APPENDIX I-J

LISTING OF LABEL FILES

LISTING OF HOUR24.LBL

```
LABEL
 ID1 ='STUDY NUMBER'
 ID2 ='HOUSEHOLD/PARTICIPANT ID'
 CMPD='COMPOUND ID'            /* 2='Chloroform'
                                   3='1,2-Dichloroethane'
                                   4='1,1,1-Trichloroethane'
                                   5='Benzene'
                                   6='Carbon Tetrachloride'
                                   7='Trichloroethylene'
                                   8='Bromodichloromethane'
                                   9='Dibromochloromethane'
                                  11='Tetrachloroethylene'
                                  12='Chlorobenzene'
                                  13='Bromoform'
                                  14='Styrene'
                                  15='m-Dichlorobenzene'
                                  16='m,p-Dichlorobenzene'
                                  17='p-Dichlorobenzene'
                                  18='o-Dichlorobenzene'
                                  19='Ethylbenzene'
                                  20='o-Xylene'
                                  21='m,p-Xylene'
                                  23='n-Decane'
                                  24='n-Dodecane'
                                  25='1,4-Dioxane'
                                  26='1,1,1,2-Tetrachloroethane'
                                  27='1,2-Dibromoethane'
                                  28='n-Octane'
                                  29='n-Undecane'
                                  30='1,1,2,2-Tetrachloroethane'
                                  31='a-Pinene'
                                  32='Limonene'
                                  33='n-Nonane' */
  CPER='CONCENTRATION FOR PERSONAL AIR'
  CKIT='CONCENTRATION FOR KITCHEN'
  CMLA='CONCENTRATION FOR MAIN LIVING AREA'
  COUT='CONCENTRATION FOR OUTDOORS'
  MPER='MEASURABILITY INDICATOR FOR PERSONAL AIR'      /*1='QUANTIFIABLE'
                                                         2='BELOW THE QL'*/
  MKIT='MEASURABILITY INDICATOR FOR KITCHEN'           /*1='QUANTIFIABLE'
                                                         2='BELOW THE QL'*/
  MMLA='MEASURABILITY INDICATOR FOR MAIN LIVING AREA'  /*1='QUANTIFIABLE'
                                                         2='BELOW THE QL'*/
  MOUT='MEASURABILITY INDICATOR FOR OUTDOORS'          /*1='QUANTIFIABLE'
                                                         2='BELOW THE QL'*/
  WPER='SAMPLING WEIGHT FOR PERSONAL AIR'
  WMLA='SAMPLING WEIGHT FOR MAIN LIVING AREA'
  WOUT='SAMPLING WEIGHT FOR OUTDOORS'
```

LISTING OF DAYNITE.LBL

```
LABEL
 ID1 ='STUDY NUMBER'
 ID2 ='HOUSEHOLD/PARTICIPANT ID'
 CMPD='COMPOUND ID'            /* 2='Chloroform'
                                   3='1,2-Dichloroethane'
                                   4='1,1,1-Trichloroethane'
                                   5='Benzene'
                                   6='Carbon Tetrachloride'
                                   7='Trichloroethylene'
                                   8='Bromodichloromethane'
                                   9='Dibromochloromethane'
                                  11='Tetrachloroethylene'
                                  12='Chlorobenzene'
                                  13='Bromoform'
                                  14='Styrene'
                                  15='m-Dichlorobenzene'
                                  16='m,p-Dichlorobenzene'
                                  17='p-Dichlorobenzene'
                                  18='o-Dichlorobenzene'
                                  19='Ethylbenzene'
                                  20='o-Xylene'
                                  21='m,p-Xylene'
                                  23='n-Decane'
                                  24='n-Dodecane'
                                  25='1,4-Dioxane'
                                  26='1,1,1,2-Tetrachloroethane'
                                  27='1,2-Dibromoethane'
                                  28='n-Octane'
                                  29='n-Undecane'
                                  30='1,1,2,2-Tetrachloroethane'
                                  31='a-Pinene'
                                  32='Limonene'
                                  33='n-Nonane' */
 CPERD='CONCENTRATION FOR PERSONAL AIR, DAYTIME'
 CKITD='CONCENTRATION FOR KITCHEN, DAYTIME'
 CMLAD='CONCENTRATION FOR MAIN LIVING AREA, DAYTIME'
 COUTD='CONCENTRATION FOR OUTDOORS, DAYTIME'
 MPERD='MEASURABILITY INDICATOR FOR PERSONAL AIR, DAYTIME'
        /*1='QUANTIFIABLE'  2='BELOW THE QL'*/
 MKITD='MEASURABILITY INDICATOR FOR KITCHEN, DAYTIME'
        /*1='QUANTIFIABLE'  2='BELOW THE QL'*/
 MMLAD='MEASURABILITY INDICATOR FOR MAIN LIVING AREA, DAYTIME'
        /*1='QUANTIFIABLE'  2='BELOW THE QL'*/
 MOUTD='MEASURABILITY INDICATOR FOR OUTDOORS, DAYTIME'
        /*1='QUANTIFIABLE'  2='BELOW THE QL'*/
 WPERD='SAMPLING WEIGHT FOR PERSONAL AIR, DAYTIME'
 WOUTD='SAMPLING WEIGHT FOR OUTDOORS, DAYTIME'
 CPERN='CONCENTRATION FOR PERSONAL AIR, NIGHTTIME'
 CKITN='CONCENTRATION FOR KITCHEN, NIGHTTIME'
 COUTN='CONCENTRATION FOR OUTDOORS, NIGHTTIME'
```

```
MPERN='MEASURABILITY INDICATOR FOR PERSONAL AIR, NIGHTTIME'
        /*1='QUANTIFIABLE'  2='BELOW THE QL'*/
MKITN='MEASURABILITY INDICATOR FOR KITCHEN, NIGHTTIME'
        /*1='QUANTIFIABLE'  2='BELOW THE QL'*/
MOUTN='MEASURABILITY INDICATOR FOR OUTDOORS, NIGHTTIME'
        /*1='QUANTIFIABLE'  2='BELOW THE QL'*/
WPERN='SAMPLING WEIGHT FOR PERSONAL AIR, NIGHTTIME'
WOUTN='SAMPLING WEIGHT FOR OUTDOORS, NIGHTTIME'
```

LISTING OF BREATH.LBL

```
LABEL
 ID1 ='STUDY NUMBER'
 ID2 ='HOUSEHOLD/PARTICIPANT ID'
 CMPD='COMPOUND ID'            /* 2='Chloroform'
                                   3='1,2-Dichloroethane'
                                   4='1,1,1-Trichloroethane'
                                   5='Benzene'
                                   6='Carbon Tetrachloride'
                                   7='Trichloroethylene'
                                   8='Bromodichloromethane'
                                   9='Dibromochloromethane'
                                  11='Tetrachloroethylene'
                                  12='Chlorobenzene'
                                  13='Bromoform'
                                  14='Styrene'
                                  15='m-Dichlorobenzene'
                                  16='m,p-Dichlorobenzene'
                                  17='p-Dichlorobenzene'
                                  18='o-Dichlorobenzene'
                                  19='Ethylbenzene'
                                  20='o-Xylene'
                                  21='m,p-Xylene'
                                  23='n-Decane'
                                  24='n-Dodecane'
                                  25='1,4-Dioxane'
                                  26='1,1,1,2-Tetrachloroethane'
                                  27='1,2-Dibromoethane'
                                  28='n-Octane'
                                  29='n-Undecane'
                                  30='1,1,2,2-Tetrachloroethane'
                                  31='a-Pinene'
                                  32='Limonene'
                                  33='n-Nonane' */
 CBRH='CONCENTRATION FOR BREATH'
 MBRH='MEASURABILITY INDICATOR FOR BREATH'      /*1='QUANTIFIABLE'
                                                  2='BELOW THE QL'*/
 WBRH='SAMPLING WEIGHT FOR BREATH'
```

LISTING OF WATER.LBL

```
LABEL
 ID1 ='STUDY NUMBER'
 ID2 ='HOUSEHOLD/PARTICIPANT ID'
 CMPD='COMPOUND ID'            /* 2='Chloroform'
                                  3='1,2-Dichloroethane'
                                  4='1,1,1-Trichloroethane'
                                  5='Benzene'
                                  6='Carbon Tetrachloride'
                                  7='Trichloroethylene'
                                  8='Bromodichloromethane'
                                  9='Dibromochloromethane'
                                 11='Tetrachloroethylene'
                                 12='Chlorobenzene'
                                 13='Bromoform'
                                 14='Styrene'
                                 15='m-Dichlorobenzene'
                                 16='m,p-Dichlorobenzene'
                                 17='p-Dichlorobenzene'
                                 18='o-Dichlorobenzene'
                                 19='Ethylbenzene'
                                 20='o-Xylene'
                                 21='m,p-Xylene'
                                 23='n-Decane'
                                 24='n-Dodecane'
                                 25='1,4-Dioxane'
                                 26='1,1,1,2-Tetrachloroethane'
                                 27='1,2-Dibromoethane'
                                 28='n-Octane'
                                 29='n-Undecane'
                                 30='1,1,2,2-Tetrachloroethane'
                                 31='a-Pinene'
                                 32='Limonene'
                                 33='n-Nonane' */
 CWAT='CONCENTRATION FOR WATER'
 MWAT='MEASURABILITY INDICATOR FOR WATER'      /*1='QUANTIFIABLE'
                                                 2='BELOW THE QL'*/
 WWAT='SAMPLING WEIGHT FOR WATER'
```

LISTING OF QUEST.LBL

```
LABEL
ID1 ='Study ID'
ID2 ='Household/Participant ID'
ID3 ='Monitoring Date (Start of Period)'
ID4 ='Stratum Code'
ID5 ='Primary Sampling Unit Code'
ID6 ='Personal Sampling Weight'
ID7 ='Household Sampling Weight'
PG1 ='Sex'                                  /* Male, Female */
PG2 ='Race'                                 /* Hispanic, Native American,
                                               Non-Hispanic Black, Asian,
                                               Non-Hispanic White, Other */

PG3 ='Age (in years)'
PG4 ='Weight (in pounds)'
PG5 ='Height (in inches)'
PG6 ='Employment Status'                    /* Employed, Housewife, Student
                                               Unemployed, Retired, Disabled */

HG1 ='Municipal Water Supply'
HG2 ='Home Treated for Pests'               /* All, Part, None */
HG3 ='Home Carpeted'
HG4 ='Carpet Cleaning in Past Year'         /* Steam, Prof. Dry Clean, Other,
                                               None */

HG5 ='Comm. Clean (Drapes/Carpet/Furn.)'
HG6 ='Attached Garage'
HG7 ='Central Air Conditioning'
HG8 ='Window Air Conditioning'
HG9 ='Portable Circulation Fans'
HG10='Ceiling Exhaust Fan'
HG11='Gas Furnace'
HG12='Fireplace'
HX7 ='Central Air Conditioning'
HX8 ='Window Air Conditioning'
HX9 ='Circulation Fans'
HX10='Exhaust Fans'
HX11='Gas Furnace'
HX12='Fireplace'
HX13='Window Fan'
HX14='Bathroom/Kitchen Exhaust'
HX15='Doors or Windows Open'
HX16='Gas or Kerosene Space Heater'
HX17='Woodstove'
HG18='Gas Stove'
HG19='Gas Water Heater'
HG20='Gas Clothes Dryer'
HG21='Electric Oven'
HX22='Gas Stove'
PX7 ='Dishwasher'
PX8 ='Clotheswasher'
PX9 ='Gas Clothes Dryer'
```

```
HG23='Paints/Varnishes/Thinners/Removers'
HG24='Gasoline/Petroleum Products'
HG25='Auto, Motorcycle, or Gas Lawnmower'
HG26='-Cides or Garden/Lawn Chemicals'
HG27='Cleaning Supplies'
HG28='New Interior Furnishings'
HG29='Room Deodorizers'
HG30='Automotive Care Products'
PG10='Painting'
PG11='Dry Cleaning'
PG12='Petroleum Plant'
PG13='Service Station/Garage/Engine Repair'
PG14='Furniture Repair/Refinishing'
PG15='Plastics Manufacture/Formulation'
PG16='Textile Mill'
PG17='Wood Processing Plant'
PG18='Printing'
PG19='Scientific Laboratory'
PG20='Dye Plant'
PG21='Hospital'
PG22='Metal Products'
PG23='Battery/Elect. Components Manufacture'
PG24='Refrigerator/AC Repair/Manufacture'
PG25='Taxi/Bus/Truck Driver'
PG26='Pest Control'
PG27='Drug Manufacture/Formulation'
PG28='Photo Developing'
PG29='Landscaping/Gardening'
PG30='Painting'
PG31='Furniture Refinishing'
PG32='Model Building'
PG33='Gardening'
PG34='House Plants'
PG35='Automobile or Bicycle Repair'
HG31='Painting'
HG32='Furniture Refinishing'
HG33='Model Building'
HG34='Gardening'
HG35='House Plants'
HG36='Automobile/Mechanical Repair'
PG36='Smoking Status'                          /* Current Smoker, Former Smoker,
                                                  Never Smoked */
PG37='Avg. No. Cigarettes/Day'                 /* <10, 10-30, >30 */
HG37='HH Members Smoke Cigarettes'             /* Participant Only, Others Only,
                                                  Both, Neither */
PG38='Avg. Hrs/Day Enclosed w/Smoker-Home'     /* 0, 1-5, 6-10, >10 */
PG39='Avg. Hrs/Day Enclosed w/Smoker-Work'     /* 0, 1-5, 6-10, >10 */
HX38='No. Cigarettes Smoked in House'
HX39='No. Cigarettes, Cigars, or Pipefuls'
HG40='Tobacco Prod. Smoked in House'
```

```
PX40='No. Cigarettes Smoked'
PD40='No. Cigarettes Smoked - Daytime'
PN40='No. Cigarettes Smoked - Nighttime'
PX41='Smoked Tobacco Products'
PD41='Smoked Tobacco Products - Day'
PN41='Smoked Tobacco Products - Night'
PX42='Enclosed Area w/Smoker(s)'
PD42='Enclosed Area w/Smokers - Day'
PN42='Enclosed Area w/Smokers - Night'
PX43='% Time Indoors at Home with Smokers'
PX44='% Time Indoors Not at Home w/ Smokers'
PX45='% Time Enclosed in Transit w/ Smokers'
HG41='Pesticides in Home'
HG42='Drycleaned Clothes in Home'
HG43='Mothballs/Crystals, Fresh/Deodorizers'
HX44='Shower/Bath'
PG46='Work with -cides'
PX47='Insecticides, Pesticides, Herbicides'
PX48='Gas Stat/Park Garage/Car Repair Shop'
PX49='Pumped Gas'
PD49='Pumped Gas - daytime'
PN49='Pumped Gas - nighttime'
PX50='Drycleaning'
PD50='Drycleaning - daytime'
PN50='Drycleaning - nighttime'
PX51='Wear Drycleaned Clothes'
PX52='Paints/Solvents'
PX53='Odorous Chemicals (Incl. Mothballs)'
PX54='Mothballs/Crystals, Fresh/Deodorizers'
PX55='Odorous, vaporizing glues/adhesives'
PX56='Toxic or Hazardous Chemicals'
PX57='Auto/Truck Exhaust'
PX58='Cleaning Solutions'
PX59='Degreasing Compounds'
PG52='Solvents'
PG53='Odorous Chemicals (Incl. Mothballs)'
PG56='Toxic or Hazardous Chemicals'
PG57='Auto/Truck Exhaust'
PG58='Cleaning Solutions'
PG59='Degreasing Compounds'
PX60='Flea Collars, Powder, Pet Shampoo'
PX61='Aerosol Personal Care Products'
PX62='Polishing/Waxing Agents'
PX63='Shower/Bath'
PG64='Drink Municipal Water'
PG65='Drink Bottled Water'
PX66='Worked at Regular occupation'          /* Yes, No, Unemployed */
PG67='Av % Day Away from Home-Weekday'
PG68='Av % Day Outdoors-Weekday'
PG69='Av % Day In Motor Vehicle-Weekday'
PG70='Av % Day Away from Home-Weekend Day'
```

```
PG71='Av % Day Outdoors-Weekend Day'
PG72='Av % Day In Motor Vehicle-Weekend Day'
PX73='% Mon.Period Indoors at Home'
PX74='% Mon.Period Indoors at Work (not home)'
PX75='% Mon.Period Indoors Not Work or Home'
PX76='% Mon.Period Outdoors at Work'
PX77='% Mon.Period Outdoors Not at Work'
PX78='% Mon.Period Indoors Not at Home'
PX79='% Mon.Period Outdoors'
PX80='% Mon.Period Enclosed in Transit';
```

VOLUME II: DATA ANALYSES

TABLE OF CONTENTS

	Page
ABSTRACT	iii
LIST OF EXHIBITS	iv
LIST OF TABLES	v
GLOSSARY OF TERMS, ABBREVIATIONS, AND SYMBOLS	vi
ACKNOWLEDGEMENTS	vii
SECTION 1 - INTRODUCTION	1
1.1 Background	1
1.2 Purpose and Organization of Volume II	2
SECTION 2 - SUMMARY AND CONCLUSIONS	4
SECTION 3 - RECOMMENDATIONS	10
SECTION 4 - ANALYSIS TYPE A: WEIGHTED ANALYSES FOR A SINGLE STUDY	11
4.1 Example A1: Personal Air 24-h Concentrations for Study 1	11
4.1.1 Analysis Objectives and Specifications	11
4.1.2 Analysis Results	12
4.2 Example A2: Indoor Air 24-h Concentrations for Study 6	24
4.2.1 Analysis Objectives and Specifications	24
4.2.2 Analysis Results	24
SECTION 5 - ANALYSIS TYPE B: WEIGHTED ANALYSES FOR SEVERAL INDEPENDENT POPULATIONS	35
5.1 Analysis Objectives and Specifications	35
5.2 Analysis Results	36
SECTION 6 - ANALYSIS TYPE C: ANALYSES FOR PAIRED OBSERVATIONS	45
6.1 Example C1: Study 4 vs. 5 Personal, Indoor, and Outdoor 24-h Concentrations	45
6.1.1 Analysis Objectives and Specifications	45
6.1.2 Analysis Results	46
6.2 Example C2: Study 1 vs. 2 Personal 24-h Concentrations	55
SECTION 7 - ANALYSIS TYPE D: ANALYSES INVOLVING DATA FROM MULTIPLE COMPOUNDS	56
7.1 Analysis Objectives and Specifications	56
7.2 Analysis Results	57
SECTION 8 - ANALYSIS TYPE E: ANALYSES BASED ON GENERAL LINEAR MODELS	71
8.1 Analysis Objectives and Specifications	71
8.2 Analysis Results	73
SECTION 9 - REFERENCES	88

TABLE OF CONTENTS (continued)

Page

APPENDICES

II-A TYPE A ANALYSIS..... A-1

II-B. TYPE B ANALYSIS..... B-1

II-C. TYPE C ANALYSIS..... C-1

II-D. TYPE D ANALYSIS..... D-1

II-E. TYPE E ANALYSIS..... E-1

ABSTRACT

The recent Woodland, CA study, sponsored by the Air Resources Board (ARB), and the series of Total Exposure Assessment Methodology (TEAM) studies, conducted in California and sponsored by the U.S. Environmental Protection Agency and the ARB, furnish residential concentration data on a number of important volatile organic compounds (VOCs), along with concomitant information. While there are some differences in the datasets associated with these studies that hinder the direct comparison of their results, there are some data analyses that can be justified and that could provide insight into indoor air pollutant levels, trends, and sources. To do so requires the consolidation of data from these various studies into a unified data base organized in a manner amenable to statistical analysis. To develop such a database was the first objective of this study. Volume I documents its structure and contents, which includes data from six studies (three 1984 TEAM studies, two 1987 follow-up TEAM studies, and the 1990 Woodland Study). The database contains five primary data files: a file derived from questionnaire response information, two files of outdoor-, indoor-, and personal-air VOC concentrations (one for daytime and nighttime monitoring periods, and one for a 24-hour monitoring period), a file of tap water VOC concentrations, and a file of breath-sample VOC concentrations.

The second study objective, to perform and document data analyses supported by these data, is addressed in this Volume II: it furnishes both substantive information and illustrations of various types of analyses that can be performed. The analyses include generation of univariate descriptive statistics and associated comparisons (weighted analyses for a single study population or for several independent study populations, and weighted or unweighted analyses of temporally paired observations), correlation analyses, and model-based analyses (regression, analysis of variance and covariance). The primary limitation of the analyses stems from the lack of adequate residential indoor VOC concentration information in the database. It is recommended that future large-scale residential studies attempt to generate more indoor data and data that are compatible with the database developed as a part of this study.

