old data and 7,589 pphm hours for the new data. The other had a missing value for the old data and a value of 5,828 pphm hours for the new data. This same discrepant case occurred for the cutpoint of 60 $\mu g/m^3$ also. This had a value of 0 for the old data and a value of 19,220 pphm hours for the new data. There were two cases where discrepancies for exceedance frequency exceeded 20% for the cutpoint of 200 $\mu g/m^3$. These two cases had missing values for the old data although excess concentration was not missing indicating an erroneous omission of exceedance frequency from the old data.

San Diego Air Basin

There were 247 station months for San Diego. All mean values differed by less than 1%. None were statistically significantly different except for dosage in excess of 150 $\mu g/m^3$. Statistics in excess of 200 $\mu g/m^3$ were identical within rounding error. All computed correlation values were 1.000. Scattergrams indicated virtually straight lines with no outliers for all ambient concentration statistics. There were no differences in excess of 20%.

San Francisco Air Basin

There were 1,012 station months for the San Francisco Air Basin. Paired t-tests were conducted. All means differed by less than 1%. None were statistically significantly different. All correlation values exceeded 0.994. Scattergrams indicated tight ellipses about straight lines with no outliers. Differences greater than 20% were manually inspected and fall within the range found in the South Coast Air Basin with the following exceptions. One station month had a value of 1,800 pphm hours for excess concentration above 100 $\mu g/m^3$ for the old method and 288 pphm hours for the new method. One other station month which had a value of 16,306 pphm hours excess concentration above 150 $\mu g/m^3$ for the old method and 3,720 pphm hours for the new method. For the cutpoint of 200 $\mu g/m^3$ there were two cases where discrepancies exceeded 20%; for one the old value for exceedance concentration was 65 hours and the new value was 120 hours. This same case caused a discrepancy for excess concentration. The other case had a value for excess concentration of 186 hours for the old data and 0 for the new. This same case causes a discrepancy for excess concentration.

Summary for TSP

The two methods gave very close agreement for TSP with the exception of the few individual station months in the South Coast and San Francisco Air Basins noted above.

Station-months for which only one method generated statistics

The new method and old method did not generate exceedance statistics for exactly the same set of combinations of pollutant and month. There are combinations for which only the old method generated values and combinations for which only the new method did. These discrepancies may be due to differences in the data the two methods assembled for the pollutant and month or to differences in the rules they used to determine if there were sufficient data. The discrepancies are tabulated below.

For gaseous pollutants, the new method used the same 75% of the hours sufficient-data rule that the old method is supposed to have used. However, an examination of the data showed that there was insufficient data present for almost all the discrepant combinations for which the new method did not generate statistics. The most plausible explanation is that the old method did not always apply the 75% of the hours rule.

For particulate pollutants, the new method did compute statistics for Februarys with only three values, instead of requiring that there be four values. This difference in the rules accounts for most of the discrepancies for TSP.

Numbers of Station-Months, 1973-76, for which only One of the Two Methods Calculated Statistics

	TSP		OX	
Air Basin	**Old Missing New Present**	**Old Present New Missing**	**Old Missing New Present**	**Old Present New Missing**
South Coast	81	0	4	70
San Diego	10	0	0	14
San Francisco	13	1	0	23

CONCLUSIONS

The primary purpose of the present comparisons was to determine if the two methods were in general agreement. From the results above we conclude that they were in general agreement for oxidants and TSP. The two methods were not compared for the other pollutants. Since the computer algorithm used for oxidants does not differ from those used for the other gaseous pollutants, and likewise the method used for TSP did not differ from the method used for SO_4, the other particulate pollutant, we inferred that the two methods would give general agreement for all pollutants.

§ D. Checking the Validity of the Interpolation Method

Assessment of precision (validation) of the interpolation methods is addressed in section 3 of chapter 2 of this final report.

§ E. Computer Programming for Ambient Concentration Data

§ 1. Overview

A two phase computer program to cumulate ambient concentrations for each study participant was written. The first phase of the program prepares the data files received from the California Air Resources Board (CARB) so that they can be efficiently merged with the study participants' work and residence history files. This first phase of the program also incorporates a missing data algorithm which imputes values to zip-code months which do not have ambient concentration data from CARB. The missing data algorithm is described in detail below. The output of this first phase of the program is a data file for each pollutant ordered by zip code month which has replaced missing values with imputed values wherever possible. The second phase of the program takes the output of the first phase and merges it with the work location and residence history files to obtain monthly ambient concentration data for each study participant, as well as total cumulative ambient concentration over three fixed length time periods. This second phase of the program also contains algorithms described below which impute values wherever it is reasonable to do so for zip code months which may be missing from the study participant's work location or residence history.

§ 2. Phase One of Cumulative Ambient Concentration Program

§ 2.1 Overview

The first phase of the cumulative ambient concentration computer program is divided into a series of sub-steps as described below. Step 1 involves the selection of a specified pollutant and ambient concentration cutpoint from each of the 15 year-specific interpolated exceedance statistics tapes (1973-1987). All zip codes having this particular pollutant level are written to 15 separate files that are year-specific. Step 2 involves the concentration of these 15 pollutant-year-specific files into a single file.

In Step 3 this combined file is sorted by zip code and year within zip code. Some zip codes in the above file only contain interpolated exceedance data for the years 1973-76 and other zip codes only contain interpolated exceedance data for the years 1977-1987. Therefore, Step 4 creates dummy zip codes: year/month data items so that every zip code has 171 months (January, 1973 - March, 1987) of original or dummy data (the created dummy data will not affect analyses as no AHSMOG study subject lived at the specified zip code during the years in question).