LIST OF EXHIBITS

Exhibit		Page
4-1	Analysis Type A: Study 1 (1984 LA County, Winter) -- 24-h Personal Air Concentrations	15
4-2	Analysis Type A: Study 6 (May-June 1990, Woodland) -- 24-h Indoor Air Concentrations	26
5-1	Analysis Type B: Study 1 vs. Study 3 vs. Study 6 -- 24-h Personal Air Concentrations	38
6-1	Analysis Type C: Study 4 vs. Study 5 Comparison	48
7-1	Analysis Type D: Intra-Media Correlation Analysis -- All Studies	58
8-1	Analysis Type E: Final Regression Results for Y3 (Log Indoor Concentration), By Compound	78

LIST OF TABLES

Table		Page
4-1	Summary of Estimated Population Distributions of 24-h Personal Air Exposures (μg/m3) -- Based on Winter 1984 LA County TEAM Data	23
4-2	Estimated Population Geometric Means of 24-h Personal Air Exposures (μg/m3), By Smoking Category -- Based on Winter 1984 LA County TEAM Data	23
4-3	Summary of Estimated Population Distributions of 24-h Indoor Air Concentrations (μg/m3) -- Based on May-June 1990 Woodland Study Data	34
4-4	Estimated Population Geometric Means of 24-h Indoor Air Concentrations (μg/m3), By Home's Smoking Status -- Based on May-June 1990 Woodland Study Data	34
5-1	Estimated Population Geometric Means of 24-h Personal Air Exposures (μg/m3), By Study	44
5-2	Estimated Geometric Means of 24-h Personal Air Exposures (μg/m3) for Subpopulations Not Exposed to Tobacco Smoke, By Study	44
6-1	Study 4 vs. Study 5 Paired Comparisons for Means of Logarithms of 24-h Indoor Concentrations and for Means of Logarithms of 24-h Indoor/Outdoor Ratios	54
7-1	Pearson Correlations Between Log Concentrations, By Media, Based on All Data Available in the Six Studies	70
8-1	Variables and Models Used for Analysis Type E Example	76
8-2	Analysis Type E: Summary of Preliminary, Full-Model Regression Results	77
8-3	Analysis Type E: Summary of Final Regression Results for Y3 (Log Indoor Concentration), By Compound	87

GLOSSARY OF TERMS, ABBREVIATIONS, AND SYMBOLS

ARB	Air Resources Board
ASCII	American Standard Code for Information Interchange
C.V.	Coefficient of variation
EPA	United States' Environmental Protection Agency
ETS	Environmental tobacco smoke
GM	Geometric mean
h	Hour
I/O	Indoor/outdoor
$\mu g/m^3$	micrograms per cubic meter
PSU	Primary sampling unit
QL	Quantifiable limit
RAE	Record of Activities and Environments
SAS	Statistical Analysis System
SD	Standard deviation
SUDAAN	Survey Data Analysis
TAD	Time Activity Diary
TEAM	Total Exposure Assessment Methodology
VOC	Volatile organic compound

1. INTRODUCTION

1.1 Background

Volume I of this report (Clayton and Perritt, 1993) provides a brief description of six California studies that generated residential measurements on a number of important volatile organic compounds (VOCs), along with concomitant information, and describes the structure and content of a database containing the studies' data. The six studies are designated as follows:

(1) Los Angeles (LA) County - February-March 1984
(2) Los Angeles County - May-June 1984
(3) Pittsburg/Antioch - June 1984
(4) Los Angeles County - January-February 1987
(5) Los Angeles County - June-July 1987
(6) Woodland - May-June 1990.

Information on the target populations, the sample designs and participant selection, and the relationship of the six studies is included in Volume I, Section 1. As indicated in Figure 1 of Volume I, the first three studies constitute a group of TEAM (Total Exposure Assessment Methodology) studies that were conducted in 1984 using the same procedures and data collection instruments. Similarly, the other two studies in LA County (Studies 4 and 5) were conducted as a group in 1987. These five studies share many common design features, while the Woodland Study, conducted later, had a more varied set of objectives than the previous five. The four LA County studies were not independent studies in that the households/participants included in the Study 2, 4, and 5 samples were (essentially) subsamples of the Study 1 sample.

Volume I also describes the types of data available, which are summarized below:

FILE NAME	DESCRIPTION OF FILE CONTENTS
QUEST.DAT	General household and personal characteristics, based on responses to Study Questionnaires (Studies 1-6); data on activities, microenvironments, and potential VOC exposures during the monitoring period, based on responses to the Record of Activities and Environments (RAE) (Studies 1-6) and to the Time Activity Diary (TAD) (Study 6)

DAYNITE.DAT	Personal, indoor, and outdoor air VOC concentration data from two consecutive monitoring periods (nighttime and daytime) -- for Studies 1-5
HOUR24.DAT	Personal, indoor, and outdoor air VOC concentration data for 24-hour monitoring periods -- constructed as a weighted average of nighttime and daytime concentration levels for Studies 1-5, and obtained directly from the 24-hour monitoring for Study 6
WATER.DAT	VOC concentration data from drinking water samples -- for Studies 1 through 5
BREATH.DAT	VOC concentration data from breath samples collected at the end of the daytime monitoring period -- for Studies 1 through 5.

1.2 Purpose and Organization of Volume II

Volume II has two purposes:

(1) to illustrate the types of statistical analyses that can be applied to these data and thereby furnish some substantive results, and

(2) to demonstrate the development of software for preparing the working data files for analysis and for carrying out the analysis.

The first purpose is addressed in the main body of the report in Sections 4 through 8, which present a series of example analyses and which follow Section 2, a brief "Summary and Conclusions" section, and Section 3, "Recommendations." The various types of analyses have been grouped into five categories, primarily on the basis that analyses of different types will require either different strategies for building the working file(s) or different types of analytical software. These examples are each discussed in two subsections: (1) analysis objectives and specifications, and (2) analysis results. The first of these subsections includes an outline of the key components of the analysis (spatial and temporal bounds, compounds, media, types of statistics, etc.), such as an outline that an analyst might prepare in response to a request from an end-user of the results. The second subsection of each

example includes an exhibit that shows the computer-generated outputs of the analysis and a brief discussion of the results.[1]

The second purpose of the report is treated primarily in the appendices, with program listings and associated discussion of the programs appearing in Appendices II-A through II-E. These correspond to Analysis Types A through E (discussed in Sections 4 through 8, respectively). All programs are written using SAS[2] or a combination of SAS and SUDAAN.[3] In each case, there are at least two main components to the programs: a section of SAS code for preparing the data files for analysis (e.g., extracting the pertinent data records, selecting variables, merging of working files), and a section of either SAS or SUDAAN code for carrying out the particular analysis. In some cases, there is also SAS code for an additional post-processing step that involves further calculations (when needed) and rearrangement and printing of the results.

---

[1]While some reformatting of the results may be desirable from the point of view of a reader interested in the substantive results, the computer-generated format is used because it is convenient for analysts attempting to relate the program logic to the program outputs.

[2]SAS is the registered trademark of SAS Institute, Inc., Cary, NC.

[3]SUDAAN is a software product developed by RTI for conducting weighted analyses on data arising from probability-based surveys in which sample members may have different selection probabilities. SUDAAN generates standard error approximations for estimated means, proportions, etc. that appropriately account for the sampling design structure. See Shah et al., 1993, for mathematical details.

2. SUMMARY AND CONCLUSIONS

The first study objective was to consolidate data from the various studies into a unified database amenable to statistical analysis. Volume I of this report (Clayton and Perritt, 1993) documented the structure and content of this database. The second objective of the study -- to perform and document data analyses supported by the data -- is addressed in this volume. It not only furnishes some substantive results but also illustrates various types of analyses that can be conducted and provides the software (SAS or a combination of SAS and SUDAAN) for carrying out these examples.

The types of data analyses illustrated are the following:

- Analysis Type A: Weighted Analyses for a Single Study -- This analysis illustrates how SUDAAN is used to perform a weighted data analysis for a single study. In particular, two examples are presented (Section 4 and Appendix II-A) that show how to produce descriptive statistics that represent estimates of population parameters such as the population mean, population geometric mean, or population percentiles. The examples also show how similar estimates for selected domains (subpopulations) can be generated and compared.

- Analysis Type B: Weighted Analyses for Several Independent Populations -- This analysis, illustrated in Section 5 and Appendix II-B, extends Type A to several independent (or independently sampled) populations. It is also a weighted analysis, and depending on how the various populations (studies) relate to one another, the analysis may be interpreted as representing geographic, seasonal, and/or annual comparisons. Such comparisons may be made overall or for comparably defined domains of each study's population.

- Analysis Type C: Analyses for Paired Observations -- This type of analysis is appropriate when a target population is sampled on more than one occasion and (some of) the same sample members are used each time. In this case, the set of observations are not independent, but are paired. If the time points of sampling represent different seasons (years), then the results are used to assess seasonal (annual) trends. If both samples are probability based, then SUDAAN is the appropriate software (a weighted analysis of differences and ratios). On the other hand, if either of the paired samples is not a probability-based sample (e.g., if a purposefully selected subsample of the first sample is chosen at the second time point), then SAS is used to perform an unweighted analysis of differences and ratios. Section 6 (Appendix II-C) presents an example of the Type C analysis.

- Analysis Type D: Analyses Involving Data from Multiple Compounds -- This type represents a broad range of analyses, such as correlation analyses, principal components analyses, and factor analyses, that are multivariate in nature and which have a general goal of gaining

an understanding of how concentration levels of different compounds and/or media relate to one another. These analyses may aid also in understanding activities and microenvironments that may influence them. These analyses are grouped together since they necessitate the same restructuring of the data files prior to statistical analysis. These analyses, which would typically be performed using SAS, may be simultaneously applied to data from multiple studies. Section 7 (Appendix II-D) provides an example of the correlation analyses.

- Analysis Type E: Analyses Based on General Linear Models -- These analyses include multiple linear regression analysis and analysis of variance and covariance -- analyses that seek to identify factors that influence concentration levels (or functions of concentration levels) and to estimate and test such effects. In general, the source data for the dependent variables will be the chemical data and the source data for (most of) the independent variables will be the questionnaire data. In most cases, SAS would be used to perform unweighted, model-based analyses, though under some circumstances, SUDAAN (PROC REGRESS) can be used to perform a weighted, probability-based analysis. Section 8 (Appendix II-E) provides an example of the unweighted type of analysis. The Type E analyses are more general than the domain comparisons conducted as a part of Analysis Types A and B in that multiple independent variables can be analyzed simultaneously. Like Type B, however, they may be used to assess geographic and/or temporal effects by including such effects in the statistical model, perhaps along with other explanatory variables.

Example A1 (Section 4.1) involved producing estimates to characterize the Study 1 (Winter 1984 LA County) 24-h personal air concentration distributions of six selected compounds (1,1,1-trichloroethane, benzene, tetrachloroethylene [perc], styrene, m,p-xylene, and m,p-dichlorobenzene).[4] These distributions were found to be highly skewed to the right, although the degree of skewness varied considerably among compounds (benzene and m,p-xylene showed the least, and 1,1,1-trichloroethane and m,p-dichlorobenzene exhibited the most). The results indicated that the median or geometric mean (GM) would be preferred over the mean as a measure of central tendency of the distributions. Parameters were also estimated separately for three subpopulations: active smokers, those exposed to environmental tobacco smoke (ETS) (i.e., passive smokers), and others (i.e., the "unexposed").

---

[4]These compounds were selected based on the ARB's interest and on the results presented in Table 3 of Volume I, which indicated that most observed concentrations for these six chemicals exceeded quantifiable limits (QLs).

Tests of the pairwise differences between the GMs revealed significantly higher GMs for the active-smoking category -- for five of the compounds relative to the unexposed category, and for three compounds (benzene, styrene, and m,p-xylene) relative to the passive exposure category.

Example A2 (Section 4.2) produced population and subpopulation estimates for 24-h indoor air concentrations for Study 6 (Woodland, 1990). The estimated mean concentrations for the six selected compounds tended to be substantially larger than either the medians or the geometric means (implying skewness); also, coefficients of variation (C.V.) were quite large. As with the prior example, the medians and GMs will therefore be better than the means for characterizing the central tendency of the distributions and will also tend to be more precisely estimated. Little or no differences in "average" indoor levels were detected between those homes in the smoking category versus those in the nonsmoking category.

Example B (Section 5) compared the 24-h personal air concentration distributions for Studies 1, 3, and 6 -- that is, for three studies for which independent, probability-based samples were available. Study 1 (LA County, Winter 1984) consistently exhibited higher GMs than either of the other studies (only the m,p-dichlorobenzene difference for Studies 1 and 6 was not detected as significant). Comparison of the GMs for Study 3 (Pittsburg/Antioch, June 1984) and for Study 6 (Woodland, May-June 1990) showed an inconsistent pattern: two compounds had significantly higher levels in Woodland, two had significantly lower levels, and two yielded nonsignificant differences. Similar comparisons of GMs from the non-ETS-exposed subpopulations of each of the three studies (i.e., excludes active smokers and those exposed to tobacco smoke) showed the same basic pattern of results.

Example C (Section 6) compared personal air, indoor air, and outdoor air levels for the two 1987 LA County Studies -- Study 4 (winter) versus Study 5 (summer). For each participant, logarithms of the winter/summer ratios of the corresponding concentrations (or of indoor/outdoor (I/O) ratios) were computed and the mean over participants was determined. A t test for such a mean was used to provide an approximate test of season differences. For indoor levels, five of six compounds had higher levels in the winter; for I/O ratios, only styrene showed a significant

difference (in the opposite direction). These results must be viewed with caution due to the small sample sizes and the nature of the samples.

To exemplify Analysis Type D (Section 7), Pearson correlation matrices for personal, indoor, and outdoor 24-h concentration data of six selected compounds (original and log scale) were developed using all available data. The log-scale correlations were generally larger and more consistent than the corresponding correlations on the original concentration scale. In general, personal and indoor log-scale correlations were quite comparable and were smaller than those for the outdoor air. The lowest log-scale correlations among indoor concentrations occurred for p-dichlorobenzene with the other compounds (ranging from 0.18 to 0.34); the highest such correlation was between benzene and m,p-xylene.

Example E (Section 8) used data from Studies 4, 5, and 6 to illustrate an exploratory analysis directed at identifying and quantifying effects that may influence indoor air concentration levels of contaminants. For each of nine VOCs, the log of the indoor concentration was modeled using the log of the outdoor concentration as a covariate and using a series of indicator variables derived from questionnaire responses as other explanatory variables. Indicator variables representing study effects (geographic and temporal effects) were also included as candidate variables in each compound's model. Stepwise regression was used to reduce the number of explanatory variates to those pertinent to the particular compound. Model performance was poor for two compounds (chlorobenzene: $R^2 = 0.13$; trichloroethylene: $R^2 = 0.23$) that had relatively low percentages of quantifiable values. For the other seven, R^2 values ranged from 0.42 to 0.70. In addition to the outdoor concentration covariate, which in all cases was a strong predictor, there were a number of other effects found to be statistically significant -- some anticipated and some not.

As indicated by the above discussion of analysis types and results, addressing even a single type of analysis objective may require different types of analyses, depending on the structure of the data (e.g., type of sampling design, independence/dependence of observations) and on what models and assumptions the analyst is required (or willing)

to make. For example, one may be interested in whether there are geographic differences or seasonal or other temporal trends in the (indoor, say) concentration levels. These types of objectives were addressed in three of the above analysis types. Type B's comparisons, for instance, entailed geographic comparisons, based upon data from three independent studies in which separate, probability-based samples had been taken. Type B could also be used to assess temporal trends, if independent samples had been drawn at two or more times from the same region. Type C comparisons took advantage of the pairwise nature of the sampling design for Studies 4 and 5 (i.e., the same houses and persons sampled on more than one occasion) to make a seasonal comparison. Finally, the Type E analysis examined differences among indoor concentrations for three studies representing two different seasons and two locations. This analysis made such comparisons by incorporating these effects into the statistical model; hence in this case the estimated geographic/temporal effects are adjusted for the other effects in the model, in contrast to the comparisons made via analysis Types B and C.

Thus, while the database affords the opportunity to investigate a number of important issues, analysts must be careful in choosing an appropriate strategy (e.g., in choosing among Types B, C, and E, and in choosing a weighted versus unweighted analysis). The table below provides an overview of the possibilities with regard to weighting:

TYPE OF DATA	STUDY 1,2,3	STUDY 4,5	STUDY 6
Questionnaire Responses	W	U	W
Breath Concentrations	W	U	-
Water Concentrations	W	U	-
12-h Personal Air Concentrations	W	U	-
12-h Indoor Air Concentrations	-	U	-
12-h Outdoor Air Concentrations	W*	U	-
24-h Personal Air Concentrations	W	U	W
24-h Indoor Air Concentrations	-	U	W
24-h Outdoor Air Concentrations	W*	U	W

W = weighted analysis can be done; U = only unweighted analysis is possible; * = unweighted analysis is recommended due to the small sample size.

Statistical inferences to target populations (e.g., reporting of a mean concentration level) are not valid for the "U" cases; rather, those cases are primarily useful for assessing trends (by making use of paired observations). Comparative analyses (e.g., comparing means) will be most meaningful when the generated statistics represent a well-defined target population -- that is, when the sampling weights in the "W" cases are appropriately utilized. There may be some instances in which the weights are ignored. These cases include exploratory analyses aimed at exploring or understanding relationships and associations among variables, especially if data from Studies 4 and/or 5 (i.e., the "U" cases) are used in conjunction with data from one or more of the other studies. The above-described examples for Analysis Types D and E illustrate such situations.

The above table also makes it evident that a number of analyses that may be desired will not be possible due to the lack of available data. This is especially true for indoor concentration data, which are available only for Studies 4, 5, and 6.

In addition to making appropriate use of the sampling weights, analysts making use of the chemical data must be careful in selecting compounds for which results will not be highly influenced by values below QLs -- that is, by avoiding chemicals with a low percentage of quantifiable values. The compounds used in the examples described in this report, for instance, were selected primarily because most of their measurements were quantifiable. Table 3 of Volume I provides guidance in this regard. If analyses must be performed for low-percent-quantifiable compounds, it may be useful to consider several ways of dealing with values below the limits in order to understand the degree of influence such values may have (see Gilbert, 1987, Chapter 14, for instance).

3. RECOMMENDATIONS

Because of the paucity of residential indoor VOC concentration data from large-scale studies, the ARB's capacity for quantitatively assessing levels, sources, and trends in such data are severely hampered. Additional VOC indoor air studies are needed to more adequately address such issues. In designing such studies, the database developed as a part of the present study should be considered, so that, to the extent feasible and reasonable, the data resulting from future studies will be compatible with its contents and thus can be augmented to it. If future studies make significant augmentations to the database, it would then be reasonable to consider other enhancements, such as increasing user-friendliness (e.g., interactive selection of studies, compounds, variables, etc.).

4. ANALYSIS TYPE A: WEIGHTED ANALYSES FOR A SINGLE STUDY

Analysis Type A is designed to illustrate how SUDAAN is used to perform a weighted data analysis for data from a single study. The analysis objective is to produce probability-based estimates of parameters for the target population and for specified domains of the population and to compare the parameters of the different domains. In this section, two examples are employed to illustrate how to produce such descriptive statistics, using SUDAAN's DESCRIPT procedure, that represent estimates of parameters such as a population mean, a population geometric mean, or a selected population percentile. Estimates are produced also for several domains related to environmental tobacco smoke (ETS) exposures.

4.1 Example A1: Personal Air 24-h Concentrations for Study 1

4.1.1 Analysis Objectives and Specifications

Example A1 involves producing estimates for the Study 1 (1984 LA County) 24-h personal air concentration distribution. The results are to be produced for the six selected compounds identified below. In addition to producing the estimates for the overall population, such parameters are to be estimated separately for three subpopulations: active smokers, passive smokers, and others (i.e., those not exposed to tobacco smoke). The means and geometric means (GMs) for the different subpopulations are to be compared on a pairwise basis (i.e., active vs. passive, active vs. unexposed, and passive vs. unexposed). Selected percentiles of the distribution are also to be estimated.

The specifications for the analysis are outlined below:

Geographic and Temporal Bounds: Study 1

Compounds: CMPD = 4, 5, 11, 14, 21, and 16

(1,1,1-trichloroethane, benzene, tetrachloroethylene [perc], styrene, m,p-xylene, and m,p-dichlorobenzene)

Media: Personal air

Time Frame: 24-hr

Population Estimates to Be Generated: Mean, Geometric Mean, Population Variance, Percentiles (10, 25, 50, 75, 90) of Personal Air 24-Hr Concentrations

Domains for Which Estimates Are to Be Generated:

(1) Overall Population

(2) Three Domains (Smokers, Others Exposed to ETS, Others Not Exposed to ETS) -- Variables PX41 and PX42 of the QUEST.DAT file are to be used to define these domains.

Type of Analysis: Weighted Analysis Using SUDAAN

4.1.2 Analysis Results

Analyses for Example A1 were accomplished by sequentially executing three programs. The first was a SAS program for preparing the input files and, in particular, for creating a new variable called SMOKE from PX41 and PX42 whose values indicate the status of each participant with respect to the three domains of interest (1=active smoker, 2=passive smoke exposure, 3=not exposed). The second program was a SUDAAN program for performing the calculations required for the statistical analysis. The third program was written in SAS and used for post-processing and printing of the results. A listing of the program logs is included in Appendix II-A, Exhibit A-1.

Exhibit 4-1 shows the tabular outputs produced from the above-described analysis steps. All concentration estimates are expressed in $\mu g/m^3$. The three portions of output are labeled as follows:

1) POPULATION ESTIMATES: MEANS, GEOMETRIC MEANS, AND STANDARD DEVIATIONS

2) POPULATION ESTIMATES: DIFFERENCES IN MEANS & GEOMETRIC MEANS, and

3) POPULATION ESTIMATES: PERCENTILES

The quantities shown in Exhibit 4-1 are defined below:

N = sample size

WSUM = sum of sampling weights (estimated population size)

CM = estimate of population mean

CMSE = estimated standard error for CM

CGM = estimate of population GM

CGMSE = estimated standard error for CGM

POPSD = estimate of population SD

CMDIF = estimated difference in two domain means

CMDIFSE = estimated standard error for CMDIF

CGMDIF = estimated difference in two domain GMs

CGMDIFSE = estimated standard error for CGMDIF

PCTILE = identification of percentile (10, 25, 50, 75, or 90) being estimated

QTILE = estimate of the selected population percentile

SEQTILE = estimated standard error for QTILE

LOWQTILE = lower 95% confidence bound for population percentile

UPQTILE = upper 95% confidence bound for population percentile.

For some combinations of compounds, domains, and percentile values, note that Exhibit 4-1 does not furnish values for SEQTILE and LOWQTILE or UPQTILE. This occurs because of the manner in which SUDAAN computes the confidence intervals for the percentile estimates -- namely, by inverse interpolation of the confidence bounds on the empirical distribution function. This sometimes results, especially when sample sizes are small, in an interval that is unbounded on one end. In those cases, no value can be provided for that endpoint and, since SEQTILE is computed from the LOWQTILE and UPQTILE values, no value is provided for SEQTILE. Computational details are given in Shah et al., 1993 (pages 44-48).

Some of the more interesting and important results contained in Exhibit 4-1 were extracted and are shown in Tables 4-1 and 4-2 (see page 23).[5] With the exception of the coefficient of variation (C.V.) shown in Table 4-1, all of the numerical results appeared directly in the exhibit.

Table 4-1 contains information relating to the overall population distribution of 24-h exposures. These distributions are obviously highly skewed to the right. This is apparent from the magnitude of the means relative to the corresponding GMs and medians, as well as from the

---

[5]The intent of preparing such tables and of providing some related discussion is to illustrate how the computer-generated results can be used to fulfill the needs of researchers and decision makers. The intent is not to provide a complete discussion of the results.

magnitude of the standard deviations to the means (and also from the percentile estimates themselves). The degree of skewness varies considerably among the six compounds, with benzene and m,p-xylene showing the least degree of skewness and with 1,1,1-trichloroethane and m,p-dichlorobenzene exhibiting substantially more skewness. These results indicate that the median or the GM are preferred over the mean as a measure of central tendency of the distributions. (See Gilbert [1987], for example, for a discussion of the effects of skewness on measures of central tendency and on the benefits of logarithmic transformations in such cases.)

Table 4-2 presents (selected) results for the three smoking categories -- namely, the GMs and the pairwise differences between the GMs. By comparing the differences to their standard errors, an approximate test of statistical significance can be performed (significance at the 0.05 level is denoted by asterisks in Table 4-2) to determine if the domains' exposures are or are not the same. Whereas the differences between the passive-exposed and unexposed categories were not found significant for any of the six compounds, the GMs for the active-smoking category were significantly different (higher) than those for the unexposed category for five of the six compounds (m,p-dichlorobenzene being the exception) and significantly different (higher) than the passive-smoking category for three compounds (benzene, styrene, and m,p-xylene).

Exhibit 4-1. Analysis Type A: Study 1 (1984 LA County, Winter) -- 24-h Personal Air Concentrations

POPULATION ESTIMATES: MEANS, GEOMETRIC MEANS, AND STANDARD DEVIATIONS

---------- TABLE=1: OVERALL POPULATION ----------

C	SMOKE	N	WSUM	CM	CMSE	CGM	CGMSE	POPSD
1,1,1-Trichloroethane		111	359492	90.69	24.74	37.03	4.62	251.22
Benzene		111	359492	17.64	1.18	15.38	1.13	9.17
Tetrachloroethylene		110	359492	15.26	1.50	10.23	0.81	19.58
Styrene		110	359492	3.54	0.25	2.74	0.27	2.43
m,p-Dichlorobenzene		109	359492	18.71	4.04	4.06	0.74	44.81
m,p-Xylene		110	359492	27.50	1.68	23.98	1.58	15.76

---------- TABLE=2: BY SMOKING CATEGORY ----------

C	SMOKE	N	WSUM	CM	CMSE	CGM	CGMSE	POPSD
1,1,1-Trichloroethane	ACTIVE	34	106552	212.29	74.79	63.03	15.99	435.61
1,1,1-Trichloroethane	PASSIVE	23	72773	39.19	5.39	29.94	4.87	26.83
1,1,1-Trichloroethane	UNEXPOSED	53	177809	39.90	5.19	29.74	3.74	31.55
Benzene	ACTIVE	34	106552	22.21	2.12	20.07	1.83	9.94
Benzene	PASSIVE	23	72773	16.17	1.52	14.80	1.58	6.21
Benzene	UNEXPOSED	53	177809	15.62	1.54	13.44	1.52	8.75
Tetrachloroethylene	ACTIVE	34	107471	20.52	3.36	13.57	1.72	23.09
Tetrachloroethylene	PASSIVE	22	70533	15.52	5.69	9.00	1.51	26.50
Tetrachloroethylene	UNEXPOSED	53	179014	12.11	1.84	9.10	1.12	12.06
Styrene	ACTIVE	34	107471	4.83	0.42	4.30	0.31	2.39
Styrene	PASSIVE	22	70533	3.18	0.51	2.54	0.36	2.19
Styrene	UNEXPOSED	53	179014	2.93	0.35	2.16	0.38	2.24
m,p-Dichlorobenzene	ACTIVE	33	106600	9.98	3.24	3.80	0.71	19.23
m,p-Dichlorobenzene	PASSIVE	22	70746	22.88	10.23	4.57	1.66	45.52
m,p-Dichlorobenzene	UNEXPOSED	53	179671	22.51	8.23	4.13	1.12	53.95
m,p-Xylene	ACTIVE	34	107471	33.13	3.17	29.99	2.29	18.31
m,p-Xylene	PASSIVE	22	70533	26.44	2.26	23.54	2.30	12.58
m,p-Xylene	UNEXPOSED	53	179014	24.74	2.06	21.28	2.11	14.31

Exhibit 4-1. (continued)

POPULATION ESTIMATES: DIFFERENCES IN MEANS & GEOMETRIC MEANS

C	CONTRST	N	WSUM	CMDIF	CMDIFSE	CGMDIF	CGMDIFSE
1,1,1-Trichloroethane	ACTIVE-PASSIVE	57	179325	173.10	76.21	33.09	17.17
1,1,1-Trichloroethane	ACTIVE-UNEXPOSED	87	284361	172.38	74.47	33.29	14.45
1,1,1-Trichloroethane	PASSIVE-UNEXPOSED	76	250582	-0.71	7.27	0.20	5.58
Benzene	ACTIVE-PASSIVE	57	179325	6.05	2.49	5.27	2.17
Benzene	ACTIVE-UNEXPOSED	87	284361	6.59	2.46	6.63	2.29
Benzene	PASSIVE-UNEXPOSED	76	250582	0.54	1.98	1.36	1.99
Tetrachloroethylene	ACTIVE-PASSIVE	56	178003	4.99	7.72	4.57	2.79
Tetrachloroethylene	ACTIVE-UNEXPOSED	87	286484	8.41	3.40	4.47	1.60
Tetrachloroethylene	PASSIVE-UNEXPOSED	75	249546	3.42	6.71	-0.10	2.00
Styrene	ACTIVE-PASSIVE	56	178003	1.65	0.67	1.76	0.51
Styrene	ACTIVE-UNEXPOSED	87	286484	1.90	0.56	2.14	0.52
Styrene	PASSIVE-UNEXPOSED	75	249546	0.25	0.67	0.38	0.50
m,p-Dichlorobenzene	ACTIVE-PASSIVE	55	177346	-12.90	11.21	-0.77	1.73
m,p-Dichlorobenzene	ACTIVE-UNEXPOSED	86	286271	-12.53	8.51	-0.32	1.18
m,p-Dichlorobenzene	PASSIVE-UNEXPOSED	75	250417	0.37	14.83	0.45	1.87
m,p-Xylene	ACTIVE-PASSIVE	56	178003	6.69	3.75	6.45	3.00
m,p-Xylene	ACTIVE-UNEXPOSED	87	286484	8.39	3.56	8.71	2.91
m,p-Xylene	PASSIVE-UNEXPOSED	75	249546	1.70	2.81	2.26	2.80

Exhibit 4-1. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=1,1,1-Trichloroethane TABLE=1: OVERALL POPULATION ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	111	359492	13.35	1.77	10.76	18.18
	25	111	359492	18.87	2.69	15.52	26.77
	50	111	359492	32.02	4.69	23.78	43.42
	75	111	359492	61.82	8.13	44.49	78.51
	90	111	359492	102.20	73.81	81.45	390.41

---------- C=1,1,1-Trichloroethane TABLE=2: BY SMOKING CATEGORY ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
ACTIVE	10	34	106552	12.90			30.31
ACTIVE	25	34	106552	26.74	5.60	14.67	38.13
ACTIVE	50	34	106552	42.94	8.80	32.26	69.08
ACTIVE	75	34	106552	97.92	70.18	55.38	349.15
ACTIVE	90	34	106552	673.69		102.83	
PASSIVE	10	23	72773	12.76			16.01
PASSIVE	25	23	72773	14.56	3.24	13.33	26.89
PASSIVE	50	23	72773	28.60	7.80	18.46	51.11
PASSIVE	75	23	72773	52.62	12.82	29.49	83.14
PASSIVE	90	23	72773	83.72		51.13	
UNEXPOSED	10	53	177809	12.01			18.52
UNEXPOSED	25	53	177809	18.60	1.97	15.64	23.90
UNEXPOSED	50	53	177809	27.25	4.12	20.32	37.57
UNEXPOSED	75	53	177809	51.12	10.34	32.44	75.73
UNEXPOSED	90	53	177809	84.55		57.02	

NOTE: Confidence intervals for percentiles are obtained by inverse interpolation of the confidence bounds on the empirical distribution function. When sample sizes are small, missing values may occur because the interval is unbounded (on one side).

Exhibit 4-1. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=Benzene TABLE=1: OVERALL POPULATION ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	111	359492	7.48	1.12	5.19	9.89
	25	111	359492	10.87	0.93	9.61	13.49
	50	111	359492	15.83	1.16	13.52	18.37
	75	111	359492	22.21	1.89	18.23	26.14
	90	111	359492	30.49	2.66	24.55	35.69

---------- C=Benzene TABLE=2: BY SMOKING CATEGORY ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
ACTIVE	10	34	106552	9.95	1.55	9.24	15.73
ACTIVE	25	34	106552	14.30	1.30	13.17	18.62
ACTIVE	50	34	106552	19.06	2.68	16.64	27.88
ACTIVE	75	34	106552	29.89	3.57	19.68	34.62
ACTIVE	90	34	106552	34.93		26.29	
PASSIVE	10	23	72773	7.26			12.83
PASSIVE	25	23	72773	11.04	2.14	7.05	15.99
PASSIVE	50	23	72773	15.73	1.94	11.72	19.84
PASSIVE	75	23	72773	19.88	1.99	15.28	23.63
PASSIVE	90	23	72773	23.64		18.25	
UNEXPOSED	10	53	177809	5.46	1.38	4.09	9.87
UNEXPOSED	25	53	177809	9.63	1.47	6.67	12.81
UNEXPOSED	50	53	177809	13.42	1.54	10.13	16.58
UNEXPOSED	75	53	177809	18.30	2.24	15.65	25.03
UNEXPOSED	90	53	177809	26.29		18.37	

Exhibit 4-1. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=Tetrachloroethylene TABLE=1: OVERALL POPULATION ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	110	359492	4.22	0.58	3.66	6.10
	25	110	359492	6.26	0.76	4.86	8.02
	50	110	359492	8.75	0.78	7.75	11.00
	75	110	359492	14.23	1.33	12.04	17.61
	90	110	359492	27.50	8.06	20.24	53.97

---------- C=Tetrachloroethylene TABLE=2: BY SMOKING CATEGORY ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
ACTIVE	10	34	107471	5.16	0.89	4.36	8.07
ACTIVE	25	34	107471	7.51	1.36	5.23	10.90
ACTIVE	50	34	107471	12.08	1.65	8.32	15.21
ACTIVE	75	34	107471	18.56	5.87	12.76	37.32
ACTIVE	90	34	107471	51.89		19.13	
PASSIVE	10	22	70533	2.80			5.13
PASSIVE	25	22	70533	5.01	1.13	3.25	7.97
PASSIVE	50	22	70533	8.72	1.35	5.18	10.82
PASSIVE	75	22	70533	10.94	1.62	8.91	15.71
PASSIVE	90	22	70533	18.04		10.88	
UNEXPOSED	10	53	179014	4.08	0.87	2.79	6.42
UNEXPOSED	25	53	179014	5.90	0.86	4.17	7.78
UNEXPOSED	50	53	179014	8.33	1.11	6.52	11.18
UNEXPOSED	75	53	179014	13.64	2.12	9.92	18.79
UNEXPOSED	90	53	179014	21.17		14.98	

Exhibit 4-1. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=Styrene TABLE=1: OVERALL POPULATION ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	110	359492	1.22	0.32	0.48	1.83
	25	110	359492	1.92	0.12	1.71	2.20
	50	110	359492	2.77	0.17	2.46	3.19
	75	110	359492	4.70	0.50	3.61	5.72
	90	110	359492	6.83	0.68	5.64	8.48

---------- C=Styrene TABLE=2: BY SMOKING CATEGORY ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
ACTIVE	10	34	107471	2.26	0.35	1.50	2.97
ACTIVE	25	34	107471	3.00	0.39	2.45	4.10
ACTIVE	50	34	107471	4.71	0.51	3.46	5.62
ACTIVE	75	34	107471	5.80	0.39	5.29	6.90
ACTIVE	90	34	107471	7.03		5.73	
PASSIVE	10	22	70533	0.69			1.93
PASSIVE	25	22	70533	1.91	0.47	0.82	2.78
PASSIVE	50	22	70533	2.79	0.36	1.95	3.46
PASSIVE	75	22	70533	3.97	0.45	2.81	4.70
PASSIVE	90	22	70533	4.74		3.14	
UNEXPOSED	10	53	179014	0.70	0.37	0.15	1.69
UNEXPOSED	25	53	179014	1.68	0.20	1.26	2.08
UNEXPOSED	50	53	179014	2.21	0.17	1.90	2.61
UNEXPOSED	75	53	179014	3.33	0.52	2.53	4.70
UNEXPOSED	90	53	179014	6.50	1.04	3.66	8.03

Exhibit 4-1. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=m,p-Dichlorobenzene TABLE=1: OVERALL POPULATION ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	109	359492	0.65	0.13	0.49	1.03
	25	109	359492	1.29	0.24	0.89	1.88
	50	109	359492	2.62	0.64	1.86	4.53
	75	109	359492	8.88	2.14	5.86	14.81
	90	109	359492	64.92	18.29	17.49	94.03

---------- C=m,p-Dichlorobenzene TABLE=2: BY SMOKING CATEGORY ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
ACTIVE	10	33	106600	1.27	0.22	0.68	1.58
ACTIVE	25	33	106600	1.77	0.23	1.33	2.29
ACTIVE	50	33	106600	2.55	0.47	1.87	3.84
ACTIVE	75	33	106600	7.29	3.11	2.70	15.73
ACTIVE	90	33	106600	16.21		8.12	
PASSIVE	10	22	70746	0.62			1.00
PASSIVE	25	22	70746	0.98	0.56	0.64	3.00
PASSIVE	50	22	70746	2.79	1.34	1.20	6.80
PASSIVE	75	22	70746	9.38	18.81	2.54	81.29
PASSIVE	90	22	70746	80.49		7.23	
UNEXPOSED	10	53	179671	0.51	0.12	0.49	0.98
UNEXPOSED	25	53	179671	1.00	0.35	0.70	2.16
UNEXPOSED	50	53	179671	3.21	1.14	1.56	6.35
UNEXPOSED	75	53	179671	9.02	9.53	4.25	44.15
UNEXPOSED	90	53	179671	84.02		17.07	

Exhibit 4-1. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=m,p-Xylene TABLE=1: OVERALL POPULATION ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	110	359492	12.38	2.09	8.88	17.64
	25	110	359492	17.90	1.63	14.04	20.86
	50	110	359492	23.03	1.60	21.00	27.68
	75	110	359492	34.98	2.46	28.41	38.70
	90	110	359492	42.70	3.77	38.57	54.35

---------- C=m,p-Xylene TABLE=2: BY SMOKING CATEGORY ----------

SMOKE	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
ACTIVE	10	34	107471	18.93	1.83	15.00	22.66
ACTIVE	25	34	107471	22.66	2.11	20.22	29.07
ACTIVE	50	34	107471	29.26	2.70	23.26	34.54
ACTIVE	75	34	107471	38.38	2.70	29.62	40.90
ACTIVE	90	34	107471	43.39	6.40	38.45	65.24
PASSIVE	10	22	70533	10.81			18.44
PASSIVE	25	22	70533	17.56	2.61	10.46	21.39
PASSIVE	50	22	70533	22.77	1.96	19.81	28.00
PASSIVE	75	22	70533	33.22	2.78	26.09	37.72
PASSIVE	90	22	70533	39.91	4.58	32.89	52.07
UNEXPOSED	10	53	179014	10.08	2.68	6.63	17.84
UNEXPOSED	25	53	179014	14.99	1.82	11.49	19.11
UNEXPOSED	50	53	179014	20.48	1.76	17.64	25.01
UNEXPOSED	75	53	179014	31.36	3.94	21.81	38.29
UNEXPOSED	90	53	179014	42.23		33.33	

TABLE 4-1. SUMMARY OF ESTIMATED POPULATION DISTRIBUTIONS OF 24-H PERSONAL AIR EXPOSURES ($\mu g/m^3$) -- BASED ON WINTER 1984 LA COUNTY TEAM DATA

COMPOUND	Mean	GM	Median	90th %ile	Std. Dev.	C.V. (%)
1,1,1-trichloroethane	90.7	37.0	32.0	102.2	251.2	277
benzene	17.6	15.4	15.8	30.5	9.2	52
tetrachloroethylene	15.3	10.2	8.8	27.5	19.6	128
styrene	3.5	2.7	2.8	6.8	2.4	69
m,p-dichlorobenzene	18.7	4.1	2.6	64.9	44.8	240
m,p-xylene	27.5	24.0	23.0	42.7	15.8	57

TABLE 4-2. ESTIMATED POPULATION GEOMETRIC MEANS OF 24-H PERSONAL AIR EXPOSURES ($\mu g/m^3$), BY SMOKING CATEGORY -- BASED ON WINTER 1984 LA COUNTY TEAM DATA

	GEOMETRIC MEANS			DIFFERENCES IN GMs		
COMPOUND	ACTIVE	PASSIVE	UNEXPOSED	ACTIVE -PASSIVE	ACTIVE -UNEXP.	PASSIVE -UNEXP.
1,1,1-trichloroethane	63.0	29.9	29.7	33.1	33.3*	0.2
benzene	20.1	14.8	13.4	5.2*	6.6*	1.4
tetrachloroethylene	13.6	9.0	9.1	4.6	4.5*	-0.1
styrene	4.3	2.6	2.2	1.8*	2.1*	0.4
m,p-dichlorobenzene	3.8	4.6	4.1	-0.8	-0.3	0.5
m,p-xylene	30.0	23.5	21.3	6.5*	8.7*	2.3

* = |difference|/std. error of difference > 2 ==> statistically significant at approximately 0.05 level

4.2 Example A2: Indoor Air 24-h Concentrations for Study 6

4.2.1 Analysis Objectives and Specifications

This example shows how to produce population and subpopulation estimates for 24-h indoor air concentrations for Study 6 (Woodland, 1990). The estimates are to be generated for two domains in addition to the overall population. The specifications are given below:

Geographic and Temporal Bounds: Study 6

Compounds: CMPD = 4, 5, 11, 14, 21, and 17
(1,1,1-trichloroethane, benzene, tetrachloroethylene [perc], styrene, m,p-xylene, and p-dichlorobenzene)

Media: Indoor air

Time Frame: 24-hr

Population Estimates to Be Generated: Mean, Geometric Mean, Population Variance, Percentiles (10, 25, 50, 75, 90) of Indoor Air 24-Hr Concentrations

Domains for Which Estimates Are to Be Generated:

(1) Overall Population

(2) Two Domains (Smoking occurs/does not occur in house) -- Variable HG40 of QUEST.DAT is to be used to define these domains.

Type of Analysis: Weighted Analysis Using SUDAAN

4.2.2 Analysis Results

The results for the 1990 Woodland Study indoor air concentration distributions are presented in Exhibit 4-2 in a format similar to the one used for Example A1. The definitions of terms given in Section 4.1.2 apply also to Exhibit 4-2. The programs used to perform this example are very similar to those described for the prior example. The main differences are that Study 6 indoor data and associated sampling weights (WMLA) were selected (rather than Study 1 personal air data with WPER as the sampling weight variable), compound 17 (p-dichlorobenzene) was used rather than compound 16 (m,p-dichlorobenzene), and the HG40 variable was used to directly define two house-related smoking

categories (rather than the three participant-related smoking categories). The program logs are shown in Appendix II-A, Exhibit A-2.

Selected results from Exhibit 4-2 were extracted and reproduced in Tables 4-3 and 4-4 (page 34), which parallel Tables 4-1 and 4-2 of the prior subsection. The estimated mean concentrations for the six selected compounds tend to be substantially larger than either the medians or the geometric means, which tend to be quite similar in magnitude. All of the coefficients of variation are large (well over 100%). Both of these properties, as well as the percentile estimates themselves, demonstrate that the indoor air distributions tend to be highly skewed. The medians and GMs will therefore be better than the means for characterizing the central tendency of the distributions; they will also tend to be more precisely estimated than the means (see standard errors in Exhibit 4-2). Little or no differences in "average" indoor levels were detected between those homes in the smoking category versus those in the nonsmoking category. The GM for styrene seemed to show a difference (t value of 0.40/0.16 = 2.5); however, the mean showed a (nonsignificant) difference in the opposite direction.