Step 5 involves the replacement of original missing data items (coded "-1.0") with imputed data as described in section E.2.2. A fixed-fielded output file is then created for each pollutant with all of the exceedance statistics for each pollutant ordered first by zip code and then by month with working hours 8 am to 5 pm and non-working hours 5 pm to 8 am for gaseous pollutants recorded separately. This output file has been generated for the years 1973 through March, 1987, for all zip codes in California where study participants resided. It is a general purpose file and could be used by other investigators since the values for given zip codes in no way depend on study subjects. No cumulations take place during phase one of the program.

§ 2.2 Missing Data from CARB

For some zip codes none of the monitoring stations within interpolation distance have ambient concentration data for certain months for one or more pollutants, thus creating missing data. It is logical to impute values for such missing months, since in this study we are only interested in cumulative ambient concentration over many years. Furthermore, multivariate statistical analyses require no missing data for any of the cases for any of the variables used in a particular analysis. If values were not imputed for missing months, all cases with one or more months of missing data would have to be excluded from statistical analyses. This could result in a large number of cases being excluded and the remaining cases could be a biased sample of the total cohort. To avoid these problems, an algorithm is used to impute values for missing months. Counts of the number of missing months for each study participant are stored in the data base, so that individuals with too many missing values can be later excluded from statistical analyses.

The algorithm incorporates a series of rules which are shown in Exhibit 1.1. These rules are invoked in a successive manner in the alphabetical order as shown. The second rule is used only if a value cannot be imputed according to the first rule, and so on. Values imputed by lower rules in the hierarchy are judged to be of lesser quality. The rules are intended to be exhaustive, so that if any data exists for any months in any of the study years, a value will eventually be imputed. For zip codes with many months of missing data, this could result in values being imputed which were not likely to be representative of that particular month. In order to avoid using such data, rules are classified as "good rules" and "bad rules". Only good rules are allowed to impute values used for final statistical analyses. Missing months which might have had values imputed according to bad rules are left missing. The number of missing zip code months occurring in the study population and the classification of good rules and bad rules are shown for TSP and Ozone in Exhibits 1.2 and 1.3. For both these pollutants some zip codes were missing values for all study months. These zip codes were always located in outlying rural areas. They were located on a map and reassigned to background zero level ambient concentration, rule U, or to a nearby zip code thought to be representative which had a value, rule V. In most cases, where zip codes were reassigned, we were able to find a nearby zip code within interpolation range which had a value; if not, the nearest zip code thought to be representative was selected.

§ 3 Phase Two of Cumulative Ambient Concentration Program

§ 3.1 Overview

Phase two of the cumulative ambient concentration program takes the output from phase one, imputes values for work location or residence history months which are missing for study participants, and cumulates ambient concentrations over specified time periods. The program counts the number of zip code months for each time period which had missing values imputed either from phase one or phase two, and counts the number of zip code months with different interpolation quality ratings. The output of phase two consists of ambient concentration statistics for each month of the cohort period, 1977-1987, as well as for six fixed time periods as described below. All missing values are substituted with imputed values so that no cases are automatically excluded by statistical analysis packages. The counter variables produced by the program will enable exclusion of cases with more than an allowable number of missing months as specified by the investigator. Details of the phase two cumulative ambient concentration program are described.

§ 3.2 Definition and Purpose of Cohort Time Periods for Cumulation

Ambient concentrations are cumulated over six different time periods. The beginning and ending dates for each of these time periods are specified below.

1. - 01/01/73 - 03/31/77

   This time period is accumulated as two separate sub-time periods:

   A. - 01/01/73 - 12/31/76
   B. - 01/01/77 - 03/31/77

   Time period one represents the baseline ambient concentration and, as in the first study, is used as a surrogate of ambient concentration 1966 through 1977. Also during this time period a cumulation of non-missing data only is created. This was correlated with the data generated by Paul Allen as a reliability check on the 1987 method of cumulation. For this correlation study time period 1A--01/01/73 - 12/31/76--was used as this corresponds to the time period used for the data generated by Paul Allen. Cumulations have also been done for time period 1B--01/01/77 - 03/31/77. Imputations based on the entire time period 1 are used for missing values for time periods 1A or 1B.

2. - 04/01/77 - 12/31/82

This time period is needed because non-fatal cancer and heart disease incidence surveillance ended by the Adventist Health Study on 12/31/82. Thus this will be the latest termination date for these disease outcomes. Note that for cancer and heart disease incidence analyses, the cumulation of ambient concentration and time for a particular study participant could end before 12/31/82 if the individual was censored (i.e. lost to follow-up) by the Adventist Health Study or experienced incidence of a specific cancer or myocardial infarction. Even though AHSMOG follow-up extends, in some cases, beyond the Adventist Health Study censoring date, it is of no use for cancer or heart disease incidence since the AHSMOG study did not conduct ascertainment for these outcomes.

3. - 04/01/77 - 03/31/87

This time period is the period of ambient concentration between first and second ascertainment of respiratory symptoms on those who completed living questionnaires in 1987.

4. - 04/01/77 - 12/31/81

5. - 01/01/82 - 03/31/87

Time periods 4 and 5 are necessary because of a discontinuity in monitoring methods for particulates which occurred in 12/31/81.

6. - 01/01/73 - 12/31/81

This time period, formed by adding time periods 1 and 4, will also be available for use in sensitivity analyses.

For time periods two through five cumulative ambient concentration for an individual study participant may not extend to the end of the time period if an individual died or was lost to follow-up by the AHSMOG study.