Exhibit 4-2. Analysis Type A: Study 6 (Woodland, May-June 1990) -- 24-h Indoor Air Concentrations

POPULATION ESTIMATES: MEANS, GEOMETRIC MEANS, AND STANDARD DEVIATIONS

---------- TABLE=1: OVERALL POPULATION ----------

C	HG40	N	WSUM	CM	CMSE	CGM	CGMSE	POPSD
1,1,1-Trichloroethane		115	13346	7.12	1.55	3.78	0.37	14.15
Benzene		124	14534	4.54	0.91	2.49	0.28	11.03
Tetrachloroethylene		124	14541	1.44	0.55	0.25	0.05	4.15
Styrene		123	14435	2.40	0.87	0.70	0.09	11.31
p-Dichlorobenzene		125	14624	16.15	3.46	1.08	0.27	51.34
m,p-Xylene		125	14624	7.41	1.22	4.44	0.39	13.43

---------- TABLE=2: BY SMOKING CATEGORY ----------

C	HG40	N	WSUM	CM	CMSE	CGM	CGMSE	POPSD
1,1,1-Trichloroethane	SMOKE	45	5099	8.24	3.01	3.78	0.67	17.85
1,1,1-Trichloroethane	NONSMOKE	64	7499	6.72	1.14	3.86	0.37	11.70
Benzene	SMOKE	47	5498	5.99	2.16	2.84	0.47	17.21
Benzene	NONSMOKE	70	8202	3.42	0.43	2.19	0.23	3.41
Tetrachloroethylene	SMOKE	46	5414	2.17	0.78	0.33	0.11	4.35
Tetrachloroethylene	NONSMOKE	71	8293	1.10	0.59	0.24	0.06	4.15
Styrene	SMOKE	47	5498	1.75	0.43	0.94	0.13	2.61
Styrene	NONSMOKE	69	8103	2.72	1.56	0.54	0.10	14.83
p-Dichlorobenzene	SMOKE	47	5498	17.08	6.58	1.23	0.41	49.53
p-Dichlorobenzene	NONSMOKE	71	8293	17.06	4.39	1.20	0.29	54.74
m,p-Xylene	SMOKE	47	5498	5.78	0.61	4.45	0.56	4.49
m,p-Xylene	NONSMOKE	71	8293	5.83	0.74	3.98	0.36	6.88

Exhibit 4-2. (continued)

POPULATION ESTIMATES: DIFFERENCES IN MEANS & GEOMETRIC MEANS

C	CONTRST	N	WSUM	CMDIF	CMDIFSE	CGMDIF	CGMDIFSE
1,1,1-Trichloroethane	SMOKE-NONSMOKE	109	12598	1.52	2.78	-0.08	0.71
Benzene	SMOKE-NONSMOKE	117	13700	2.57	2.08	0.65	0.44
Tetrachloroethylene	SMOKE-NONSMOKE	117	13707	1.07	0.73	0.09	0.12
Styrene	SMOKE-NONSMOKE	116	13601	-0.96	1.67	0.40	0.16
p-Dichlorobenzene	SMOKE-NONSMOKE	118	13791	0.02	8.38	0.03	0.46
m,p-Xylene	SMOKE-NONSMOKE	118	13791	-0.06	0.83	0.47	0.60

Exhibit 4-2. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=1,1,1-Trichloroethane TABLE=1: OVERALL POPULATION ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	115	13346	1.47	0.15	1.13	1.76
	25	115	13346	1.90	0.16	1.70	2.35
	50	115	13346	3.09	0.50	2.50	4.57
	75	115	13346	6.66	0.92	4.51	8.29
	90	115	13346	11.73	2.45	9.20	19.30

---------- C=1,1,1-Trichloroethane TABLE=2: BY SMOKING CATEGORY ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
SMOKE	10	45	5099	1.51	0.13	1.31	1.84
SMOKE	25	45	5099	1.87	0.13	1.68	2.23
SMOKE	50	45	5099	2.58	0.59	2.17	4.59
SMOKE	75	45	5099	5.45	1.45	3.85	9.83
SMOKE	90	45	5099	11.64		5.38	
NONSMOKE	10	64	7499	1.32	0.18	1.14	1.90
NONSMOKE	25	64	7499	1.92	0.28	1.60	2.76
NONSMOKE	50	64	7499	3.44	0.61	2.31	4.85
NONSMOKE	75	64	7499	7.23	1.11	4.87	9.44
NONSMOKE	90	64	7499	11.46		9.50	

Exhibit 4-2. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=Benzene TABLE=1: OVERALL POPULATION ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	124	14534	0.82	0.12	0.62	1.11
	25	124	14534	1.48	0.15	1.12	1.74
	50	124	14534	2.18	0.33	1.87	3.25
	75	124	14534	4.73	1.00	3.35	7.46
	90	124	14534	8.53	1.58	5.32	11.86

---------- C=Benzene TABLE=2: BY SMOKING CATEGORY ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
SMOKE	10	47	5498	0.85	0.24	0.61	1.61
SMOKE	25	47	5498	1.61	0.27	1.14	2.25
SMOKE	50	47	5498	2.43	0.57	1.79	4.16
SMOKE	75	47	5498	4.73	1.30	3.02	8.39
SMOKE	90	47	5498	9.66		4.63	
NONSMOKE	10	70	8202	0.77	0.13	0.56	1.09
NONSMOKE	25	70	8202	1.21	0.17	1.01	1.69
NONSMOKE	50	70	8202	2.00	0.21	1.72	2.58
NONSMOKE	75	70	8202	4.26	0.86	2.90	6.45
NONSMOKE	90	70	8202	7.87		5.38	

Exhibit 4-2. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=Tetrachloroethylene TABLE=1: OVERALL POPULATION ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	124	14541				0.08
	25	124	14541	0.13	0.02	0.09	0.16
	50	124	14541	0.25	0.06	0.17	0.42
	75	124	14541	0.73	0.18	0.47	1.20
	90	124	14541	1.80	1.62	0.86	7.58

---------- C=Tetrachloroethylene TABLE=2: BY SMOKING CATEGORY ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
SMOKE	10	46	5414	0.01			0.13
SMOKE	25	46	5414	0.13	0.02	0.09	0.17
SMOKE	50	46	5414	0.23	0.08	0.16	0.51
SMOKE	75	46	5414	1.01		0.30	
SMOKE	90	46	5414	8.47		0.87	
NONSMOKE	10	71	8293				0.11
NONSMOKE	25	71	8293	0.13	0.02	0.10	0.18
NONSMOKE	50	71	8293	0.28	0.09	0.18	0.55
NONSMOKE	75	71	8293	0.74	0.13	0.49	1.05
NONSMOKE	90	71	8293	1.32	0.20	0.76	1.58

Exhibit 4-2. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=Styrene TABLE=1: OVERALL POPULATION ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	123	14435	0.16	0.03	0.10	0.23
	25	123	14435	0.33	0.05	0.26	0.48
	50	123	14435	0.74	0.12	0.51	0.99
	75	123	14435	1.83	0.31	1.12	2.41
	90	123	14435	3.42	0.87	2.41	5.99

---------- C=Styrene TABLE=2: BY SMOKING CATEGORY ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
SMOKE	10	47	5498	0.27	0.07	0.17	0.46
SMOKE	25	47	5498	0.47	0.06	0.40	0.63
SMOKE	50	47	5498	0.79	0.23	0.58	1.52
SMOKE	75	47	5498	2.02	0.51	0.90	3.01
SMOKE	90	47	5498	3.30		2.24	
NONSMOKE	10	69	8103	0.13			0.21
NONSMOKE	25	69	8103	0.25	0.05	0.17	0.37
NONSMOKE	50	69	8103	0.54	0.15	0.33	0.95
NONSMOKE	75	69	8103	1.41	0.43	0.88	2.64
NONSMOKE	90	69	8103	3.03		1.95	

Exhibit 4-2. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=p-Dichlorobenzene TABLE=1: OVERALL POPULATION ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	125	14624	0.09	0.04	0.05	0.20
	25	125	14624	0.27	0.06	0.20	0.43
	50	125	14624	1.13	0.30	0.45	1.70
	75	125	14624	3.35	0.82	2.07	5.44
	90	125	14624	24.26	18.25	6.18	81.49

---------- C=p-Dichlorobenzene TABLE=2: BY SMOKING CATEGORY ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
SMOKE	10	47	5498	0.08	0.06	0.01	0.25
SMOKE	25	47	5498	0.28	0.06	0.21	0.46
SMOKE	50	47	5498	1.16	0.42	0.37	2.12
SMOKE	75	47	5498	4.26	2.27	2.03	11.41
SMOKE	90	47	5498	38.20		4.49	
NONSMOKE	10	71	8293	0.12	0.03	0.09	0.21
NONSMOKE	25	71	8293	0.29	0.08	0.19	0.51
NONSMOKE	50	71	8293	1.20	0.32	0.47	1.81
NONSMOKE	75	71	8293	2.92	1.26	1.77	6.96
NONSMOKE	90	71	8293	24.58	32.69	7.05	141.96

Exhibit 4-2. (continued)

POPULATION ESTIMATES: PERCENTILES

---------- C=m,p-Xylene TABLE=1: OVERALL POPULATION ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
	10	125	14624	1.59	0.17	1.39	2.09
	25	125	14624	2.47	0.23	2.04	2.98
	50	125	14624	4.08	0.37	3.35	4.90
	75	125	14624	7.52	0.95	5.70	9.60
	90	125	14624	11.85	2.43	10.40	20.42

---------- C=m,p-Xylene TABLE=2: BY SMOKING CATEGORY ----------

HG40	PCTILE	N	WSUM	QTILE	SEQTILE	LOWQTILE	UPQTILE
SMOKE	10	47	5498	1.77	0.44	0.83	2.65
SMOKE	25	47	5498	2.33	0.48	2.04	4.04
SMOKE	50	47	5498	4.24	0.91	2.68	6.44
SMOKE	75	47	5498	7.65	0.93	5.56	9.41
SMOKE	90	47	5498	10.71	1.04	8.20	12.50
NONSMOKE	10	71	8293	1.30	0.21	1.19	2.04
NONSMOKE	25	71	8293	2.24	0.32	1.85	3.18
NONSMOKE	50	71	8293	3.78	0.35	3.15	4.61
NONSMOKE	75	71	8293	6.94	1.16	4.82	9.60
NONSMOKE	90	71	8293	10.86	1.55	8.43	14.81

TABLE 4-3. SUMMARY OF ESTIMATED POPULATION DISTRIBUTIONS OF 24-H INDOOR AIR CONCENTRATIONS ($\mu g/m^3$) -- BASED ON JUNE 1990 WOODLAND STUDY DATA

COMPOUND	Mean	GM	Median	90th %ile	Std. Dev.	C.V. (%)
1,1,1-trichloroethane	7.12	3.78	3.09	11.73	14.15	199
benzene	4.54	2.49	2.18	8.53	11.03	243
tetrachloroethylene	1.44	0.25	0.25	1.80	4.15	288
styrene	2.40	0.70	0.74	3.42	11.31	471
p-dichlorobenzene	16.15	1.08	1.13	24.26	51.34	318
m,p-xylene	7.41	4.44	4.08	11.85	13.43	181

TABLE 4-4. ESTIMATED POPULATION GEOMETRIC MEANS OF 24-H INDOOR AIR CONCENTRATIONS ($\mu g/m^3$), BY HOME'S SMOKING STATUS -- BASED ON JUNE 1990 WOODLAND STUDY DATA

	GEOMETRIC MEANS		
COMPOUND	SMOKE	NONSMOKE	DIFFERENCE
1,1,1-trichloroethane	3.78	3.86	-0.08
benzene	2.84	2.19	0.65
tetrachloroethylene	0.33	0.24	0.09
styrene	0.94	0.54	0.40*
m,p-dichlorobenzene	1.23	1.20	0.03
m,p-xylene	4.45	3.98	0.47

* = |difference|/std. error of difference > 2 ==> statistically significant at approximately 0.05 level

5. ANALYSIS TYPE B: WEIGHTED ANALYSES FOR SEVERAL INDEPENDENT POPULATIONS

Analysis Type B is an extension of Type A to several independent studies. This type of analysis can be used to compare two or more independent (or independently sampled) populations. Depending on how the target population definitions differ among the studies, this type of analysis may be interpreted as providing geographic, seasonal, and/or annual comparisons. The types of statistics produced are like those for Analysis Type A, but in this case comparisons are made between studies and/or between comparable subpopulations of different studies.

The remainder of this section provides programs and results illustrating Analysis Type B; the example utilizes personal air 24-h concentration data from three independent populations -- namely, from Studies 1, 3, and 6.

5.1 Analysis Objectives and Specifications

This analysis has similar goals to that for Example A1, except that it applies to three independent studies. Also, the population standard deviation and percentile estimates, except for the median, are not to be produced. The specifications are as follows:

Geographic and Temporal Bounds: Studies 1, 3, and 6 to be compared

Compounds: CMPD = 4, 5, 11, 14, 21, and 16 & 17

(1,1,1-trichloroethane, benzene, tetrachloroethylene [perc], styrene, m,p-xylene, and m,p- or p-dichlorobenzene)

Media: Personal air

Time Frame: 24-hr

Population Estimates to Be Generated/Compared: Mean, Geometric Mean, and Median of Personal Air 24-Hr Concentrations

Domains for Which Estimates Are to Be Generated:

(1) 3 Overall Populations -- Compared to One Another

(2) Three Domains (Smokers, Others Exposed to ETS, Others Not Exposed to ETS) -- Variables PX41 and PX42 will be used to define these domains. Corresponding domains of each study to be compared to one another.

Type of Analysis: Weighted Analysis Using SUDAAN

5.2 Analysis Results

Appendix II-B shows the program listings used to prepare the data files, to conduct the analysis, and to generate the analysis results. The results are given in Exhibit 5-1, which has four sections of output entitled as follows:

1) POPULATION ESTIMATES: MEANS, GEOMETRIC MEANS, AND MEDIANS -- TABLE 1: BY STUDY

2) POPULATION ESTIMATES: MEANS, GEOMETRIC MEANS, AND MEDIANS -- TABLE 2: BY STUDY AND SMOKING CATEGORY

3) POPULATION ESTIMATES: DIFFERENCES IN MEANS AND GEOMETRIC MEANS

4) POPULATION ESTIMATES, BY SMOKING CATEGORY: DIFFERENCES IN MEANS AND GEOMETRIC MEANS

The quantities shown in the exhibit are defined below:

N = sample size

WSUM = sum of sampling weights (estimated population size)

CM = estimate of population mean

CMSE = estimated standard error for CM

CGM = estimate of population GM

CGMSE = estimated standard error for CGM

CMED = estimate of population median

CMEDSE = estimate of standard deviation of CMED

LOWCMED = lower 95% confidence bound for population median

UPCMED = upper 95% confidence bound for population median

CMDIF = estimated difference in two domain means

CMDIFSE = estimated standard error for CMDIF

CGMDIF = estimated difference in two domain GMs

CGMDIFSE = estimated standard error for CGMDIF

A portion of the Exhibit 5-1 results are reproduced in Table 5-1 (page 44). It gives the estimates of the overall population GMs of the 24-h personal exposures for the three studies and provides pairwise comparisons among the population GMs. Study 1 (LA County, winter) consistently exhibited higher GMs than either of the other studies (only

the m,p-dichlorobenzene difference for Studies 1 and 6 was not detected as significant). Comparison of the GMs for Study 3 (Pittsburg/Antioch, June 1984) and for Study 6 (Woodland, May-June 1990) showed an inconsistent pattern: two compounds had significantly higher levels in Woodland, two had significantly lower levels, and two yielded nonsignificant differences.

Table 5-2 (page 44) furnishes results (also extracted from Exhibit 5-1) that parallel those of Table 5-1; these results apply to the "unexposed" subpopulations of each of the three studies (i.e., excludes active smokers and those exposed to tobacco smoke) rather than to the overall population. The basic pattern of results for this subpopulation is like that for the overall population.

Exhibit 5-1. Analysis Type B: Study 1 vs. Study 3 vs. Study 6 -- 24-h Personal Air Concentrations

POPULATION ESTIMATES: MEANS, GEOMETRIC MEANS, AND MEDIANS

---------- TABLE=1: BY STUDY ----------

C	STUDY	SMOKE	N	WSUM	CM	CMSE	CGM	CGMSE	CMED	CMEDSE	LOWCMED	UPCMED
1,1,1-Trichloroethane	LA,WINTER84		111	359492	90.69	24.74	37.03	4.62	32.02	4.67	24.26	43.04
1,1,1-Trichloroethane	P/A,JUNE84		67	90696	15.34	5.84	6.83	0.95	5.54	0.57	4.17	6.48
1,1,1-Trichloroethane	WL,JUNE90		92	31221	21.58	5.56	6.23	0.92	4.14	1.19	2.99	7.79
Benzene	LA,WINTER84		111	359492	17.64	1.18	15.38	1.13	15.83	1.17	13.58	18.30
Benzene	P/A,JUNE84		67	90696	7.46	0.95	6.13	0.71	5.49	0.82	4.55	7.86
Benzene	WL,JUNE90		93	31470	5.01	0.80	3.36	0.32	3.03	0.23	2.76	3.69
Tetrachloroethylene	LA,WINTER84		110	359492	15.26	1.50	10.23	0.81	8.75	0.78	7.83	10.95
Tetrachloroethylene	P/A,JUNE84		67	90696	5.38	1.88	2.47	0.34	2.17	0.33	1.54	2.85
Tetrachloroethylene	WL,JUNE90		93	31470	1.62	0.40	0.50	0.08	0.37	0.07	0.29	0.55
Styrene	LA,WINTER84		110	359492	3.54	0.25	2.74	0.27	2.77	0.17	2.47	3.18
Styrene	P/A,JUNE84		66	90696	0.97	0.12	0.75	0.11	0.86	0.15	0.53	1.14
Styrene	WL,JUNE90		93	31470	2.43	0.73	1.26	0.09	1.19	0.18	0.92	1.64
m,p-Dichlorobenzene	LA,WINTER84		109	359492	18.71	4.04	4.06	0.74	2.62	0.65	1.87	4.51
m,p-Dichlorobenzene	P/A,JUNE84		67	90696	5.91	2.40	1.09	0.30	0.87	0.29	0.40	1.57
m,p-Dichlorobenzene	WL,JUNE90		93	31470	21.26	4.52	2.69	0.41	1.49	0.52	1.08	3.15
m,p-Xylene	LA,WINTER84		110	359492	27.50	1.68	23.98	1.58	23.03	1.57	21.09	27.42
m,p-Xylene	P/A,JUNE84		67	90696	11.10	2.00	8.64	1.30	8.14	1.75	6.67	13.73
m,p-Xylene	WL,JUNE90		93	31470	9.28	1.31	6.18	0.64	5.90	0.68	5.06	7.78

Exhibit 5-1. (continued)

POPULATION ESTIMATES: MEANS, GEOMETRIC MEANS, AND MEDIANS

TABLE=2: BY STUDY AND SMOKING CATEGORY

C	STUDY	SMOKE	N	WSUM	CM	CMSE	CGM	CGMSE	CMED	CMEDSE	LOWCMED	UPCMED
1,1,1-Trichloroethane	LA,WINTER84	ACTIVE	34	106552	212.29	74.79	63.03	15.99	42.94	8.77	32.62	67.91
1,1,1-Trichloroethane	LA,WINTER84	PASSIVE	23	72773	39.19	5.39	29.94	4.87	28.60	8.01	18.74	50.95
1,1,1-Trichloroethane	LA,WINTER84	UNEXPOSED	53	177809	39.90	5.19	29.74	3.74	27.25	4.20	20.56	37.45
1,1,1-Trichloroethane	P/A,JUNE84	ACTIVE	20	24082	35.03	20.53	10.26	3.87	6.17	2.53	3.48	13.65
1,1,1-Trichloroethane	P/A,JUNE84	PASSIVE	18	25916	8.60	2.17	5.63	1.44	5.16	1.55	2.44	8.66
1,1,1-Trichloroethane	P/A,JUNE84	UNEXPOSED	29	40698	7.97	0.95	6.07	0.54	4.86	0.41	4.20	5.86
1,1,1-Trichloroethane	WL,JUNE90	ACTIVE	21	7344	5.60	2.01	3.29	0.67	2.61	0.46	1.96	3.79
1,1,1-Trichloroethane	WL,JUNE90	PASSIVE	28	9353	26.59	10.61	7.19	1.87	4.67	2.50	2.95	13.00
1,1,1-Trichloroethane	WL,JUNE90	UNEXPOSED	43	14524	26.44	8.69	7.85	1.69	5.09	3.73	2.78	17.79
Benzene	LA,WINTER84	ACTIVE	34	106552	22.21	2.12	20.07	1.83	19.06	2.68	16.75	27.52
Benzene	LA,WINTER84	PASSIVE	23	72773	16.17	1.52	14.80	1.58	15.73	1.86	11.91	19.38
Benzene	LA,WINTER84	UNEXPOSED	53	177809	15.62	1.54	13.44	1.52	13.42	1.59	10.15	16.54
Benzene	P/A,JUNE84	ACTIVE	20	24082	9.11	1.27	7.62	1.18	6.69	1.53	4.44	10.60
Benzene	P/A,JUNE84	PASSIVE	18	25916	7.39	1.40	6.04	0.97	5.64	0.97	4.11	8.02
Benzene	P/A,JUNE84	UNEXPOSED	29	40698	6.53	1.22	5.43	0.94	4.56	1.21	3.78	8.65
Benzene	WL,JUNE90	ACTIVE	21	7344	5.65	1.84	3.89	0.63	3.05	0.73	2.65	5.57
Benzene	WL,JUNE90	PASSIVE	28	9353	4.76	0.73	3.64	0.46	3.07	0.48	2.38	4.33
Benzene	WL,JUNE90	UNEXPOSED	44	14772	4.85	1.38	2.96	0.44	2.97	0.18	2.59	3.30
Tetrachloroethylene	LA,WINTER84	ACTIVE	34	107471	20.52	3.36	13.57	1.72	12.08	1.57	8.41	14.74
Tetrachloroethylene	LA,WINTER84	PASSIVE	22	70533	15.52	5.69	9.00	1.51	8.72	1.36	5.32	10.81
Tetrachloroethylene	LA,WINTER84	UNEXPOSED	53	179014	12.11	1.84	9.10	1.12	8.33	1.13	6.57	11.13
Tetrachloroethylene	P/A,JUNE84	ACTIVE	20	24082	10.58	6.75	3.00	1.24	2.13	1.15	0.83	5.46
Tetrachloroethylene	P/A,JUNE84	PASSIVE	18	25916	4.67	1.35	2.72	0.86	2.47	0.51	0.89	2.95
Tetrachloroethylene	P/A,JUNE84	UNEXPOSED	29	40698	2.75	0.48	2.06	0.28	1.73	0.55	1.40	3.60
Tetrachloroethylene	WL,JUNE90	ACTIVE	21	7344	3.70	1.32	0.85	0.37	0.31	0.95	0.23	4.05
Tetrachloroethylene	WL,JUNE90	PASSIVE	28	9353	0.73	0.16	0.42	0.10	0.45	0.18	0.21	0.93
Tetrachloroethylene	WL,JUNE90	UNEXPOSED	44	14772	1.15	0.45	0.44	0.07	0.36	0.07	0.26	0.56

Exhibit 5-1. (continued)

POPULATION ESTIMATES: MEANS, GEOMETRIC MEANS, AND MEDIANS

---------- TABLE=2: BY STUDY AND SMOKING CATEGORY ----------
(continued)

C	STUDY	SMOKE	N	WSUM	CM	CMSE	CGM	CGMSE	CMED	CMEDSE	LOWCMED	UPCMED
Styrene	LA,WINTER84	ACTIVE	34	107471	4.83	0.42	4.30	0.31	4.71	0.52	3.52	5.61
Styrene	LA,WINTER84	PASSIVE	22	70533	3.18	0.51	2.54	0.36	2.79	0.37	1.95	3.42
Styrene	LA,WINTER84	UNEXPOSED	53	179014	2.93	0.35	2.16	0.38	2.21	0.17	1.90	2.58
Styrene	P/A,JUNE84	ACTIVE	20	24684	1.32	0.17	1.09	0.16	1.06	0.11	0.73	1.19
Styrene	P/A,JUNE84	PASSIVE	18	26260	0.90	0.20	0.69	0.17	0.76	0.25	0.28	1.27
Styrene	P/A,JUNE84	UNEXPOSED	28	39752	0.79	0.15	0.64	0.13	0.61	0.15	0.48	1.08
Styrene	WL,JUNE90	ACTIVE	21	7344	1.69	0.17	1.37	0.19	1.52	0.25	0.86	1.87
Styrene	WL,JUNE90	PASSIVE	28	9353	1.61	0.30	1.14	0.23	0.92	0.23	0.68	1.58
Styrene	WL,JUNE90	UNEXPOSED	44	14772	3.33	1.46	1.28	0.12	1.19	0.19	0.88	1.64
m,p-Dichlorobenzene	LA,WINTER84	ACTIVE	33	106600	9.98	3.24	3.80	0.71	2.55	0.47	1.91	3.78
m,p-Dichlorobenzene	LA,WINTER84	PASSIVE	22	70746	22.88	10.23	4.57	1.66	2.79	1.37	1.22	6.72
m,p-Dichlorobenzene	LA,WINTER84	UNEXPOSED	53	179671	22.51	8.23	4.13	1.12	3.21	1.15	1.58	6.22
m,p-Dichlorobenzene	P/A,JUNE84	ACTIVE	20	24082	11.50	7.32	1.77	1.21	1.66	0.98	0.21	4.15
m,p-Dichlorobenzene	P/A,JUNE84	PASSIVE	18	25916	3.48	1.55	0.84	0.28	0.71	0.49	0.19	2.15
m,p-Dichlorobenzene	P/A,JUNE84	UNEXPOSED	29	40698	4.15	2.42	0.98	0.25	0.89	0.24	0.44	1.40
m,p-Dichlorobenzene	WL,JUNE90	ACTIVE	21	7344	31.63	13.81	3.00	1.18	1.14	0.70	0.45	3.28
m,p-Dichlorobenzene	WL,JUNE90	PASSIVE	28	9353	10.57	8.44	1.63	0.71	1.41	0.71	0.59	3.45
m,p-Dichlorobenzene	WL,JUNE90	UNEXPOSED	44	14772	22.87	3.57	3.41	0.80	2.36	1.57	0.98	7.29
m,p-Xylene	LA,WINTER84	ACTIVE	34	107471	33.13	3.17	29.99	2.29	29.26	2.99	24.13	36.18
m,p-Xylene	LA,WINTER84	PASSIVE	22	70533	26.44	2.26	23.54	2.30	22.77	1.92	19.96	27.70
m,p-Xylene	LA,WINTER84	UNEXPOSED	53	179014	24.74	2.06	21.28	2.11	20.48	1.81	17.54	24.81
m,p-Xylene	P/A,JUNE84	ACTIVE	20	24082	11.50	1.50	9.51	1.35	10.62	1.76	6.42	13.52
m,p-Xylene	P/A,JUNE84	PASSIVE	18	25916	11.13	3.88	7.89	1.98	7.02	2.55	4.26	14.51
m,p-Xylene	P/A,JUNE84	UNEXPOSED	29	40698	10.85	2.40	8.65	1.77	8.73	1.81	6.82	14.09
m,p-Xylene	WL,JUNE90	ACTIVE	21	7344	7.03	1.15	5.76	0.76	5.74	1.01	4.06	8.11
m,p-Xylene	WL,JUNE90	PASSIVE	28	9353	9.19	2.22	5.69	1.40	5.25	1.10	3.45	7.88
m,p-Xylene	WL,JUNE90	UNEXPOSED	44	14772	10.45	2.12	6.73	0.83	6.18	0.99	4.86	8.83

Exhibit 5-1. (continued)

POPULATION ESTIMATES: DIFFERENCES IN MEANS & GEOMETRIC MEANS

C	CONTRST	N	WSUM	CMDIF	CMDIFSE	CGMDIF	CGMDIFSE
1,1,1-Trichloroethane	LA,WINTER84-P/A,JUNE84	178	450188	75.35	25.42	30.20	4.72
1,1,1-Trichloroethane	LA,WINTER84-WL,JUNE90	203	390714	69.10	25.36	30.81	4.71
1,1,1-Trichloroethane	P/A,JUNE84-WL,JUNE90	159	121917	-6.25	8.06	0.60	1.32
Benzene	LA,WINTER84-P/A,JUNE84	178	450188	10.17	1.52	9.26	1.33
Benzene	LA,WINTER84-WL,JUNE90	204	390962	12.63	1.43	12.03	1.17
Benzene	P/A,JUNE84-WL,JUNE90	160	122166	2.45	1.24	2.77	0.78
Tetrachloroethylene	LA,WINTER84-P/A,JUNE84	177	450188	9.89	2.40	7.76	0.88
Tetrachloroethylene	LA,WINTER84-WL,JUNE90	203	390962	13.64	1.55	9.72	0.81
Tetrachloroethylene	P/A,JUNE84-WL,JUNE90	160	122166	3.76	1.92	1.96	0.34
Styrene	LA,WINTER84-P/A,JUNE84	176	450188	2.58	0.28	1.99	0.30
Styrene	LA,WINTER84-WL,JUNE90	203	390962	1.11	0.77	1.49	0.29
Styrene	P/A,JUNE84-WL,JUNE90	159	122166	-1.47	0.74	-0.50	0.15
m,p-Dichlorobenzene	LA,WINTER84-P/A,JUNE84	176	450188	12.80	4.70	2.97	0.80
m,p-Dichlorobenzene	LA,WINTER84-WL,JUNE90	202	390962	-2.54	6.07	1.38	0.85
m,p-Dichlorobenzene	P/A,JUNE84-WL,JUNE90	160	122166	-15.35	5.12	-1.59	0.51
m,p-Xylene	LA,WINTER84-P/A,JUNE84	177	450188	16.40	2.61	15.33	2.04
m,p-Xylene	LA,WINTER84-WL,JUNE90	203	390962	18.23	2.13	17.80	1.70
m,p-Xylene	P/A,JUNE84-WL,JUNE90	160	122166	1.83	2.39	2.47	1.45

Exhibit 5-1. (continued)

POPULATION ESTIMATES, BY SMOKING CATEGORY: DIFFERENCES IN MEANS & GEOMETRIC MEANS

C	SMOKE	CONTRST	N	WSUM	CMDIF	CMDIFSE	CGMDIF	CGMDIFSE
1,1,1-Trichloroethane	ACTIVE	LA,WINTER84-P/A,JUNE84	54	130634	177.26	77.55	52.77	16.45
1,1,1-Trichloroethane	ACTIVE	LA,WINTER84-WL,JUNE90	55	113896	206.69	74.81	59.74	16.00
1,1,1-Trichloroethane	ACTIVE	P/A,JUNE84-WL,JUNE90	41	31427	29.43	20.63	6.97	3.92
1,1,1-Trichloroethane	PASSIVE	LA,WINTER84-P/A,JUNE84	41	98689	30.59	5.81	24.30	5.08
1,1,1-Trichloroethane	PASSIVE	LA,WINTER84-WL,JUNE90	51	82126	12.60	11.90	22.75	5.22
1,1,1-Trichloroethane	PASSIVE	P/A,JUNE84-WL,JUNE90	46	35269	-17.99	10.83	-1.55	2.36
1,1,1-Trichloroethane	UNEXPOSED	LA,WINTER84-P/A,JUNE84	82	218507	31.93	5.28	23.67	3.77
1,1,1-Trichloroethane	UNEXPOSED	LA,WINTER84-WL,JUNE90	96	192333	13.46	10.12	21.89	4.10
1,1,1-Trichloroethane	UNEXPOSED	P/A,JUNE84-WL,JUNE90	72	55222	-18.47	8.74	-1.77	1.77
Benzene	ACTIVE	LA,WINTER84-P/A,JUNE84	54	130634	13.11	2.47	12.45	2.18
Benzene	ACTIVE	LA,WINTER84-WL,JUNE90	55	113896	16.57	2.81	16.18	1.93
Benzene	ACTIVE	P/A,JUNE84-WL,JUNE90	41	31427	3.46	2.24	3.73	1.34
Benzene	PASSIVE	LA,WINTER84-P/A,JUNE84	41	98689	8.77	2.07	8.76	1.86
Benzene	PASSIVE	LA,WINTER84-WL,JUNE90	51	82126	11.40	1.69	11.16	1.65
Benzene	PASSIVE	P/A,JUNE84-WL,JUNE90	46	35269	2.63	1.58	2.39	1.08
Benzene	UNEXPOSED	LA,WINTER84-P/A,JUNE84	82	218507	9.09	1.97	8.01	1.78
Benzene	UNEXPOSED	LA,WINTER84-WL,JUNE90	97	192582	10.78	2.07	10.48	1.58
Benzene	UNEXPOSED	P/A,JUNE84-WL,JUNE90	73	55470	1.69	1.85	2.47	1.04
Tetrachloroethylene	ACTIVE	LA,WINTER84-P/A,JUNE84	54	131553	9.93	7.54	10.57	2.13
Tetrachloroethylene	ACTIVE	LA,WINTER84-WL,JUNE90	55	114815	16.82	3.61	12.72	1.76
Tetrachloroethylene	ACTIVE	P/A,JUNE84-WL,JUNE90	41	31427	6.89	6.87	2.15	1.30
Tetrachloroethylene	PASSIVE	LA,WINTER84-P/A,JUNE84	40	96448	10.85	5.85	6.28	1.74
Tetrachloroethylene	PASSIVE	LA,WINTER84-WL,JUNE90	50	79886	14.80	5.70	8.59	1.52
Tetrachloroethylene	PASSIVE	P/A,JUNE84-WL,JUNE90	46	35269	3.94	1.36	2.31	0.86
Tetrachloroethylene	UNEXPOSED	LA,WINTER84-P/A,JUNE84	82	219711	9.36	1.90	7.04	1.15
Tetrachloroethylene	UNEXPOSED	LA,WINTER84-WL,JUNE90	97	193786	10.96	1.89	8.67	1.12
Tetrachloroethylene	UNEXPOSED	P/A,JUNE84-WL,JUNE90	73	55470	1.60	0.66	1.63	0.29
Styrene	ACTIVE	LA,WINTER84-P/A,JUNE84	54	132155	3.51	0.46	3.21	0.35
Styrene	ACTIVE	LA,WINTER84-WL,JUNE90	55	114815	3.14	0.46	2.93	0.36
Styrene	ACTIVE	P/A,JUNE84-WL,JUNE90	41	32029	-0.37	0.24	-0.28	0.24
Styrene	PASSIVE	LA,WINTER84-P/A,JUNE84	40	96792	2.28	0.55	1.85	0.40
Styrene	PASSIVE	LA,WINTER84-WL,JUNE90	50	79886	1.57	0.59	1.40	0.43
Styrene	PASSIVE	P/A,JUNE84-WL,JUNE90	46	35613	-0.71	0.36	-0.45	0.29
Styrene	UNEXPOSED	LA,WINTER84-P/A,JUNE84	81	218766	2.14	0.38	1.53	0.40
Styrene	UNEXPOSED	LA,WINTER84-WL,JUNE90	97	193786	-0.40	1.50	0.88	0.40
Styrene	UNEXPOSED	P/A,JUNE84-WL,JUN	72	54525	-2.53	1.47	-0.64	0.17

Exhibit 5-1. (continued)

POPULATION ESTIMATES, BY SMOKING CATEGORY: DIFFERENCES IN MEANS & GEOMETRIC MEANS

C	SMOKE	CONTRST	N	WSUM	CMDIF	CMDIFSE	CGMDIF	CGMDIFSE
m,p-Dichlorobenzene	ACTIVE	LA,WINTER84-P/A,JUNE84	53	130682	-1.53	8.00	2.03	1.41
m,p-Dichlorobenzene	ACTIVE	LA,WINTER84-WL,JUNE90	54	113944	-21.65	14.18	0.81	1.38
m,p-Dichlorobenzene	ACTIVE	P/A,JUNE84-WL,JUNE90	41	31427	-20.13	15.63	-1.23	1.69
m,p-Dichlorobenzene	PASSIVE	LA,WINTER84-P/A,JUNE84	40	96662	19.40	10.34	3.74	1.68
m,p-Dichlorobenzene	PASSIVE	LA,WINTER84-WL,JUNE90	50	80099	12.31	13.26	2.94	1.80
m,p-Dichlorobenzene	PASSIVE	P/A,JUNE84-WL,JUNE90	46	35269	-7.09	8.58	-0.80	0.76
m,p-Dichlorobenzene	UNEXPOSED	LA,WINTER84-P/A,JUNE84	82	220369	18.36	8.58	3.15	1.15
m,p-Dichlorobenzene	UNEXPOSED	LA,WINTER84-WL,JUNE90	97	194444	-0.36	8.97	0.71	1.38
m,p-Dichlorobenzene	UNEXPOSED	P/A,JUNE84-WL,JUNE90	73	55470	-18.72	4.31	-2.44	0.84
m,p-Xylene	ACTIVE	LA,WINTER84-P/A,JUNE84	54	131553	21.63	3.50	20.48	2.66
m,p-Xylene	ACTIVE	LA,WINTER84-WL,JUNE90	55	114815	26.09	3.37	24.23	2.42
m,p-Xylene	ACTIVE	P/A,JUNE84-WL,JUNE90	41	31427	4.46	1.89	3.75	1.55
m,p-Xylene	PASSIVE	LA,WINTER84-P/A,JUNE84	40	96448	15.31	4.49	15.66	3.04
m,p-Xylene	PASSIVE	LA,WINTER84-WL,JUNE90	50	79886	17.25	3.17	17.86	2.69
m,p-Xylene	PASSIVE	P/A,JUNE84-WL,JUNE90	46	35269	1.94	4.47	2.20	2.43
m,p-Xylene	UNEXPOSED	LA,WINTER84-P/A,JUNE84	82	219711	13.89	3.16	12.63	2.75
m,p-Xylene	UNEXPOSED	LA,WINTER84-WL,JUNE90	97	193786	14.30	2.95	14.55	2.27
m,p-Xylene	UNEXPOSED	P/A,JUNE84-WL,JUNE90	73	55470	0.40	3.21	1.92	1.95

TABLE 5-1. ESTIMATED POPULATION GEOMETRIC MEANS OF 24-H PERSONAL AIR EXPOSURES ($\mu g/m^3$), BY STUDY[a]

COMPOUND	GEOMETRIC MEANS			DIFFERENCES IN GMs		
	STUDY 1	STUDY 3	STUDY 6	STUDY 1 - 3	STUDY 1 - 6	STUDY 3 - 6
1,1,1-trichloroethane	37.03	6.83	6.23	30.20*	30.81*	0.60
benzene	15.38	6.13	3.36	9.26*	12.03*	2.77*
tetrachloroethylene	10.23	2.47	0.50	7.76*	9.72*	1.96*
styrene	2.74	0.75	1.26	1.99*	1.49*	-0.50*
m,p-dichlorobenzene[b]	4.06	1.09	2.69	2.97*	1.38	-1.59*
m,p-xylene	23.98	8.64	6.18	15.33*	17.80*	2.47

* = |difference|/std. error of difference > 2 ==> statistically significant at approximately 0.05 level

a Study 1 = Los Angeles (LA) County - February-March 1984;
Study 3 = Pittsburg/Antioch - June 1984;
Study 6 = Woodland - May-June 1990.

b For Study 6, p-dichlorobenzene only.

TABLE 5-2. ESTIMATED GEOMETRIC MEANS OF 24-H PERSONAL AIR EXPOSURES ($\mu g/m^3$) FOR SUBPOPULATIONS NOT EXPOSED TO TOBACCO SMOKE, BY STUDY[a]

COMPOUND	GEOMETRIC MEANS			DIFFERENCES IN GMs		
	STUDY 1	STUDY 3	STUDY 6	STUDY 1 - 3	STUDY 1 - 6	STUDY 3 - 6
1,1,1-trichloroethane	29.74	6.07	7.85	23.67*	21.89*	-1.77
benzene	13.44	5.43	2.96	8.01*	10.48*	2.47*
tetrachloroethylene	9.10	2.06	0.44	7.04*	8.67*	1.63*
styrene	2.16	0.64	1.28	1.53*	0.88*	-0.64*
m,p-dichlorobenzene[b]	4.13	0.98	3.41	3.15*	0.71	-2.44*
m,p-xylene	21.28	8.65	6.73	12.63*	14.55*	1.92

* = |difference|/std. error of difference > 2 ==> statistically significant at approximately 0.05 level

a Study 1 = Los Angeles (LA) County - February-March 1984;
Study 3 = Pittsburg/Antioch - June 1984;
Study 6 = Woodland - May-June 1990.

b For Study 6, p-dichlorobenzene only.

6. ANALYSIS TYPE C: ANALYSES FOR PAIRED OBSERVATIONS

Analysis Type C applies to situations in which a target population is sampled on more than one occasion and (some of) the same sample members are used each time. In this case, the set of observations are not independent, but are paired. If the time points of sampling represent different seasons (years), then the results are used to assess seasonal (annual) trends. When the paired sample is not probability-based (i.e., either component), then SAS can be employed to generate an unweighted analysis of differences and/or ratios. Studies 1, 4, and 5 form such a paired-sample situation in which population estimates are not meaningful since Studies 4 and 5 were purposefully selected samples (subsamples of the Study 1 sample). Section 6.1 illustrates this case. If the paired sample (i.e., both components) is a probability-based sample, then SUDAAN can be applied to produce a weighted analysis of paired differences and/or ratios. The set-up for such an example (Studies 1 and 2 for 24-h personal air concentrations of six selected compounds) is given in section 6.2.

6.1 Example C1: Study 4 vs. 5 Personal, Indoor, and Outdoor 24-h Concentrations

6.1.1 Analysis Objectives and Specifications

Personal air, indoor air, and outdoor air levels are to be compared for the two 1987 LA County Studies -- Study 4 (winter) versus Study 5 (summer). The specifications are shown below:

Geographic and Temporal Bounds: Studies 4 and 5 to be compared

Compounds: CMPD = 4, 5, 11, 14, 21, and 16 & 17
(1,1,1-trichloroethane, benzene, tetrachloroethylene [perc], styrene, m,p-xylene, and m,p- or p-dichlorobenzene)

Media: Personal air, Indoor Air, Outdoor Air

Time Frame: 24-hr

Statistics to Be Generated (for personal air, indoor air, outdoor air and indoor/outdoor (I/O) ratios): Mean of Study 4 minus Study 5 Differences, Mean of Study 4 to Study 5 Ratios, Mean of Logarithms of Study 4 to Study 5 Ratios

Domains for Which Estimates Are to Be Generated: Overall Population
Type of Analysis: Unweighted Analysis Using SAS

Comparisons of the means of the indoor/outdoor ratios are included, in addition to comparisons of the means of the concentration measures themselves, in order to ascertain if a seasonal difference in indoor levels stems, at least partially, from a similar seasonal change in outdoor levels. Note that only overall estimates are to be produced (due to the small sample sizes). The results should not be interpreted as providing population-based estimates since the samples were purposefully drawn subsamples of the Study 1 sample.

6.1.2 Analysis Results

A listing of the program for performing the Type C example is included in Appendix II-C, Exhibit C-1. As a part of the file preparation, the following variables were created for conducting the analysis:

TYPE = type of concentration variable (1 = personal, 2 = indoor, 3 = outdoor, and 4 = I/O ratio)

Y1 = Study 4 concentration or indoor/outdoor ratio

Y2 = Study 5 concentration or indoor/outdoor ratio

LY1 = ln(Y1)

LY2 = ln(Y2)

DIFF = Y1-Y2

RATIO = Y1/Y2

LRATIO = ln(RATIO) = ln(Y1) - ln(Y2)

The pairwise comparisons performed for the 1987 LA County Studies (Study 4 [winter] versus Study 5 [summer]) are summarized in Exhibit 6-1. The MEANS procedure was applied to each of the variables listed above (other than TYPE) and produced the following statistics for each such variable: minimum value, maximum value, mean, standard deviation, standard error (of the mean), t statistic value (for testing that the mean is zero), and the significance probability associated with the t test. Although the software generates t statistics and associated probabilities for all of the above concentration-related quantities,

these should be ignored for all of the quantities except for LRATIO. (The statistics and associated probabilities are useful only (1) if the hypothesis tested by the t-test is meaningful and (2) if the underlying distribution of the data is approximately normal. Tests for Y1, Y2, LY1, LY2, and RATIO should be ignored for reason (1). Tests for DIFF may be meaningful, but should be ignored for reason (2) -- that is, differences of concentrations from long-tailed distributions are unlikely to be approximately normally distributed. On the other hand, the hypothesis that the mean of LRATIO is zero, equivalent to the hypothesis that the geometric mean of the ratios is one, is meaningful; moreover, for LRATIO, it appears reasonable to assume that the underlying data are approximately normally distributed.) For these reasons, we considered LRATIO to be the key variable for assessing seasonal differences, and hence consider the t test associated with the LRATIO mean as providing an approximate test of season differences.

The results for the indoor air concentrations and the I/O ratios are reproduced in Table 6-1 (page 54). The results imply that indoor levels for five of the compounds were higher in the winter. This was not true for the I/O ratios: styrene exhibited an I/O ratio difference in the opposite direction and the other compounds did not show statistically significant differences. Together, these results suggest that the apparent seasonal differences in indoor levels may be largely due to differences in outdoor levels. The results of this analysis should be viewed with caution due to the small sample sizes and the nature of the samples. It should also be noted that participants, rather than houses, were matched between the studies (see Appendix II-C); some differences might be expected if houses were matched.