§ 3.3 Daily Time Periods for Gaseous Pollutants

For gaseous pollutants--OX, NO_2, SO_2--cumulative ambient concentrations are calculated according to three distinct daily time segments according to where the study participant was located during this part of the day. These three time segment locations are:

1. Residence location--non-working hours, 6 pm to 8 am
2. Residence location--working hours, 8 am to 6 pm
3. Work location--working hours, 8 am to 6 pm

For a given month, an individual's total cumulative ambient concentrations for that month was the sum of the cumulative ambient concentrations for daily time segment one and three above if work locations were present--that is, an individual reported working more than five miles from home. If an individual did not report working more than five miles from home, the cumulative ambient concentration for the month was the sum of the cumulative ambient concentrations for daily time segments one and two above.

§ 3.4 Work versus Home Location for Particulate Pollutants--Daily Weighting Factors

The particulate pollutants, TSP and SO_4, are monitored for 24 hour periods. Therefore, ambient concentration was cumulated for residence location for the entire month and for work location for the entire month. If work locations were present and differed from home locations, the total cumulative ambient concentration for the month was obtained by a weighted addition of the home ambient concentration and work ambient concentration using the formula:

Total ambient concentration = 9/24 x Work ambient concentration + 15/24 x Home ambient concentration.

§ 3.5 Types of Data

The phase two cumulative ambient concentration program must cope with 13 different types of data according to the output of phase one or the nature of the zip codes and zip code flags in the work location and residence history files. These 13 different data types are:

1-5. Good data with quality ratings A, B, C, H, I.
6. Missing data imputed by good rules.
7. Missing data from CARB--only bad rules could have been used to impute values or no values could be imputed.
8. Zip code of residence or work location missing.
9. Zip code out of state but classified as high smog.
10. Zip code out of state but classified as low smog.
11. Out of country classified as high smog.
12. Out of country classified as low smog.
13. Zip code in state but no fixed location or location unknown, code 99999 used.

The different types of data are treated by the phase two cumulative ambient concentration program in the following ways:

- Types 10 and 12 are given an ambient concentration of 0
- Types 1-6, 10, and 12 are treated as good data
- Types 7, 8, 9, 11, and 13 are treated as missing values

The output of phase 2 of the program does not differentiate between the 13 different types of data but does retain counts of the number of zip code daily time segments for each of the 13 types of data for each of the six time periods for each study participant in the analysis file. To allow separate counting of different data types for work location and home location, the day is divided into three time segments with one time segment allocated to working hours and two time segments allocated to non-working hours. Counts can be converted back to zip code months by dividing by three.

§ 3.6 Missing Values for Work Location

Work location missing values are replaced by residence location values.

§ 3.7 Missing Value Imputation Rule for Residence Time Period One

Missing values for residence during this time period are replaced with a good data months average. The average of the good data months is computed by cumulating ambient concentration over the good data months and dividing by the number of good data months. The total cumulative ambient concentration for this time period is obtained by multiplying the total number of months in the time period by this good data months average, effectively replacing missing value months with the average for good data months.

§ 3.8 Missing Value Imputation Rule for Residence for All Other Time Periods

Missing values which occur in all other time periods are replaced by the average over the 12 most recent months having good data. This same rule is used when individual months between 04/01/77 and 03/31/87 are treated as individual time periods for the Cox regression time dependent covariate models.

§ 3.9 Types of Missing Residence Data from Study Subjects

A few study subjects who were alive in 1987 and completed the questionnaire or who were deceased and were reported on in 1987 by surrogates, had some work location or residence history months where the location was unknown. Study subjects for whom no response was obtained, either by living or surrogate questionnaire in 1987, often had several months of missing residence history data since the last follow-up mailing for the Adventist Health Study was in March, 1983. If these study subjects were deceased and the last residence according to the death certificate was the same as the last residence in the Adventist Health Study, we assumed that they had not moved between last contact by Adventist Health Study and time of death. Otherwise, a code for unknown residence was assigned.

§ 4. Merging 1977 and 1987 Residence History Files

During the computer merging of the work location and residence history files obtained by the 1977 questionnaire with those obtained by the 1987 questionnaire, some discrepancies were found. A listing of all study participants with such discrepancies was made and their 1977 and 1987 questionnaires manually examined in an attempt to reconcile the discrepancies. In most cases, we were able to reconcile the discrepancies. However, in a few cases study participants had to be telephoned.

REFERENCES

Hayes, T.P., Kinney, J.R., Wheeler, N.J. (1984). California surface wind climatology, California Air Resources Board.

Table G.1

Comparison of 1973-1976 "Old" versus "New" Average Values of Ambient Concentration Statistics for South Coast, San Diego, and San Francisco Air Basins

Previous "Old" Computations by Paul Allen versus Current "New" Computations
Averages are over all Months and all Stations

	SOUTH COAST				SAN DIEGO				SAN FRANCISCO			
Statistic	n	Mean Old	Mean New	Correlation	n	Mean Old	Mean New	Correlation	n	Mean Old	Mean New	Correlation
<u>Oxidants</u>												
Total Concentration	1961	2180.8	2178.0***	1.000	257	1927.7	1932.8	0.997	1059	1249.0	1249.0	0.999
Excess Hrs. 10	1961	38.3	35.2***	0.998	257	7.1	6.0***	0.989	1059	2.0	1.7***	0.990
Excess Concentration	1961	183.4	181.3***	0.999	257	18.3	18.0	0.998	1059	4.6	4.5	0.996
Excess Hrs. 15	1961	13.5	12.2***	0.998	257	1.0	0.8***	0.989	1059	0.2	0.2***	0.981
Excess Concentration	1961	60.8	60.0***	0.999	257	2.8	2.7	0.998	1059	0.4	0.4***	0.986
Excess Hrs. 20	1961	4.5	4.0***	0.997	257	0.2	0.1***	0.997	1059	0.02	0.02	0.863
Excess Concentration	1961	18.7	18.4***	0.999	257	0.5	0.5	0.988	1059	0.02	0.03**	0.897
Excess Hrs. 25	1961	1.4	1.2***	0.995	257	0.03	0.04	0.982	1059	0.0	0.0	---+
Excess Concentration	1961	5.1	5.0**	0.999	257	0.1	0.1	0.990	1059	0.0	0.0	---+