Exhibit 6-1. Analysis Type C: Study 4 vs. Study 5 Comparison

CMPD=1,1,1-Trichloroethane TYPE=PERSONAL

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
29	Y1	1.810	281.270	38.451	55.046	10.222	3.762	0.0008
	Y2	3.170	321.050	22.992	58.578	10.878	2.114	0.0436
	LY1	0.593	5.639	3.043	1.107	0.206	14.801	0.0001
	LY2	1.154	5.772	2.342	0.954	0.177	13.223	0.0001
	DIFF	-194.19	271.890	15.459	68.122	12.650	1.222	0.2319
	RATIO	0.115	29.986	4.507	6.591	1.224	3.682	0.0010
	LRATIO	-2.167	3.401	0.701	1.324	0.246	2.851	0.0081

CMPD=1,1,1-Trichloroethane TYPE=INDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
25	Y1	2.560	62.080	18.756	14.836	2.967	6.321	0.0001
	Y2	1.900	154.920	16.096	29.524	5.905	2.726	0.0118
	LY1	0.940	4.128	2.630	0.822	0.164	15.994	0.0001
	LY2	0.642	5.043	2.268	0.864	0.173	13.130	0.0001
	DIFF	-129.49	54.060	2.660	31.448	6.290	0.423	0.6762
	RATIO	0.164	12.942	2.430	2.840	0.568	4.278	0.0003
	LRATIO	-1.807	2.560	0.362	1.066	0.213	1.698	0.1024

CMPD=1,1,1-Trichloroethane TYPE=OUTDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
25	Y1	1.080	26.950	9.546	7.191	1.438	6.637	0.0001
	Y2	1.160	19.590	5.677	3.993	0.799	7.109	0.0001
	LY1	0.077	3.294	1.953	0.848	0.170	11.519	0.0001
	LY2	0.148	2.975	1.534	0.653	0.131	11.756	0.0001
	DIFF	-13.540	19.840	3.869	8.178	1.636	2.365	0.0264
	RATIO	0.293	7.986	2.302	2.098	0.420	5.485	0.0001
	LRATIO	-1.226	2.078	0.419	0.973	0.195	2.153	0.0416

CMPD=1,1,1-Trichloroethane TYPE=I/O RATIO

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
17	Y1	0.944	12.139	2.874	3.117	0.756	3.802	0.0016
	Y2	1.026	14.712	3.026	3.273	0.794	3.812	0.0015
	LY1	-0.058	2.496	0.704	0.775	0.188	3.743	0.0018
	LY2	0.025	2.689	0.819	0.687	0.167	4.915	0.0002
	DIFF	-13.769	8.736	-0.152	4.684	1.136	-0.134	0.8953
	RATIO	0.064	8.466	1.485	1.989	0.482	3.077	0.0072
	LRATIO	-2.747	2.136	-0.115	1.031	0.250	-0.460	0.6518

Exhibit 6-1. (continued)

CMPD=Benzene TYPE=PERSONAL

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
21	Y1	2.360	67.350	19.115	16.755	3.656	5.228	0.0001
	Y2	1.550	20.360	7.403	3.939	0.859	8.613	0.0001
	LY1	0.859	4.210	2.643	0.820	0.179	14.771	0.0001
	LY2	0.438	3.014	1.876	0.532	0.116	16.167	0.0001
	DIFF	-5.800	56.330	11.712	15.705	3.427	3.417	0.0027
	RATIO	0.289	9.961	2.912	2.341	0.511	5.701	0.0001
	LRATIO	-1.241	2.299	0.767	0.848	0.185	4.146	0.0005

CMPD=Benzene TYPE=INDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
23	Y1	1.470	40.350	9.966	9.173	1.913	5.210	0.0001
	Y2	1.510	36.280	5.875	7.065	1.473	3.988	0.0006
	LY1	0.385	3.698	1.954	0.864	0.180	10.850	0.0001
	LY2	0.412	3.591	1.441	0.747	0.156	9.251	0.0001
	DIFF	-4.970	26.360	4.091	6.319	1.317	3.105	0.0052
	RATIO	0.228	13.260	2.313	2.554	0.532	4.343	0.0003
	LRATIO	-1.477	2.585	0.513	0.812	0.169	3.028	0.0062

CMPD=Benzene TYPE=OUTDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
25	Y1	0.860	19.550	6.279	4.350	0.870	7.217	0.0001
	Y2	0.870	10.630	4.080	2.729	0.546	7.475	0.0001
	LY1	-0.151	2.973	1.609	0.722	0.144	11.145	0.0001
	LY2	-0.139	2.364	1.182	0.709	0.142	8.333	0.0001
	DIFF	-5.760	8.970	2.200	3.924	0.785	2.803	0.0099
	RATIO	0.332	5.239	1.972	1.346	0.269	7.325	0.0001
	LRATIO	-1.103	1.656	0.426	0.769	0.154	2.771	0.0106

CMPD=Benzene TYPE=I/O RATIO

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
15	Y1	0.751	10.465	1.998	2.455	0.634	3.151	0.0071
	Y2	0.459	6.726	2.010	1.804	0.466	4.317	0.0007
	LY1	-0.286	2.348	0.364	0.707	0.183	1.991	0.0663
	LY2	-0.778	1.906	0.442	0.691	0.178	2.479	0.0265
	DIFF	-2.290	3.739	-0.013	1.302	0.336	-0.038	0.9702
	RATIO	0.528	3.498	1.074	0.751	0.194	5.536	0.0001
	LRATIO	-0.638	1.252	-0.078	0.522	0.135	-0.582	0.5697

Exhibit 6-1. (continued)

CMPD=Tetrachloroethylene TYPE=PERSONAL

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
20	Y1	0.400	78.780	13.409	17.744	3.968	3.380	0.0031
	Y2	0.370	251.780	15.046	55.743	12.464	1.207	0.2422
	LY1	-0.916	4.367	1.989	1.190	0.266	7.476	0.0001
	LY2	-0.994	5.529	0.963	1.308	0.292	3.293	0.0038
	DIFF	-173.00	31.610	-1.636	41.211	9.215	-0.178	0.8610
	RATIO	0.313	23.228	5.064	5.983	1.338	3.785	0.0013
	LRATIO	-1.162	3.145	1.026	1.148	0.257	3.997	0.0008

CMPD=Tetrachloroethylene TYPE=INDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
19	Y1	0.650	44.890	7.290	9.958	2.284	3.191	0.0051
	Y2	0.560	4.540	2.329	0.898	0.206	11.303	0.0001
	LY1	-0.431	3.804	1.540	0.889	0.204	7.555	0.0001
	LY2	-0.580	1.513	0.756	0.474	0.109	6.953	0.0001
	DIFF	-0.700	42.970	4.961	10.094	2.316	2.142	0.0461
	RATIO	0.750	23.380	3.677	5.535	1.270	2.896	0.0096
	LRATIO	-0.288	3.152	0.784	0.893	0.205	3.828	0.0012

CMPD=Tetrachloroethylene TYPE=OUTDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
25	Y1	0.450	11.480	3.780	3.079	0.616	6.137	0.0001
	Y2	0.410	2.640	1.616	0.688	0.138	11.744	0.0001
	LY1	-0.799	2.441	0.985	0.902	0.180	5.462	0.0001
	LY2	-0.892	0.971	0.353	0.570	0.114	3.100	0.0049
	DIFF	-1.310	9.320	2.164	2.907	0.581	3.722	0.0011
	RATIO	0.428	7.425	2.512	1.865	0.373	6.737	0.0001
	LRATIO	-0.849	2.005	0.632	0.812	0.162	3.891	0.0007

CMPD=Tetrachloroethylene TYPE=I/O RATIO

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
13	Y1	0.702	22.445	3.097	5.876	1.630	1.901	0.0816
	Y2	0.596	3.260	1.590	0.815	0.226	7.037	0.0001
	LY1	-0.354	3.111	0.488	0.918	0.255	1.919	0.0791
	LY2	-0.517	1.182	0.350	0.496	0.138	2.543	0.0258
	DIFF	-2.260	21.718	1.507	6.159	1.708	0.882	0.3950
	RATIO	0.307	30.862	3.297	8.297	2.301	1.433	0.1774
	LRATIO	-1.182	3.430	0.139	1.129	0.313	0.442	0.6661

Exhibit 6-1. (continued)

CMPD=Styrene TYPE*PERSONAL

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
26	Y1	0.220	162.720	9.553	31.322	6.143	1.555	0.1325
	Y2	0.190	3.670	1.418	0.920	0.181	7.854	0.0001
	LY1	-1.514	5.092	1.102	1.177	0.231	4.774	0.0001
	LY2	-1.661	1.300	0.118	0.738	0.145	0.818	0.4213
	DIFF	-1.100	161.720	8.135	31.393	6.157	1.321	0.1984
	RATIO	0.550	162.720	9.302	31.441	6.166	1.509	0.1440
	LRATIO	-0.598	5.092	0.984	1.161	0.228	4.321	0.0002

CMPD=Styrene TYPE=INDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
23	Y1	0.420	8.510	2.334	1.928	0.402	5.806	0.0001
	Y2	0.330	4.280	1.283	1.167	0.243	5.272	0.0001
	LY1	-0.868	2.141	0.571	0.760	0.158	3.607	0.0016
	LY2	-1.109	1.454	-0.084	0.804	0.168	-0.502	0.6203
	DIFF	-2.770	8.110	1.052	2.369	0.494	2.129	0.0447
	RATIO	0.132	21.275	3.282	4.282	0.893	3.676	0.0013
	LRATIO	-2.028	3.058	0.656	1.112	0.232	2.828	0.0098

CMPD=Styrene TYPE=OUTDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
24	Y1	0.040	4.370	1.304	1.153	0.235	5.540	0.0001
	Y2	0.050	1.860	0.405	0.398	0.081	4.979	0.0001
	LY1	-3.219	1.475	-0.159	1.060	0.216	-0.735	0.4699
	LY2	-2.996	0.621	-1.309	0.947	0.193	-6.772	0.0001
	DIFF	-1.240	3.440	0.900	1.103	0.225	3.996	0.0006
	RATIO	0.333	15.000	4.857	4.152	0.848	5.730	0.0001
	LRATIO	-1.099	2.708	1.150	1.030	0.210	5.469	0.0001

CMPD=Styrene TYPE=I/O RATIO

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
16	Y1	0.677	7.062	2.455	2.099	0.525	4.678	0.0003
	Y2	0.894	13.909	5.444	4.088	1.022	5.326	0.0001
	LY1	-0.389	1.955	0.617	0.740	0.185	3.331	0.0046
	LY2	-0.112	2.633	1.395	0.846	0.211	6.600	0.0001
	DIFF	-12.461	3.902	-2.989	4.008	1.002	-2.983	0.0093
	RATIO	0.104	2.468	0.673	0.673	0.168	3.998	0.0012
	LRATIO	-2.262	0.903	-0.779	0.888	0.222	-3.507	0.0032

Exhibit 6-1. (continued)

CMPD=p-Dichlorobenzene TYPE=PERSONAL

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
27	Y1	0.310	211.010	23.596	45.725	8.800	2.681	0.0126
	Y2	0.250	152.850	9.629	29.113	5.603	1.719	0.0976
	LY1	-1.171	5.352	1.598	1.822	0.351	4.557	0.0001
	LY2	-1.386	5.029	0.720	1.602	0.308	2.337	0.0274
	DIFF	-104.22	205.180	13.967	49.456	9.518	1.467	0.1543
	RATIO	0.128	36.194	5.646	8.305	1.598	3.533	0.0016
	LRATIO	-2.054	3.589	0.878	1.395	0.269	3.269	0.0030

CMPD=p-Dichlorobenzene TYPE=INDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
23	Y1	0.480	177.620	18.635	43.016	8.970	2.078	0.0496
	Y2	0.180	13.940	2.683	4.148	0.865	3.102	0.0052
	LY1	-0.734	5.180	1.195	1.704	0.355	3.363	0.0028
	LY2	-1.715	2.635	0.001	1.384	0.289	0.005	0.9960
	DIFF	-4.210	171.900	15.951	40.613	8.468	1.884	0.0729
	RATIO	0.264	31.052	5.385	6.529	1.361	3.956	0.0007
	LRATIO	-1.332	3.436	1.194	1.028	0.214	5.571	0.0001

CMPD=p-Dichlorobenzene TYPE=OUTDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
25	Y1	0.050	10.760	2.182	2.688	0.538	4.058	0.0005
	Y2	0.060	4.020	0.699	0.809	0.162	4.321	0.0002
	LY1	-2.996	2.376	0.071	1.302	0.260	0.274	0.7862
	LY2	-2.813	1.391	-0.794	0.941	0.188	-4.216	0.0003
	DIFF	-0.970	9.020	1.482	2.544	0.509	2.913	0.0076
	RATIO	0.179	22.111	4.518	5.376	1.075	4.203	0.0003
	LRATIO	-1.723	3.096	0.865	1.214	0.243	3.563	0.0016

CMPD=p-Dichlorobenzene TYPE=I/O RATIO

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
15	Y1	0.401	41.450	9.560	13.918	3.594	2.660	0.0187
	Y2	0.729	56.364	6.601	14.039	3.625	1.821	0.0901
	LY1	-0.915	3.724	1.187	1.574	0.406	2.921	0.0112
	LY2	-0.317	4.032	1.010	1.136	0.293	3.446	0.0039
	DIFF	-14.914	34.439	2.959	10.814	2.792	1.060	0.3071
	RATIO	0.183	14.329	2.486	3.737	0.965	2.576	0.0220
	LRATIO	-1.700	2.662	0.177	1.201	0.310	0.570	0.5779

Exhibit 6-1. (continued)

CMPD=m,p-Xylene TYPE=PERSONAL

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
29	Y1	3.760	215.840	43.204	42.166	7.830	5.518	0.0001
	Y2	3.390	217.450	20.732	38.274	7.107	2.917	0.0069
	LY1	1.324	5.375	3.429	0.843	0.157	21.912	0.0001
	LY2	1.221	5.382	2.615	0.716	0.133	19.671	0.0001
	DIFF	-124.45	197.710	22.472	47.963	8.907	2.523	0.0176
	RATIO	0.428	11.905	3.023	2.594	0.482	6.277	0.0001
	LRATIO	-0.849	2.477	0.814	0.785	0.146	5.585	0.0001

CMPD=m,p-Xylene TYPE=INDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
24	Y1	5.660	79.000	25.959	17.296	3.531	7.353	0.0001
	Y2	2.980	51.570	12.054	10.016	2.044	5.896	0.0001
	LY1	1.733	4.369	3.039	0.699	0.143	21.305	0.0001
	LY2	1.092	3.943	2.263	0.663	0.135	16.726	0.0001
	DIFF	-23.380	73.660	13.905	19.215	3.922	3.545	0.0017
	RATIO	0.350	14.794	3.028	2.926	0.597	5.068	0.0001
	LRATIO	-1.051	2.694	0.776	0.851	0.174	4.470	0.0002

CMPD=m,p-Xylene TYPE=OUTDOOR

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
25	Y1	1.650	49.910	16.440	11.711	2.342	7.019	0.0001
	Y2	1.270	26.160	8.509	6.176	1.235	6.889	0.0001
	LY1	0.501	3.910	2.543	0.777	0.155	16.369	0.0001
	LY2	0.239	3.264	1.870	0.799	0.160	11.709	0.0001
	DIFF	-11.030	28.500	7.931	10.583	2.117	3.747	0.0010
	RATIO	0.501	8.669	2.726	2.311	0.462	5.898	0.0001
	LRATIO	-0.691	2.160	0.673	0.837	0.167	4.019	0.0005

CMPD=m,p-Xylene TYPE=I/O RATIO

N Obs	Variable	Minimum	Maximum	Mean	Std Dev	Std Error	T	Prob>|T|
16	Y1	0.716	4.683	1.594	0.996	0.249	6.403	0.0001
	Y2	0.340	7.814	2.171	1.921	0.480	4.522	0.0004
	LY1	-0.334	1.544	0.336	0.492	0.123	2.732	0.0154
	LY2	-1.079	2.056	0.504	0.738	0.184	2.734	0.0154
	DIFF	-4.244	2.410	-0.577	1.582	0.395	-1.460	0.1649
	RATIO	0.215	8.091	1.292	1.871	0.468	2.762	0.0145
	LRATIO	-1.537	2.091	-0.168	0.809	0.202	-0.832	0.4185

TABLE 6-1. STUDY 4 VS. STUDY 5 PAIRED COMPARISONS FOR MEANS OF LOGARITHMS OF 24-H INDOOR CONCENTRATIONS AND FOR MEANS OF LOGARITHMS OF 24-H INDOOR/OUTDOOR RATIOS

	LRATIO for Indoor Conc.[a]				LRATIO for I/O Ratios[b]			
COMPOUND	n	Mean	Std. Error	t	n	Mean	Std. Error	t
1,1,1-trichloroethane	25	0.362	0.231	1.70	17	-0.115	0.250	-0.46
benzene	23	0.513	0.169	3.03*	15	-0.078	0.135	-0.58
tetrachloroethylene	19	0.784	0.205	3.83*	13	0.139	0.313	0.44
styrene	23	0.656	0.232	2.83*	16	-0.779	0.222	-3.51*
p-dichlorobenzene	23	1.194	0.214	5.57*	15	0.177	0.310	0.57
m,p-xylene	24	0.776	0.174	4.47*	16	-0.168	0.202	-0.83

* = statistically significant at the 0.01 level.

a LRATIO = ln(RATIO) = ln(Y1) - ln(Y2), where Y1 = Study 4 concentration and Y2 = Study 5 concentration.

b LRATIO = ln(RATIO) = ln(Y1) - ln(Y2), where Y1 = Study 4 indoor/outdoor ratio and Y2 = Study 5 indoor/outdoor ratio.

6.2 Example C2: Study 1 vs. 2 Personal 24-h Concentrations

Details for this example are not given. This is because the file preparation is like that described in Appendix II-C for Example C1 (except for the selection of studies to be included) and the analysis is like that for Analysis Type A (Section 4 and Appendix II-A), except that it is applied to concentration differences (or logarithms thereof) rather than to concentrations (or logarithms thereof). The set-up is outlined below:

Geographic and Temporal Bounds: Studies 1 and 2 to be compared

Compounds: CMPD = 4, 5, 11, 14, 21, and 16 (1,1,1-trichloroethane, benzene, tetrachloroethylene [perc], styrene, m,p-xylene, and m,p-dichlorobenzene)

Media: Personal air

Time Frame: 24-hr

Population Estimates to Be Generated: Mean and Median of Personal Air Differences; Mean, Geometric Mean, and Median of Personal Air Ratios.

Domains for Which Estimates Are to Be Generated: Overall Population

Type of Analysis: Weighted Analysis Using SUDAAN, with weights coming from Study 2

---

7. ANALYSIS TYPE D: ANALYSES INVOLVING DATA FROM MULTIPLE COMPOUNDS

Analyses of Type D differ from the above types in that they are multivariate in nature. Their general goal is to help elucidate how concentration levels of different compounds (in the same medium or in different media) relate to one another. Whereas the file preparation for the previously described analyses involves mainly a selection of data records and perhaps merging of the chemical data with the questionnaire data, the Type D analyses will generally require a restructuring of the chemical data to place all relevant data items from a given sample member into a single data record prior to analysis. Although the example described below deals with correlation analyses, the file restructuring is the same that will be needed for other types of multivariate analyses (e.g., principal components).

7.1 Analysis Objectives and Specifications

This analysis illustrates how to construct correlation matrices for the concentration data. Data from all studies were used; this ignores the fact that the samples for Studies 1, 2, 4, and 5 were not independent and the fact that Studies 4 and 5 were purposeful samples.

Geographic and Temporal Bounds: All Studies

Compounds: CMPD = 4, 5, 11, 14, 21, and 16 & 17
(1,1,1-trichloroethane, benzene, tetrachloroethylene [perc], styrene, m,p-xylene, and m,p- or p-dichlorobenzene)

Media: Personal air, Indoor, and Outdoor

Time Frame: 24-hr

Statistics to Be Generated: Unweighted Mean Concentrations, by Compound and Medium, and Correlations (and Associated Sample Sizes), by Medium (i.e., three 6x6 correlation matrices)

Domains for Which Analysis Is to Be Conducted: None

Type of Analysis: Unweighted Analysis Using SAS

7.2 Analysis Results

Appendix II-D describes the programs for this analysis example and includes a program listing (Exhibit D-1). Outputs generated are given in Exhibit 7-1, which shows six sets of summary statistics and six correlation matrices; the first three are for original concentration data values -- for personal, indoor, and outdoor air. The next three are for log-concentration-scale data. These correlation analyses reveal that the Pearson correlations of the logs of indoor concentrations are consistently, and sometimes dramatically, larger than the corresponding correlations on the original concentration scale. This is undoubtedly due to the highly skewed nature of the concentration distributions. It suggests that the log-scale correlations should generally be used, since the bivariate distributions on the log scale will tend to be more nearly normal. (Scatterplots are recommended to explore such assumptions.) An alternative to the log-scale correlations is the use of Spearman (rank) correlations, which would be scale-invariant. The log-scale correlations, which are reproduced in Table 7-1 (page 70), are consistently lower than the corresponding correlations for the outdoor air, as would be expected if indoor sources impact the indoor levels (i.e., indoor, and personal, concentrations include additional sources of variability due to variations in peoples' activities and their homes' characteristics). For the most part, the personal and indoor log-scale correlations are quite comparable. For the indoor air, the lowest log-scale correlations occur for p-dichlorobenzene with the other compounds (ranging from 0.18 to 0.34); the highest log-scale correlation, 0.83, was between benzene and m,p-xylene.