c:\final\append-G.rep May 13, 1994

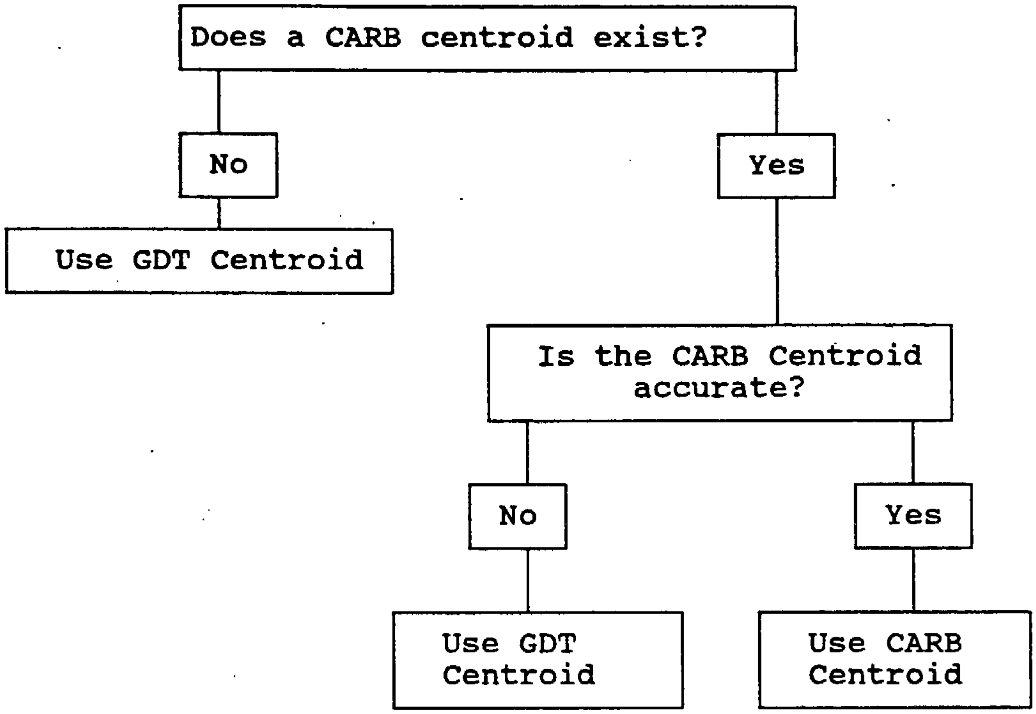
(Table G.1, con't)

	SOUTH COAST				SAN DIEGO				SAN FRANCISCO			
Statistic	n	Mean		Correl-ation	n	Mean		Correl-ation	n	Mean		Correl-ation
		Old	New			Old	New			Old	New	
TSP												
Total Concentration	1392	73640.8	73636.5	1.000	247	64224.0	64224.0	1.000	1012	46573.4	46518.7	0.999
Excess Hrs. 60 μg/m³	1392	570.5	570.5	1.000	247	550.7	550.7	1.000	1012	322.3	322.0	0.998
Excess Concentration	1392	32732.5	32742.0	1.000	247	23175.3	23175.3	1.000	1012	11149.8	11116.6	0.999
Excess Hrs. 100 μg/m³	1392	303.4	303.3	1.000	247	204.5	204.5	1.000	1012	93.2	93.0	0.998
Excess Concentration	1392	15163.0	15165.8	1.000	247	8260.6	8260.6	1.000	1012	3703.9	3684.0	0.999
Excess Hrs. 150 μg/m³	1392	105.3	105.3	1.000	247	51.6	51.6	1.000	1012	24.3	24.2	0.994
Excess Concentration	1392	5546.6	5546.5***	1.000	247	2398.31	2398.27***	1.000	1012	1200.6	1188.2	0.998
Excess Hrs. 200 μg/m³	1392	35.5	36.0	0.988	247	18.9	18.9	1.000	1012	8.1	8.0	0.992
Excess Concentration	1392	2323.2	2323.2	1.000	247	798.9	798.9	1.000	1012	493.9	490.1	0.999

* New mean significantly different from old mean ($p < .05$)
** New mean significantly different from old mean ($p < .01$)
*** New mean significantly different from old mean ($p < .001$)
\+ Could not be computed as all values were zero