Exhibit 7-1. Analysis Type D: Intra-Media Correlation Analysis -- All Studies

24-HR PERSONAL AIR CONCENTRATION DATA

Simple Statistics

Variable	N	Mean	Std Dev	Sum
CP4	397	42.417254	139.625600	16840
CP5	388	10.910928	9.984461	4233.440000
CP11	382	8.854581	23.193622	3382.450000
CP14	392	2.856403	8.842805	1119.710000
CP16	394	15.525533	38.408501	6117.060000
CP21	397	20.912746	23.642299	8302.360000

Variable	Minimum	Maximum	Label
CP4	1.040000	1841.640000	1,1,1-Trichloroethane
CP5	0.350000	67.350000	Benzene
CP11	0	266.910000	Tetrachloroethylene
CP14	0.020000	162.720000	Styrene
CP16	0	308.180000	m,p- or p-Dichlorobenzene
CP21	0.580000	217.450000	m,p-Xylene

Exhibit 7-1. (continued)

24-HR PERSONAL AIR CONCENTRATION DATA

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / Number of Observations

	CP4	CP5	CP11	CP14	CP16	CP21
CP4 1,1,1-Trichloroethane	1.00000 0.0 397	0.25024 0.0001 387	0.36050 0.0001 381	0.09626 0.0572 391	-0.01300 0.7972 393	0.24068 0.0001 396
CP5 Benzene	0.25024 0.0001 387	1.00000 0.0 388	0.38419 0.0001 372	0.12429 0.0151 382	0.06034 0.2381 384	0.79035 0.0001 387
CP11 Tetrachloroethylene	0.36050 0.0001 381	0.38419 0.0001 372	1.00000 0.0 382	0.07892 0.1241 381	-0.01416 0.7835 379	0.45300 0.0001 382
CP14 Styrene	0.09626 0.0572 391	0.12429 0.0151 382	0.07892 0.1241 381	1.00000 0.0 392	-0.00854 0.8667 389	0.14629 0.0037 392
CP16 m,p- or p-Dichlorobenzene	-0.01300 0.7972 393	0.06034 0.2381 384	-0.01416 0.7835 379	-0.00854 0.8667 389	1.00000 0.0 394	0.05231 0.3004 394
CP21 m,p-Xylene	0.24068 0.0001 396	0.79035 0.0001 387	0.45300 0.0001 382	0.14629 0.0037 392	0.05231 0.3004 394	1.00000 0.0 397

Exhibit 7-1. (continued)

24-HR INDOOR AIR CONCENTRATION DATA

Simple Statistics

Variable	N	Mean	Std Dev	Sum
CI4	188	11.021223	17.393749	2071.990000
CI5	192	6.111875	11.388695	1173.480000
CI11	190	2.485105	5.053671	472.170000
CI14	192	2.309063	9.847583	443.340000
CI16	192	16.190469	49.346253	3108.570000
CI21	196	12.079286	16.718562	2367.540000

Variable	Minimum	Maximum	Label
CI4	0	154.920000	1,1,1-Trichloroethane
CI5	0	130.000000	Benzene
CI11	0	44.890000	Tetrachloroethylene
CI14	0	135.000000	Styrene
CI16	0	295.000000	m,p- or p-Dichlorobenzene
CI21	0	125.930000	m,p-Xylene

Exhibit 7-1. (continued)

24-HR INDOOR AIR CONCENTRATION DATA

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / Number of Observations

	CI4	CI5	CI11	CI14	CI16	CI21
CI4 1,1,1-Trichloroethane	1.00000 0.0 188	0.15343 0.0387 182	0.23824 0.0013 180	0.02433 0.7438 183	0.02777 0.7098 182	0.21142 0.0038 186
CI5 Benzene	0.15343 0.0387 182	1.00000 0.0 192	0.18717 0.0110 184	0.12131 0.0981 187	0.19328 0.0082 186	0.45628 0.0001 190
CI11 Tetrachloroethylene	0.23824 0.0013 180	0.18717 0.0110 184	1.00000 0.0 190	0.01535 0.8348 187	0.13033 0.0762 186	0.31404 0.0001 190
CI14 Styrene	0.02433 0.7438 183	0.12131 0.0981 187	0.01535 0.8348 187	1.00000 0.0 192	0.00816 0.9113 189	0.15047 0.0372 192
CI16 m,p- or p-Dichlorobenzene	0.02777 0.7098 182	0.19328 0.0082 186	0.13033 0.0762 186	0.00816 0.9113 189	1.00000 0.0 192	0.04258 0.5576 192
CI21 m,p-Xylene	0.21142 0.0038 186	0.45628 0.0001 190	0.31404 0.0001 190	0.15047 0.0372 192	0.04258 0.5576 192	1.00000 0.0 196

Exhibit 7-1. (continued)

24-HR OUTDOOR AIR CONCENTRATION DATA

Simple Statistics

Variable	N	Mean	Std Dev	Sum
C04	182	9.535110	15.129255	1735.390000
C05	181	5.316464	5.878398	962.280000
C011	181	3.117735	4.703572	564.310000
C014	181	1.163812	1.659776	210.650000
C016	178	1.119045	1.897971	199.190000
C021	181	10.650994	11.393220	1927.830000

Variable	Minimum	Maximum	Label
C04	0	119.090000	1,1,1-Trichloroethane
C05	0.460000	28.450000	Benzene
C011	0	27.640000	Tetrachloroethylene
C014	0	9.090000	Styrene
C016	0	14.860000	m,p- or p-Dichlorobenzene
C021	0.380000	50.640000	m,p-Xylene

Exhibit 7-1. (continued)

24-HR OUTDOOR AIR CONCENTRATION DATA

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / Number of Observations

	C04	C05	C011	C014	C016	C021
C04 1,1,1-Trichloroethane	1.00000 0.0 182	0.82745 0.0001 181	0.80639 0.0001 181	0.74478 0.0001 181	0.28699 0.0001 178	0.67655 0.0001 181
C05 Benzene	0.82745 0.0001 181	1.00000 0.0 181	0.84639 0.0001 180	0.86881 0.0001 180	0.42173 0.0001 177	0.85840 0.0001 181
C011 Tetrachloroethylene	0.80639 0.0001 181	0.84639 0.0001 180	1.00000 0.0 181	0.84732 0.0001 180	0.34115 0.0001 177	0.72514 0.0001 180
C014 Styrene	0.74478 0.0001 181	0.86881 0.0001 180	0.84732 0.0001 180	1.00000 0.0 181	0.37649 0.0001 177	0.76244 0.0001 180
C016 m,p- or p-Dichlorobenzene	0.28699 0.0001 178	0.42173 0.0001 177	0.34115 0.0001 177	0.37649 0.0001 177	1.00000 0.0 178	0.53099 0.0001 177
C021 m,p-Xylene	0.67655 0.0001 181	0.85840 0.0001 181	0.72514 0.0001 180	0.76244 0.0001 180	0.53099 0.0001 177	1.00000 0.0 181

Exhibit 7-1. (continued)

LN(24-HR PERSONAL AIR CONCENTRATION DATA)

Simple Statistics

Variable	N	Mean	Std Dev	Sum
CP4	397	2.636515	1.326692	1046.696499
CP5	388	2.018052	0.897209	783.004060
CP11	382	1.018521	1.603595	389.075150
CP14	392	0.425080	1.043198	166.631288
CP16	394	0.852159	1.908489	335.750532
CP21	397	2.613261	0.947113	1037.464484

Variable	Minimum	Maximum	Label
CP4	0.039221	7.518412	1,1,1-Trichloroethane
CP5	-1.049822	4.209903	Benzene
CP11	-5.298317	5.586912	Tetrachloroethylene
CP14	-3.912023	5.092031	Styrene
CP16	-5.298317	5.730684	m,p- or p-Dichlorobenzene
CP21	-0.544727	5.381969	m,p-Xylene

Exhibit 7-1. (continued)

LN(24-HR PERSONAL AIR CONCENTRATION DATA)

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / Number of Observations

	CP4	CP5	CP11	CP14	CP16	CP21
CP4 1,1,1-Trichloroethane	1.00000 0.0 397	0.52886 0.0001 387	0.61732 0.0001 381	0.52654 0.0001 391	0.22843 0.0001 393	0.59250 0.0001 396
CP5 Benzene	0.52886 0.0001 387	1.00000 0.0 388	0.62844 0.0001 372	0.58416 0.0001 382	0.20338 0.0001 384	0.83126 0.0001 387
CP11 Tetrachloroethylene	0.61732 0.0001 381	0.62844 0.0001 372	1.00000 0.0 382	0.45168 0.0001 381	0.18639 0.0003 379	0.62634 0.0001 382
CP14 Styrene	0.52654 0.0001 391	0.58416 0.0001 382	0.45168 0.0001 381	1.00000 0.0 392	0.26754 0.0001 389	0.63555 0.0001 392
CP16 m,p- or p-Dichlorobenzene	0.22843 0.0001 393	0.20338 0.0001 384	0.18639 0.0003 379	0.26754 0.0001 389	1.00000 0.0 394	0.22148 0.0001 394
CP21 m,p-Xylene	0.59250 0.0001 396	0.83126 0.0001 387	0.62634 0.0001 382	0.63555 0.0001 392	0.22148 0.0001 394	1.00000 0.0 397

Exhibit 7-1. (continued)

LN(24-HR INDOOR AIR CONCENTRATION DATA)

Simple Statistics

Variable	N	Mean	Std Dev	Sum
CI4	188	1.788114	1.074843	336.165429
CI5	192	1.203609	1.089315	231.092873
CI11	190	-0.552378	2.086181	-104.951822
CI14	192	-0.126873	1.328946	-24.359669
CI16	192	0.309640	2.202443	59.450914
CI21	196	1.921321	1.055764	376.578919

Variable	Minimum	Maximum	Label
CI4	-2.813411	5.042909	1,1,1-Trichloroethane
CI5	-3.352407	4.867534	Benzene
CI11	-5.298317	3.804215	Tetrachloroethylene
CI14	-5.298317	4.905275	Styrene
CI16	-5.298317	5.686975	m,p- or p-Dichlorobenzene
CI21	-1.742969	4.835726	m,p-Xylene

Exhibit 7-1. (continued)

LN(24-HR INDOOR AIR CONCENTRATION DATA)

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / Number of Observations

	CI4	CI5	CI11	CI14	CI16	CI21
CI4 1,1,1-Trichloroethane	1.00000 0.0 188	0.60393 0.0001 182	0.59637 0.0001 180	0.51940 0.0001 183	0.33064 0.0001 182	0.60297 0.0001 186
CI5 Benzene	0.60393 0.0001 182	1.00000 0.0 192	0.43496 0.0001 184	0.62889 0.0001 187	0.17714 0.0156 186	0.82897 0.0001 190
CI11 Tetrachloroethylene	0.59637 0.0001 180	0.43496 0.0001 184	1.00000 0.0 190	0.44786 0.0001 187	0.34151 0.0001 186	0.48751 0.0001 190
CI14 Styrene	0.51940 0.0001 183	0.62889 0.0001 187	0.44786 0.0001 187	1.00000 0.0 192	0.33055 0.0001 189	0.66812 0.0001 192
CI16 m,p- or p-Dichlorobenzene	0.33064 0.0001 182	0.17714 0.0156 186	0.34151 0.0001 186	0.33055 0.0001 189	1.00000 0.0 192	0.25564 0.0003 192
CI21 m,p-Xylene	0.60297 0.0001 186	0.82897 0.0001 190	0.48751 0.0001 190	0.66812 0.0001 192	0.25564 0.0003 192	1.00000 0.0 196

Exhibit 7-1. (continued)

LN(24-HR OUTDOOR AIR CONCENTRATION DATA)

Simple Statistics

Variable	N	Mean	Std Dev	Sum
C04	182	1.536637	1.167744	279.667897
C05	181	1.159408	1.010994	209.852933
C011	181	0.110566	1.653438	20.012523
C014	181	-0.788882	1.501593	-142.787688
C016	178	-0.861850	1.589882	-153.409319
C021	181	1.756703	1.185263	317.963185

Variable	Minimum	Maximum	Label
C04	-2.813411	4.779880	1,1,1-Trichloroethane
C05	-0.776529	3.348148	Benzene
C011	-5.298317	3.319264	Tetrachloroethylene
C014	-5.298317	2.207175	Styrene
C016	-5.298317	2.698673	m,p- or p-Dichlorobenzene
C021	-0.967584	3.924742	m,p-Xylene

Exhibit 7-1. (continued)

LN(24-HR OUTDOOR AIR CONCENTRATION DATA)

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / Number of Observations

	C04	C05	C011	C014	C016	C021
C04 1,1,1-Trichloroethane	1.00000 0.0 182	0.89336 0.0001 181	0.82834 0.0001 181	0.76156 0.0001 181	0.62612 0.0001 178	0.85327 0.0001 181
C05 Benzene	0.89336 0.0001 181	1.00000 0.0 181	0.82961 0.0001 180	0.80919 0.0001 180	0.72744 0.0001 177	0.93249 0.0001 181
C011 Tetrachloroethylene	0.82834 0.0001 181	0.82961 0.0001 180	1.00000 0.0 181	0.78805 0.0001 180	0.65138 0.0001 177	0.85804 0.0001 180
C014 Styrene	0.76156 0.0001 181	0.80919 0.0001 180	0.78805 0.0001 180	1.00000 0.0 181	0.65643 0.0001 177	0.80444 0.0001 180
C016 m,p- or p-Dichlorobenzene	0.62612 0.0001 178	0.72744 0.0001 177	0.65138 0.0001 177	0.65643 0.0001 177	1.00000 0.0 178	0.77476 0.0001 177
C021 m,p-Xylene	0.85327 0.0001 181	0.93249 0.0001 181	0.85804 0.0001 180	0.80444 0.0001 180	0.77476 0.0001 177	1.00000 0.0 181

TABLE 7-1. PEARSON CORRELATIONS BETWEEN LOG CONCENTRATIONS, BY MEDIA, BASED ON ALL DATA AVAILABLE IN THE SIX STUDIES

COMPOUNDS		CORRELATIONS x 100 PERSONAL	INDOOR	OUTDOOR
1,1,1-trichloroethane	benzene	53	60	89
	tetrachloroethylene	62	60	83
	styrene	53	52	76
	m,p- or p-dichlorobenzene	23	33	63
	m,p-xylene	59	60	85
benzene	1,1,1-trichloroethane	53	60	89
	tetrachloroethylene	63	43	83
	styrene	58	63	81
	m,p- or p-dichlorobenzene	20	18	73
	m,p-xylene	83	83	93
tetrachloroethylene	1,1,1-trichloroethane	62	60	89
	benzene	63	43	83
	styrene	45	45	79
	m,p- or p-dichlorobenzene	19	34	65
	m,p-xylene	63	49	86
styrene	1,1,1-trichloroethane	53	52	76
	benzene	58	63	81
	tetrachloroethylene	45	45	79
	m,p- or p-dichlorobenzene	27	33	66
	m,p-xylene	64	67	80
m,p- or p-dichlorobenzene	1,1,1-trichloroethane	23	33	63
	benzene	20	18	73
	tetrachloroethylene	19	34	65
	styrene	27	33	66
	m,p-xylene	22	26	77
m,p-xylene	1,1,1-trichloroethane	59	60	85
	benzene	83	83	93
	tetrachloroethylene	63	49	86
	styrene	64	67	80
	m,p- or p-dichlorobenzene	22	26	77

8. ANALYSIS TYPE E: ANALYSES BASED ON GENERAL LINEAR MODELS

These analyses include multiple linear regression analysis and analysis of variance and covariance -- analyses that seek to identify factors that influence concentration levels (or functions of concentration levels) and to estimate and test such effects. These analyses are more general than the domain comparisons conducted as a part of Analysis Types A and B in that multiple independent variables can be treated simultaneously. In general, the source data for the dependent variables will be the chemical data and the source data for (most of) the independent variables will be the questionnaire data. In most cases, SAS would be used to perform unweighted, model-based analyses, though under some circumstances, SUDAAN (PROC REGRESS) can be used to perform a weighted, probability-based analysis.

8.1 Analysis Objectives and Specifications

The purpose of the example illustrated in this section is to demonstrate how to conduct exploratory analyses directed at identifying and quantifying effects that may influence indoor air concentration levels of contaminants. Data from Studies 4, 5, and 6 are to be used since they provide indoor air data.[6] The pairing of observations and resultant lack of independence in the Study 4 and 5 data are ignored in the analysis (i.e., observations from all three studies are treated as though they are independent). The first goal of the analysis is to arrive at a meaningful dependent variable for characterizing indoor air levels; the second is to determine a "good" model that relates the variable to household characteristics, primarily those dealing with pollutant-generating sources/activities or activities that may reduce indoor levels (e.g., leaving windows open).

Specifications for the analysis are outlined below:

---

[6]Correlations and plots using Study 4 and 5 personal air nighttime data and indoor air nighttime data suggested that the former could not be expected to be a good surrogate for the latter. Hence, just those studies having actual indoor measurements were included in the analysis.

Geographic and Temporal Bounds: Studies 4, 5, 6

Compounds: CMPD = 4, 5, 7, 11, 12, 14, 20, 21, and 16 & 17 (1,1,1-trichloroethane, benzene, trichloroethylene, tetrachloroethylene [perc], chlorobenzene, styrene, o-xylene, m,p-xylene, and m,p- or p-dichlorobenzene)

Note: Compounds 7 and 12 may cause problems in that they have 42% and 8% quantifiable, respectively.

Media: Indoor air, as dependent variable; outdoor air as possible covariate in some models

Time Frame: 24-hr

Domains for Which Analysis Is to Be Conducted: Overall

Type of Analysis: Unweighted Regression Analyses Using SAS

Four candidate dependent variables were considered; these are defined in Table 8-1 (page 76) and are designated as Y1, Y2, Y3, and Y4. Sample sizes for Y2 and Y4 are considerably smaller than for Y1 or Y3, since they are functions of both indoor and outdoor concentration data and since the outdoor observations were taken at only a subset of the homes where indoor data were available. Of course, if outdoor levels are to be used as covariates in modeling Y1 or Y3, then the available observations will again be truncated to those having both indoor and outdoor data.

To arrive at a model, a list of candidate questionnaire variables was first constructed and frequencies across the three pertinent studies were calculated for each variable. Based on these frequencies, several variables were dropped from consideration and several others were combined to form more comprehensive variables (e.g., use of fireplaces and use of woodstoves were combined to form a single variable indicating if either was used). The final list of 24 candidate regressors (referred to here as X variables) is indicated in Table 8-1, which includes 22 questionnaire-derived indicator variables plus two variables indicating season and location (LOC1 and LOC2).

Also shown in this table are the forms of the candidate models, which can be explained as follows:

Case 1.	Model 1, applied to Y1:	Relate indoor air concentrations to X variables, ignoring outdoor air levels
Case 2.	Model 1, applied to Y3:	Relate logs of indoor air concentrations to X variables, ignoring outdoor air levels
Case 3.	Model 2, applied to Y1:	Relate indoor air concentrations to X variables, using outdoor air levels as a covariate
Case 4.	Model 3, applied to Y3:	Relate logs of indoor air concentrations to X variables, using logs of outdoor air levels as a covariate
Case 5.	Model 4, applied to Y2:	Relate indoor/outdoor concentration ratios to X variables
Case 6.	Model 4, applied to Y4:	Relate logs of indoor/outdoor concentration ratios to X variables

After deciding upon one of these ways for representing the relationship, the next goal of the analysis was to use a variable selection technique to reduce the number of regressors.

8.2 Analysis Results

The analysis involved three phases, as described below. Appendix II-E describes the analysis and includes the program listing (Exhibit E-1) for all three phases. It should be emphasized that the results are representative of an exploratory analysis and that further checks (residual plots, outlier checks, collinearity checks, etc.) on "final" models should be conducted.

Preliminary Modeling Results. The preliminary modeling involved using SAS to fit models for all six of the above-described cases, in each case using all of the available X variables in the model. SAS outputs from these preliminary regressions are quite extensive and are

not given here; the results relating to model performance are summarized in Table 8-2 (page 77). It gives the R^2 values and sample sizes for each case and compound. It is clear that Y3 is usually a better choice than Y1; also, despite the sacrifice in sample size, it is obvious that adjustment for outdoor levels is generally beneficial. Moreover, examination of the coefficient on the COUT or LCOUT variables indicated that it was generally significantly different from 1, suggesting that the covariance model would be preferred to the modeling of the I/O ratio. Hence Case 4 (Model 3 applied to Y3) was selected for the next phase of the analysis, the application of stepwise regression.

Stepwise Regression Results. Exhibit E-2 in Appendix II-E gives the complete SAS stepwise regression results for one compound -- 1,1,1-trichloroethane. It provides results for each step of the procedure -- i.e., the sequential introduction of variables into the model (and, where appropriate, the deletion of previously introduced variables). After arriving at a final model, the results of the procedure are summarized. Similar results for the other compounds were obtained but are simply summarized (Exhibit E-3 of Appendix II-E) by showing the last portion of the stepwise procedure's output.

Final Regression Results. The form of the final regression models is identical to those chosen via the stepwise algorithm. The models were refit, however, in order to utilize the maximum number of observations for each compound's model (since the stepwise procedure requires data values on all candidate regressors while these refit models require data values only for those regressors actually in the chosen model).

Exhibit 8-1 and Table 8-3 (page 87) provide the results of the final regressions for relating the 24-h indoor air concentrations (log scale) of the nine compounds to outdoor levels and selected indicator variables (see Table 8-1). Exhibit 8-1 shows, by compound, the simple summary statistics for Y3 and the independent variables; the analysis of variance and associated statistics (R^2, F, etc.); and the parameter estimates and their corresponding significance levels. Table 8-3 summarizes the results of Exhibit 8-1 and furnishes an overview of the patterns of the final models.

In terms of the amount of variability explained by the models (see Table 8-3), the model performance ranged from poor for two compounds (chlorobenzene: $R^2 = 0.13$; trichloroethylene: $R^2 = 0.23$) to good for the remaining seven (R^2 values in the range 0.42 to 0.70). The poor performance for chlorobenzene and trichloroethylene was anticipated, since the percentages of values quantifiable for the two compounds (unweighted estimates) were 8 and 42 percent, respectively.

The outdoor concentration covariate was always a strong predictor for the indoor concentration level. In addition to the covariate, there were a number of consistent and anticipated results. These include positive effects (i.e., increased levels predicted) for HG40R (tobacco smoking) on benzene levels; for HG25R (storage of gasoline-powered vehicles) on benzene and o-xylene (and styrene) levels; and for HG42R (drycleaned clothes) on tetrachloroethylene (and 1,1,1-trichloroethane) levels. Other effects that might be anticipated that were found were the positive effects for HX7R (air conditioner use) and HG6R (attached garage) on several of the compounds. A number of the other effects (or the direction of the effects) that were found statistically significant were not anticipated -- for example, the sometimes negative influences of HG23R and HG24R. Such examples point out that care must be exercised when interpreting these kinds of modeling results since a designated variable may be somewhat confounded with other variables (either other variables explicitly included in the model, or other potential and perhaps unknown variables left out of the model). For instance, one might speculate that homes with paints, etc. stored in the home (HG23R = 1) are more likely to be detached homes and/or larger homes having better ventilation than houses and apartments without such storage.