FIGURE 1

DECISION TREE FOR ASSIGNING CENTROIDS TO ZIP CODES, 1973-76



c:\final\append-G.rep May 13, 1994

FIGURE 2

DECISION TREE FOR ASSIGNING CENTROIDS TO ZIP CODES, 1977-87

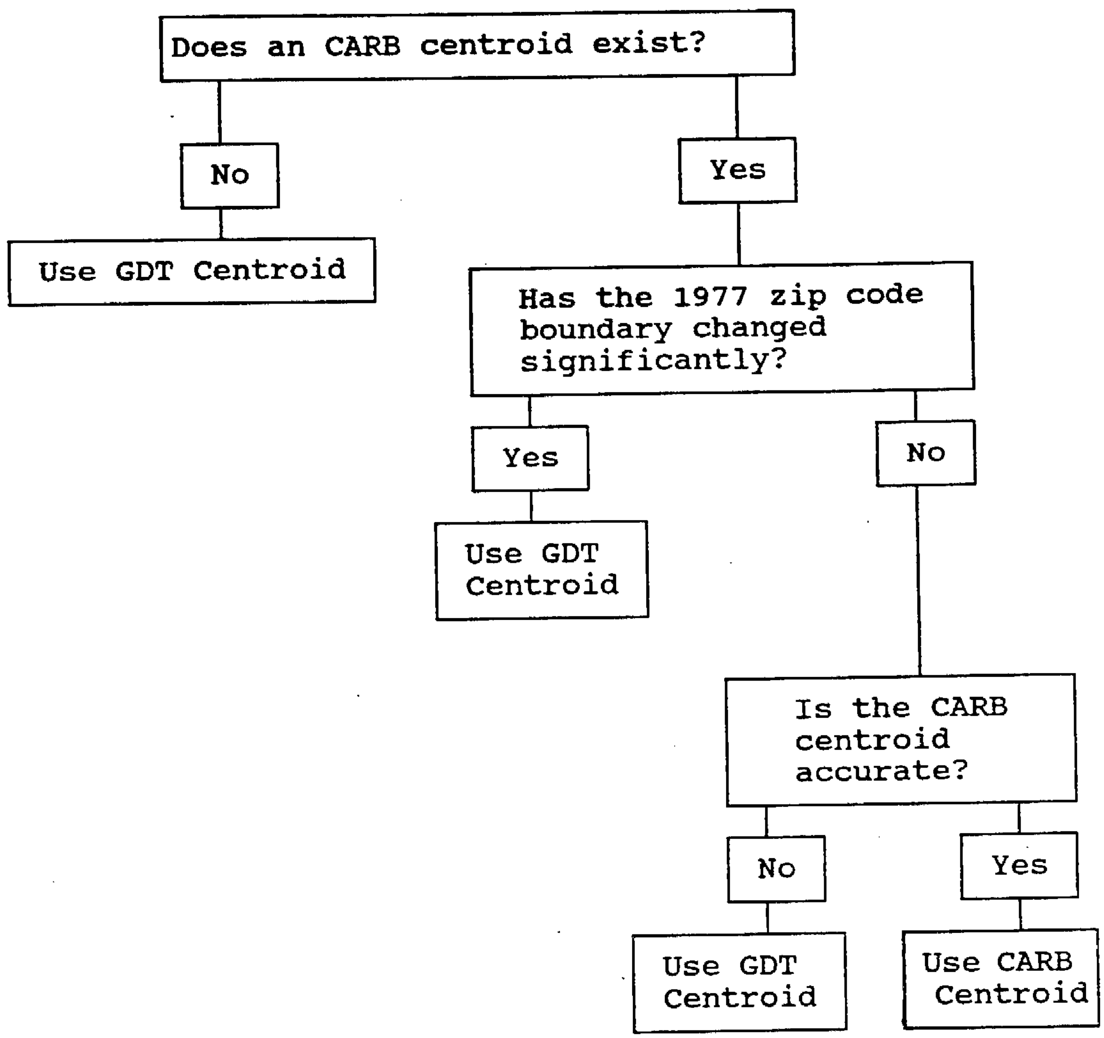


FIGURE 3

DECISION TREE FOR CHOOSING FOR EACH ZIPCODE THE "SUPERSET" OF STATIONS RELEVANT TO INTERPOLATIONS

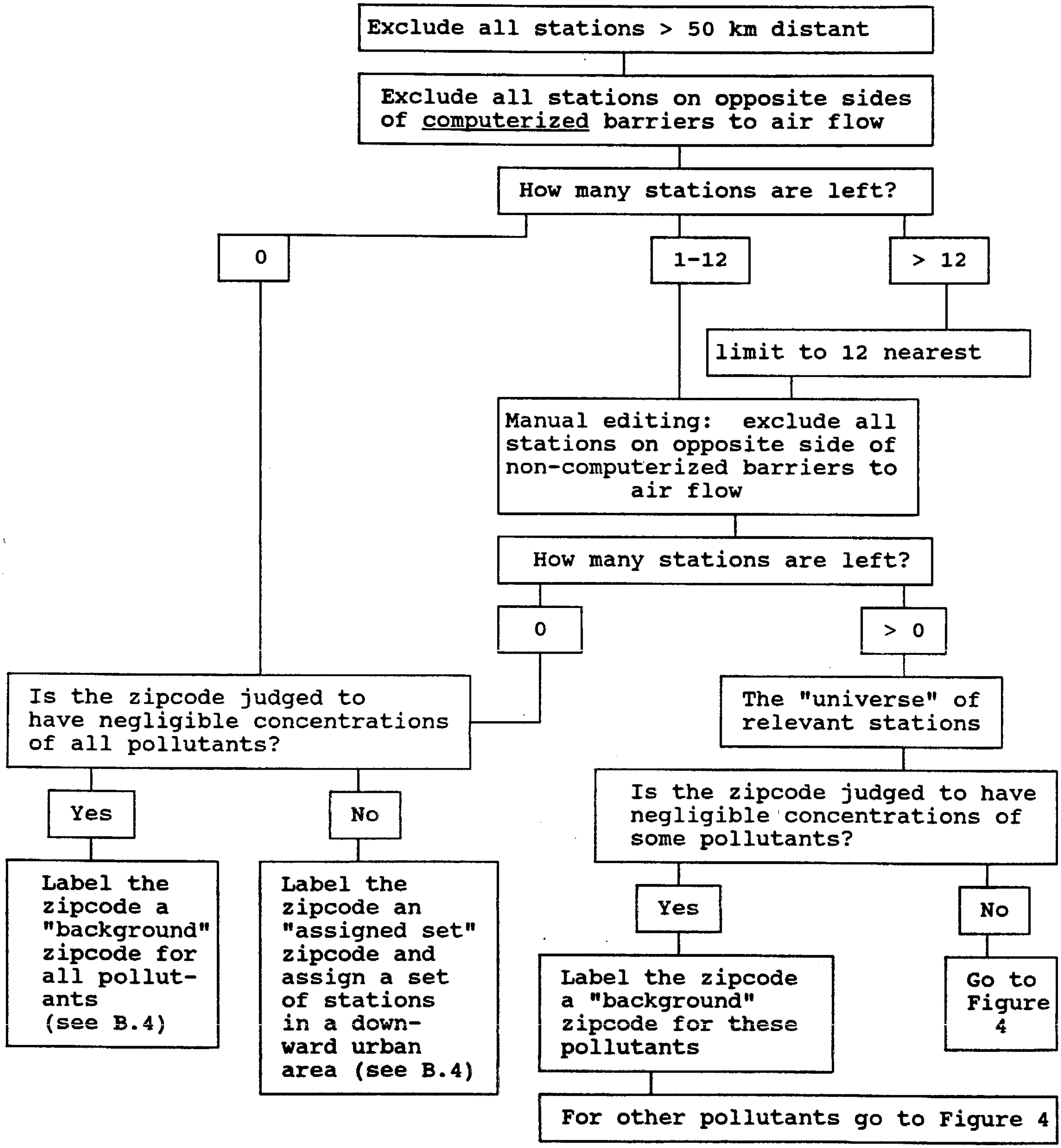
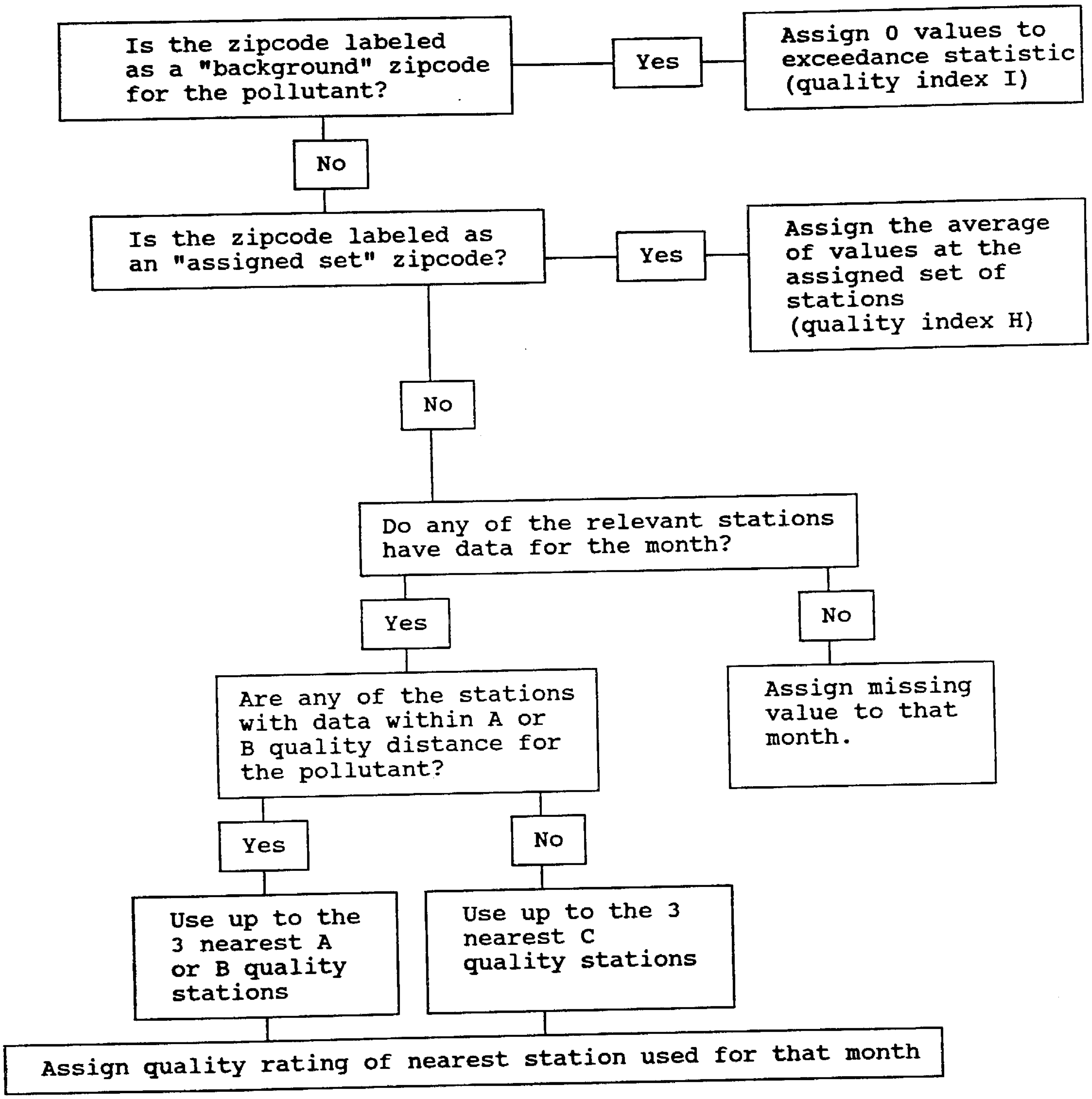


FIGURE 4

DECISION TREE FOR DECIDING FOR A ZIPCODE, POLLUTANT, AND MONTH WHICH DATA FROM RELEVANT STATIONS WILL BE USED IN THE INTERPOLATION



Appendix G

EXHIBIT 1.1 Rules for Imputing Values for Missing Data from CARB

G = Good Rule
B = Bad Rule (Bad rules were replaced with previous 12 month average in Phase II Cumulation)

(G) A : If more than 75% of the months are 0.0, then impute a 0.0 value to all missing months (i.e. -1.0 ----> 0.0).

(G) B : For quality codes = 'a' or 'b', average over months on either end of the missing string, but do not jump over more than two months to find a valid data item. If missing string does not have a valid data item on one end, then use only the other valid end. Quality codes 'a' and 'b' are treated equally. Use whichever is closest to the missing string.

(G) C : For quality codes = 'c', average over months on either end of the missing string, but do not jump over more than two months to find a valid data item. If missing string does not have a valid data item on one end, then use only the other valid end.

(G) D : For quality codes = 'a' or 'b', average over same month but adjacent years, but do not jump over more than two years to find a valid data item. If missing data does not have a valid data item in adjacent year on one end, then use only the other valid adjacent year.

(G) E : For quality codes = 'c', average over the some month, but use adjacent years, but do not jump over more than two years to find a valid data item. If missing data does not have a valid data item in adjacent year on one end, then use only the other valid adjacent year.

(G) F : For quality codes = 'a' or 'b', compute the average of the nearest months on either end of the missing value looking over the entire span of years but do not jump over more than two months in a given year to find a valid month. (i.e. Look over entire span of years looking both left and right no more than 3 months for the nearest valid month. A maximum of four months will be averaged for the missing month -- two from years less than the year with missing data and two from years greater than the year with missing data. Less than 4 months will be obtained if the missing item is near the boundary of the year:month matrix or if no valid month could be found within 3 months for a given year and a given direction.)

(G) G : Same as rule (F) except for quality code = 'c'.

Exhibit 1.1 (cont'd)

(B) H : For quality codes = 'a' or 'b', compute the average of all valid months for all years and impute that average into all missing items only if at least 4 different month names (e.g. Jan, Feb, Mar, Apr) are represented.

(B) I : Same as rule (H) except for quality code = 'c'.

(B) J : For quality codes = 'a' or 'b', compute the average of all valid months for all years and impute that average into all missing items (-1.0 -----> avg.) with no restriction as to the number of months represented by the valid data.

(B) K : For quality codes = 'c', compute the average of all valid months for all years and impute that average into all missing items (-1.0 -----> avg.) with no restriction as to the number of months represented by the valid data.

(G) L : Same as rule (B) except for quality 'h'.

(G) O : Same as rule (D) except for quality 'h'.

(G) P : Same as rule (F) except for quality 'h'.

(B) Q : Same as rule (H) except for quality 'h'.

(B) R : Same as rule (J) except for quality 'h'.

(G) S : For quality codes = 'I' (all of which would be 0.0), place a 0.0 into all missing items (-1.0 -----> 0.0).

(B) T : None of the above rules apply because there are no valid data for any year for that zip code (i.e. All months for all years for this zip code = -1.0).

(G) U : Recoded from rule T to background value of 0.0 after looking on zip code maps.

(B) V : Recoded to another nearby zip code after looking on zip code maps.

==

(B) M : All months for a given year were not even on the data tapes from CARB. (i.e. '73 to '77 missing or '78 to '87 missing).

(G) N : Valid data - no rule applied (data >= 0.0).

Note: Rules A - L, O - T were applied in alphabetical order in that a later rule was applied only when all previous rules failed.

Exhibit 1.2 Summary of Quality Codes and Missing Data for TSP, Hours in Excess of 100 $\mu g/m^3$.

DATE: 12 June, 1990

SUBJECT: Interpolated Exceedance Statistics 'Rule' program
(After missing zip codes were converted to 'V' & 'U')

Test Pollutant: '41005' = TSP - 100 Total hours

1304 Unique zip codes
* 171 Months / zip code

224,984 (Zip code:year:month) data items per pollutant

204,422 Non-missing data (>= 0.0)

44,115 for quality area 'A'
63,926 for quality area 'B'
83,998 for quality area 'C'
6,701 for quality area 'H'
5,682 for quality area 'I'

13,068 Given background values of 0.0 (Rule 'U')

0 Copied from another nearby zip code (Rule 'V')

0 for quality area 'A'
0 for quality area 'B'
0 for quality area 'C'
0 for quality area 'H'
0 for quality area 'I'

5,494 Missing data imputed to have the following quality areas because of the rule program. (Note: All original missing data on John Moore's tapes had data= -1.0 and a quality of 'X'. The rule program carried along the quality of the corresponding data used in the imputation.)

573 imputed data items given quality 'A'
954 imputed data items given quality 'B'
2,565 imputed data items given quality 'C'
1,402 imputed data items given quality 'H'
0 imputed data items given quality 'I'

0 Missing data items remain quality 'X'

Exhibit 1.2 (cont'd)

DATE: 12 June, 1990

SUBJECT: Interpolated Exceedance Statistics 'Rule' program
After missing zip codes were converted to 'V' & 'U')

Test Pollutant: '41005' = TSP - 100 Total hours

1304	Unique zip codes
* 171	Months / zip code (12 months each for '73 to '86 + 3 months for '87)
222,984	(Zip code:Year:month) data items per pollutant

204,422 Non-missing data (>= 0.0) (Rule 'N')

18,562 Missing data (coded as -1.0 on John Moore's tapes) to be imputed using the following rules.

490	Imputed using rule "A"
383	Imputed using rule "B"
1,095	Imputed using rule "C"
659	Imputed using rule "D"
921	Imputed using rule "E"
390	Imputed using rule "F"
343	Imputed using rule "G"
0	Imputed using rule "H"
0	Imputed using rule "I"
0	Imputed using rule "J"
0	Imputed using rule "K"
358	Imputed using rule "L"
496	Imputed using rule "O"
359	Imputed using rule "P"
0	Imputed using rule "Q"
0	Imputed using rule "R"
0	Imputed using rule "S"
0	Not Imputed using above rules because every month of every year for given zip code was missing. No nearby zip code could be found. (Rule 'T')
13,068	Background data=0.0 (Rule 'U')
0	Copied from another zip code (Rule 'V')

Exhibit 1.3 Summary of Quality Codes and Missing Data for Ozone, Working Hours in Excess of 10 pphm.

DATE: 16 September, 1993

SUBJECT: Interpolated Exceedance Statistics 'Rule' program

Pollutant: '10101' = Oz 10 pphm (working hours)
(newly revised data)

1312 Unique zip codes
* 171 Months / zip code

224,352 (Zip code:year:month) data items per pollutant

200,767 Non-missing data (>= 0.0)

154,804 for quality area 'A'
23,902 for quality area 'B'
8,812 for quality area 'C'
6,760 for quality area 'H'
6,489 for quality area 'I'

342 Given background values of 0.0 (Rule "U")

2,052 Copied from another nearby zip code (Rule "V")
698 for quality area 'A'
717 for quality area 'B'
466 for quality area 'C'
0 for quality area 'H'
171 for quality area 'I'

9,840 Missing years not supplied (Rule 'M')

11,351 Missing data imputed to have the following quality areas because of the rule program. (Note: All original missing data on John Moore's tapes had data= -1.0 and a quality of 'X'. The rule program carried along the quality of the corresponding data used in the imputation.)

1,476 imputed data items given quality 'A'
3,992 imputed data items given quality 'B'
2,653 imputed data items given quality 'C'
3,230 imputed data items given quality 'H'
0 imputed data items given quality 'I'
0 imputed data items remain quality 'X'

Exhibit 1.3 (cont'd)

DATE: 16 September, 1993

SUBJECT: Interpolated Exceedance Statistics 'Rule' program

Pollutant: '10101' = Oz, 10 pphm (working hours)
(newly revised data)

1312 Unique zip codes
* 171 Months / zip code (12 months each for '73 to '86 + 3 months for '87)

224,352 (Zip code:Year:month) data items per pollutant

200,767 Non-missing data (>= 0.0)

9,840 Year Missing. These are zip codes that had no data for years '73 to '76 or zip codes that had no data for years '77 to '86. These data should not be of concern because apparently none of the AHSMOG subjects lived in these zip codes during these years. (Rule "M")

11,351 Missing data (coded as -1.0 on John Moore's tapes) to be imputed using the following rules.

287 Imputed using rule "A"
1,178 Imputed using rule "B"
674 Imputed using rule "C"
2,037 Imputed using rule "D"
903 Imputed using rule "E"
2,028 Imputed using rule "F"
885 Imputed using rule "G"
21 Imputed using rule "H"
0 Imputed using rule "I"
0 Imputed using rule "J"
112 Imputed using rule "K"
726 Imputed using rule "L"
1,438 Imputed using rule "O"
1,062 Imputed using rule "P"
0 Imputed using rule "Q"
0 Imputed using rule "R"
0 Imputed using rule "S"
0 Not Imputed using above rules because every month of every year for given zip code was missing. (Rule "T")
342 Background data=0.0 (Rule "U")
2,052 Copied from another zip code (Rule "V")

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
PB94218740	May 1994	Final Report

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Incidence of Respiratory Symptoms and Chronic Disease in a Non-Smoking Population as a Function of Long-Term Cumulative Exposure to Ambient Air Pollutants (Adventist Health Study of Smog Follow-Up Study) Volume 1: Final Report	A932-160
6. AUTHOR(S) David E. Abbey	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Loma Linda Center for Health Promotion Loma Linda University Loma Linda, CA 92350	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) California Air Resources Board Research Division 2020 L Street Sacramento, CA 95814	10. SPONSORING/MONITORING AGENCY REPORT NUMBER ARB/R-94/542

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Release unlimited. Available from National Technical Information Service. 5285 Port Royal Road Springfield, VA 22161	

13. ABSTRACT *(Maximum 200 Words)*

Data on occurrence and severity of chronic bronchitis and asthma, on occurrence of cancer, heart disease, and mortality, and on numerous lifestyle and air pollution exposure covariates were collected from a cohort of 6340 non-smoking California Seventh-day Adventists in 1987. Monthly pollutant exposures were estimated from ambient concentrations of seven air pollutants--ozone, NO2, SO2, TSP, PM10 estimated from TSP, PM2.5 estimated from visibility, and SO4--experienced by subjects during 1966-87. Statistical analyses estimated relationships between these health endpoints and long-term average ambient concentrations, adjusting for covariates. TSP, the pollutant most strongly related to health endpoints, was significantly related to occurrence of and increases in severity of chronic bronchitis and asthma and to occurrence of cancers in females. PM10 was associated with occurrence of chronic bronchitis and severity of asthma, and PM2.5 was significantly but less strongly associated with several respiratory endpoints. Ozone and SO4 were significantly associated with occurrence and severity of asthma. There were no significant associations between any health endpoint, SO2, and NO2. Multipollutant analyses were performed to investigate whether any of the significant relationships between pollutants and health endpoints were due to surrogate relationships with other pollutants.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Air pollution, respiratory diseases, malignant neoplasms, mortality.	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

CARB-RD Version (Org. 5-93) of Federal Standard Form 298 (Rev. 2-89)