TABLE 8-1. VARIABLES AND MODELS USED FOR ANALYSIS TYPE E EXAMPLE

DEFINITIONS OF VARIABLES:

COUT = outdoor concentration
LCOUT = ln(COUT)
Y1 = indoor concentration
Y2 = Y1/COUT
Y3 = ln(Y1)
Y4 = ln(Y2)
LOC1 = 1 if Study = 4; = 0 otherwise
LOC2 = 1 if Study = 6; = 0 otherwise
HG5R = 1 if commercial cleaning (drapes/carpet); = 0 otherwise (based on HG for Studies 4 and 5, on HG4 for Study 6)
HG6R = 1 if attached garage; = 0 otherwise
HX7R = 1 if central (HX7) or window (HX8) air conditioning; = 0 otherwise
HX9R = 1 if circulation fans; = 0 otherwise
HX10R = 1 if exhaust fans; = 0 otherwise
HX11R = 1 gas furnace; = 0 otherwise
HX12R = 1 if fireplace (HX12) or woodstove (HX17); = 0 otherwise
HX15R = 1 if doors or windows open; = 0 otherwise
HX16R = 1 if gas or kerosene space heater; = 0 otherwise
HG19R = 1 if gas water heater; = 0 otherwise
HX22R = 1 if gas stove; = 0 otherwise
HG23R = 1 if paints/varnishes/thinners/removers; = 0 otherwise
HG24R = 1 if gasoline/petroleum products; = 0 otherwise
HG25R = 1 if auto, motorcycle, or gas lawnmower; = 0 otherwise
HG26R = 1 if -cides or garden/lawn chemicals; = 0 otherwise
HG27R = 1 if cleaning supplies; = 0 otherwise
HG31R = 1 if painting (HG31), furniture refinishing (HG32), or model building (HG33); = 0 otherwise
HG40R = 1 if tobacco prod. smoked in house; = 0 otherwise
HG41R = 1 if pesticides used in home; = 0 otherwise
HG42R = 1 if drycleaned clothes in home; = 0 otherwise
HX44R = 1 if shower/bath; = 0 otherwise
PX61R = 1 if aerosol personal care products; = 0 otherwise.

DEFINITIONS OF PRELIMINARY REGRESSION MODELS:

Model 1: Y1 or Y3 = $b_0 + \Sigma b_j X_j$,
where X_j = LOC1, LOC2, HG5R,...,PX61R.

Model 2: Y1 = $b_0 + a(\text{COUT}) + \Sigma b_j X_j$,
where X_j = LOC1, LOC2, HG5R,...,PX61R.

Model 3: Y3 = $b_0 + a(\text{LCOUT}) + \Sigma b_j X_j$,
where X_j = LOC1, LOC2, HG5R,...,PX61R.

Model 4: Y2 or Y4 = $b_0 + \Sigma b_j X_j$,
where X_j = LOC1, LOC2, HG5R,...,PX61R.

NOTE: Models 2, 3, and 4 require data on both indoor and outdoor concentrations, whereas Model 1 requires only indoor data.

TABLE 8-2. SUMMARY OF PRELIMINARY, FULL-MODEL REGRESSIONS FOR ANALYSIS TYPE E

COMPOUND	SAMPLE SIZE FOR MODEL 1	SAMPLE SIZE FOR MODEL 2,3,4	R^2 VALUES BY MODEL & DEPENDENT VAR. MODEL 1 Y1	MODEL 1 Y3	MODEL 2 Y1	MODEL 3 Y3	MODEL 4 Y2	MODEL 4 Y4
1,1,1-trichloroethane	167	104	0.22	0.40	0.33	0.59	0.30	0.42
benzene	169	99	0.22	0.36	0.43	0.56	0.36	0.31
trichloroethylene	174	101	0.22	0.28	0.29	0.43	0.32	0.49
tetrachloroethylene	167	99	0.34	0.48	0.57	0.74	0.41	0.35
chlorobenzene*	134	70	0.14	0.23	0.28	0.26	0.30	0.24
styrene	169	102	0.15	0.32	0.53	0.51	0.39	0.40
p-dichlorobenzene	169	99	0.13	0.16	0.64	0.60	0.30	0.47
o-xylene	172	102	0.41	0.52	0.44	0.69	0.44	0.38
m,p-xylene	173	102	0.44	0.56	0.46	0.73	0.48	0.40

* No data were available for Study 4; hence LOC1 was excluded from models.

Exhibit 8-1. Analysis Type E: Final Regression Results for Y3 (Log Indoor Concentration), By Compound

CMPD=1,1,1-Trichloroethane

Descriptive Statistics

Variables	Sum	Mean	Uncorrected SS
INTERCEP	106	1	106
LCOUT	127.7474792	1.2051648982	273.65232361
LOC2	44	0.4150943396	44
HG6R	54	0.5094339623	54
HX11R	15	0.141509434	15
HX12R	5	0.0471698113	5
HG26R	40	0.3773584906	40
HG42R	11	0.1037735849	11
Y3	211.41214138	1.9944541639	530.63466279

Variables	Variance	Std Deviation
INTERCEP	0	0
LCOUT	1.1399575795	1.0676879598
LOC2	0.2451033243	0.4950791092
HG6R	0.2522911051	0.5022858799
HX11R	0.1226415094	0.350202098
HX12R	0.0453728661	0.2130090752
HG26R	0.2371967655	0.4870285058
HG42R	0.0938903863	0.306415382
Y3	1.0379317821	1.0187893708

Dependent Variable: Y3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	7	57.06869	8.15267	15.390	0.0001
Error	98	51.91414	0.52974		
C Total	105	108.98284			

Root MSE	0.72783	R-square	0.5236
Dep Mean	1.99445	Adj R-sq	0.4896
C.V.	36.49268		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	1.514415	0.23062976	6.566	0.0001
LCOUT	1	0.511153	0.10269016	4.978	0.0001
LOC2	1	-0.473637	0.23477202	-2.017	0.0464
HG6R	1	-0.310839	0.14914341	-2.084	0.0397
HX11R	1	-0.751780	0.25498430	-2.948	0.0040
HX12R	1	0.685862	0.40055954	1.712	0.0900
HG26R	1	0.536626	0.17143812	3.130	0.0023
HG42R	1	0.872115	0.24190441	3.605	0.0005

Exhibit 8-1. (continued)

CMPD=Benzene

Descriptive Statistics

Variables	Sum	Mean	Uncorrected SS
INTERCEP	101	1	101
LCOUT	83.538992818	0.8271187408	152.60995162
HX7R	19	0.1881188119	19
HX9R	29	0.2871287129	29
HG25R	42	0.4158415842	42
HG40R	38	0.3762376238	38
Y3	135.72332813	1.343795328	282.53999801

Variables	Variance	Std Deviation
INTERCEP	0	0
LCOUT	0.8351328508	0.9138560339
HX7R	0.1542574257	0.3927561912
HX9R	0.2067326733	0.4546786484
HG25R	0.2453465347	0.495324676
HG40R	0.237029703	0.4868569636
Y3	1.0015562378	1.0007778164

Dependent Variable: Y3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	47.77291	9.55458	17.328	0.0001
Error	95	52.38271	0.55140		
C Total	100	100.15562			

Root MSE	0.74256	R-square	0.4770
Dep Mean	1.34380	Adj R-sq	0.4495
C.V.	55.25850		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	0.545724	0.16348335	3.338	0.0012
LCOUT	1	0.704307	0.09581600	7.351	0.0001
HX7R	1	0.581328	0.21874591	2.658	0.0092
HX9R	1	-0.368998	0.17881239	-2.064	0.0418
HG25R	1	0.275718	0.15573881	1.770	0.0799
HG40R	1	0.259043	0.16061304	1.613	0.1101

Exhibit 8-1. (continued)

CMPD=Trichloroethylene

Descriptive Statistics

Variables	Sum	Mean	Uncorrected SS
INTERCEP	103	1	103
LCOUT	-335.1533132	-3.253915662	1354.0457776
HX7R	19	0.1844660194	19
HG6R	50	0.4854368932	50
HG42R	10	0.0970873786	10
Y3	-115.347463	-1.119878282	400.31463063

Variables	Variance	Std Deviation
INTERCEP	2.514329E-16	1.5856635E-8
LCOUT	2.58318787	1.6072298747
HX7R	0.1519131925	0.3897604296
HG6R	0.2522368171	0.502231836
HG42R	0.0885208452	0.2975245288
Y3	2.658230509	1.6304080805

Dependent Variable: Y3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	61.52658	15.38165	7.191	0.0001
Error	98	209.61293	2.13891		
C Total	102	271.13951			

Root MSE	1.46250	R-square	0.2269
Dep Mean	-1.11988	Adj R-sq	0.1954
C.V.	-130.59459		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-0.277637	0.37038187	-0.750	0.4553
LCOUT	1	0.449862	0.10176231	4.421	0.0001
HX7R	1	1.238659	0.42110440	2.941	0.0041
HG6R	1	0.631504	0.28895107	2.186	0.0312
HG42R	1	0.891214	0.49455027	1.802	0.0746

Exhibit 8-1. (continued)

CMPD=Tetrachloroethylene

Descriptive Statistics

Variables	Sum	Mean	Uncorrected SS
INTERCEP	101	1	101
LCOUT	-33.79877246	-0.334641311	283.02586391
HG5R	29	0.2871287129	29
HG6R	51	0.504950495	51
HX9R	29	0.2871287129	29
HX22R	47	0.4653465347	47
HG23R	62	0.6138613861	62
HG24R	28	0.2772277228	28
HG42R	11	0.1089108911	11
Y3	8.9794449905	0.0889053959	287.86278095

Variables	Variance	Std Deviation
INTERCEP	0	0
LCOUT	2.7171539837	1.648379199
HG5R	0.2067326733	0.4546786484
HG6R	0.2524752475	0.5024691508
HX9R	0.2067326733	0.4546786484
HX22R	0.2512871287	0.5012854763
HG23R	0.2394059406	0.4892912636
HG24R	0.2023762376	0.4498624652
HG42R	0.098019802	0.3130811428
Y3	2.8706445984	1.6942976711

Dependent Variable: Y3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	8	201.04406	25.13051	26.877	0.0001
Error	92	86.02040	0.93500		
C Total	100	287.06446			

Root MSE	0.96696	R-square	0.7003
Dep Mean	0.08891	Adj R-sq	0.6743
C.V.	1087.62380		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	0.043414	0.22706535	0.191	0.8488
LCOUT	1	0.643195	0.06690084	9.614	0.0001
HG5R	1	0.388726	0.21863977	1.778	0.0787
HG6R	1	0.979861	0.23054650	4.250	0.0001
HX9R	1	-0.629195	0.23253856	-2.706	0.0081
HX22R	1	0.520070	0.20592171	2.526	0.0133
HG23R	1	-0.547068	0.22059521	-2.480	0.0150
HG24R	1	-0.617783	0.26368290	-2.343	0.0213
HG42R	1	0.918853	0.31254952	2.940	0.0042

Exhibit 8-1. (continued)

CMPD=Chlorobenzene

Descriptive Statistics

Variables	Sum	Mean	Uncorrected SS
INTERCEP	73	1	73
LCOUT	-237.6017608	-3.254818641	804.51910856
HX7R	19	0.2602739726	19
HG41R	24	0.3287671233	24
Y3	-215.0155101	-2.945417946	686.50106374

Variables	Variance	Std Deviation
INTERCEP	0	0
LCOUT	0.4328953919	0.6579478642
HX7R	0.1952054795	0.4418206417
HG41R	0.2237442922	0.4730161649
Y3	0.7387572443	0.8595098861

Dependent Variable: Y3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	6.75328	2.25109	3.345	0.0240
Error	69	46.43724	0.67300		
C Total	72	53.19052			

Root MSE	0.82037	R-square	0.1270
Dep Mean	-2.94542	Adj R-sq	0.0890
C.V.	-27.85234		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-1.979415	0.48982090	-4.041	0.0001
LCOUT	1	0.298461	0.14742495	2.024	0.0468
HX7R	1	0.510816	0.23110874	2.210	0.0304
HG41R	1	-0.387867	0.21550148	-1.800	0.0763

Exhibit 8-1. (continued)

CMPD=Styrene

Descriptive Statistics

Variables	Sum	Mean	Uncorrected SS
INTERCEP	104	1	104
LCOUT	-122.9234594	-1.181956341	342.57609234
HX7R	19	0.1826923077	19
HX9R	30	0.2884615385	30
HG23R	62	0.5961538462	62
HG25R	42	0.4038461538	42
Y3	1.2852041551	0.0123577323	90.906126137

Variables	Variance	Std Deviation
INTERCEP	0	0
LCOUT	1.9153973794	1.3839788219
HX7R	0.1507654966	0.3882853289
HX9R	0.2072442121	0.4552408287
HG23R	0.2430918596	0.4930434662
HG25R	0.2430918596	0.4930434662
Y3	0.8824295527	0.9393772153

Dependent Variable: Y3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	38.04157	7.60831	14.108	0.0001
Error	98	52.84867	0.53927		
C Total	103	90.89024			

Root MSE	0.73435	R-square	0.4185
Dep Mean	0.01236	Adj R-sq	0.3889
C.V.	5942.44565		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	0.637367	0.13435840	4.744	0.0001
LCOUT	1	0.344095	0.05454946	6.308	0.0001
HX7R	1	0.723140	0.21364161	3.385	0.0010
HX9R	1	-0.496140	0.16786884	-2.956	0.0039
HG23R	1	-0.608609	0.16430358	-3.704	0.0004
HG25R	1	0.385111	0.16202214	2.377	0.0194

Exhibit 8-1. (continued)

CMPD=p-Dichlorobenzene

Descriptive Statistics

Variables	Sum	Mean	Uncorrected SS
INTERCEP	102	1	102
LCOUT	-125.258081	-1.228020401	432.00583485
LOC2	45	0.4411764706	45
HG5R	28	0.2745098039	28
HX12R	5	0.0490196078	5
HX15R	99	0.9705882353	99
Y3	54.659705194	0.5358794627	388.67904376

Variables	Variance	Std Deviation
INTERCEP	0	0
LCOUT	2.7543203563	1.6596145204
LOC2	0.2489807804	0.4989797395
HG5R	0.201125995	0.4484707292
HX12R	0.0470782372	0.2169751996
HX15R	0.0288293535	0.1697920891
Y3	3.5582973298	1.8863449658

Dependent Variable: Y3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	183.36528	36.67306	20.001	0.0001
Error	96	176.02275	1.83357		
C Total	101	359.38803			

Root MSE	1.35409	R-square	0.5102
Dep Mean	0.53588	Adj R-sq	0.4847
C.V.	252.68628		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-0.890460	0.78723224	-1.131	0.2608
LCOUT	1	0.889257	0.10051141	8.847	0.0001
LOC2	1	1.417994	0.33985961	4.172	0.0001
HG5R	1	-1.182304	0.30972684	-3.817	0.0002
HX12R	1	1.114712	0.62307554	1.789	0.0768
HX15R	1	2.228227	0.80034080	2.784	0.0065

Exhibit 8-1. (continued)

CMPD=o-Xylene

Descriptive Statistics

Variables	Sum	Mean	Uncorrected SS
INTERCEP	104	1	104
LCOUT	67.650003872	0.6504808065	152.74705713
LOC1	31	0.2980769231	31
LOC2	44	0.4230769231	44
HX7R	19	0.1826923077	19
HG23R	62	0.5961538462	62
HG25R	42	0.4038461538	42
Y3	126.11328289	1.2126277201	237.64911379

Variables	Variance	Std Deviation
INTERCEP	0	0
LCOUT	1.0557478452	1.0274959101
LOC1	0.2112584018	0.4596285476
LOC2	0.2464525765	0.4964399023
HX7R	0.1507654966	0.3882853289
HG23R	0.2430918596	0.4930434662
HG25R	0.2430918596	0.4930434662
Y3	0.822530593	0.9069347237

Dependent Variable: Y3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	53.59065	8.93178	27.831	0.0001
Error	97	31.13000	0.32093		
C Total	103	84.72065			

Root MSE	0.56650	R-square	0.6326
Dep Mean	1.21263	Adj R-sq	0.6098
C.V.	46.71714		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	0.827648	0.15414232	5.369	0.0001
LCOUT	1	0.485956	0.08394269	5.789	0.0001
LOC1	1	0.380894	0.15989285	2.382	0.0192
LOC2	1	-0.333093	0.18128962	-1.837	0.0692
HX7R	1	0.693221	0.17578626	3.944	0.0002
HG23R	1	-0.263107	0.12879431	-2.043	0.0438
HG25R	1	0.313164	0.12902570	2.427	0.0171

Exhibit 8-1. (continued)

CMPD=m,p-Xylene

Descriptive Statistics

Variables	Sum	Mean	Uncorrected SS
INTERCEP	104	1	104
LCOUT	153.52829638	1.476233619	363.49210237
LOC1	31	0.2980769231	31
HX7R	19	0.1826923077	19
HG6R	52	0.5	52
HX11R	15	0.1442307692	15
HG26R	39	0.375	39
HX44R	101	0.9711538462	101
Y3	219.52387694	2.110806509	560.34581525

Variables	Variance	Std Deviation
INTERCEP	0	0
LCOUT	1.3286259202	1.1526603664
LOC1	0.2112584018	0.4596285476
HX7R	0.1507654966	0.3882853289
HG6R	0.2524271845	0.5024213217
HX11R	0.124626587	0.3530249099
HG26R	0.2366504854	0.4864673529
HX44R	0.0282860344	0.1681845247
Y3	0.9414891933	0.9703036604

Dependent Variable: Y3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	7	66.94969	9.56424	30.581	0.0001
Error	96	30.02370	0.31275		
C Total	103	96.97339			

Root MSE	0.55924	R-square	0.6904
Dep Mean	2.11081	Adj R-sq	0.6678
C.V.	26.49403		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	0.432228	0.34101121	1.267	0.2080
LCOUT	1	0.582229	0.06884141	8.458	0.0001
LOC1	1	0.513418	0.16726921	3.069	0.0028
HX7R	1	0.563202	0.16575027	3.398	0.0010
HG6R	1	0.322217	0.11548250	2.790	0.0064
HX11R	1	-0.316717	0.16999030	-1.863	0.0655
HG26R	1	-0.295062	0.13213854	-2.233	0.0279
HX44R	1	0.574946	0.33303155	1.726	0.0875

TABLE 8-3. ANALYSIS TYPE E: SUMMARY OF FINAL REGRESSION RESULTS FOR Y3 (LOG INDOOR CONCENTRATION), BY COMPOUND

	COMPOUND NUMBER[a]								
VARIABLE[b]	4	5	7	11	12	14	16	20	21
Intercept[c]	++++	++++	-	+	----	++++	-	++++	+
LCOUT	++++	++++	++++	++++	++	++++	++++	++++	++++
LOC1[c]					(na)			+++	++++
LOC2[c]	--						++++	--	
HG5R				++			----		
HX7R		++++	++++		+++	++++		++++	++++
HG6R	--		+++	++++					++++
HX9R		--		----		----			
HX11R	----								--
HX12R	++						++		
HX15R							++++		
HX22R				+++					
HG23R				---		----		----	
HG24R				---					
HG25R		+++				++++		+++	
HG26R	++++								---
HG40R		+++							
HG41R					--				
HG42R	++++		++	++++					
HX44R									++
n	106	101	103	101	73	104	102	104	104
R^2	0.52	0.48	0.23	0.70	0.13	0.42	0.51	0.63	0.69

NOTE: Sign (+ or -) indicates if the estimated coefficient for the variable is positive or negative. The number of signs indicates significance: ++++ or ---- signifies 0.01 level; +++ or --- signifies 0.05 level; ++ or -- signifies 0.10 level; + or - indicates that the term is in the model but is not significant at the 0.10 level.

a Compounds are as follows: 4 = 1,1,1-trichloroethane; 5 = benzene; 7 = trichloroethylene; 11 = tetrachloroethylene; 12 = chlorobenzene; 14 = styrene; 16 = p-dichlorobenzene; 20 = o-xylene; 21 = m,p-xylene.

b Variables are defined in Table 8-1. Variables HX10R, HX16R, HX19R, HG27R, HG31R, and PX61R are not listed, as they were not included in any of the final models.

c Intercept is an estimate of the Study 5 mean, adjusted for other effects in the model. Coefficients on LOC1 and LOC2 represent, respectively, the Study 4 vs. Study 5 effect and the Study 6 vs. Study 5 effect.

9. REFERENCES

1. Clayton, C.A. and Perritt, R.L. (1993). Data Base Development and Data Analysis for California Indoor Exposure Studies -- Volume I: Data Base. RTI Final Report prepared for California Air Resources Board (Contract A133-187).

2. Gilbert, R.O. (1987). Statistical Methods for Environmental Pollution Monitoring. New York: Van Nostrand Reinhold Company.

3. Hartwell, T.D., Pellizzari, E.D., Perritt, R.L., Whitmore, R.W., and Zelon, H.S. (1987). Comparison of volatile organic levels between sites and seasons for the Total Exposure Assessment Methodology (TEAM) study. Atmos. Environ. 21, 2413-2424.

4. Hartwell, T.D., Pellizzari, E.D., Perritt, R.L., Whitmore, R.W., Zelon, H.S., Sheldon, L.S. and Sparacino, C.M. (1987). Results from the Total Exposure Assessment Methodology (TEAM) study in selected communities in Northern and Southern California. Atmos. Environ. 21, 1995-2004.

5. Hartwell, T.D., Perritt, R.L., Whitmore, Pellizzari, E.D., and Michael, L.C. (1992). Results from the 1987 Total Exposure Assessment Methodology (TEAM) study in Southern California. Atmos. Environ. 26A, 1519-1527, 1992.

6. Pellizzari, E.D., Hartwell, T.D., Perritt, R.L., Sparacino, C.M., Sheldon, L.S., Zelon, H.S., Whitmore, R.W. and Wallace, L.A. (1986). Comparison of indoor and outdoor residential levels of volatile organic chemicals in five U.S. geographic areas. Environ. Int. 12, 619-623.

7. Pellizzari, E.D., Michael, L.C., Perritt, R.L., Smith, D.J., Hartwell, T.D., and Sebestik, J. (1990). Comparison of Indoor and Outdoor Toxic Air Pollutant Levels in Several Southern California Communities. RTI Final Report. EPA 68-02-4544.

8. Pellizzari, E.D., Perritt, R.L., Hartwell, T.D., Michael, L.C., Whitmore, R., Handy, R.W., Smith, D., and Zelon, H.S. (1985). Total Exposure Assessment Methodology (TEAM) Study: Selected Communities in Northern and Southern California. RTI Final Report. EPA 68-02-3679.

9. Perritt, R.L. and Clayton, C.A. (1990). Documentation of Data Obtained in the 1981-1987 Total Exposure Assessment Methodology Studies. RTI Final Report. EPA 68-02-4544.

10. Shah, B.V., Folsom, R.E., LaVange, L.M., Wheeless, S.C., Boyle, K.E., and Williams, R.L. (1993). Statistical Methods and Mathematical Algorithms Used in SUDAAN. Research Triangle Institute Report.

11. Sheldon, L., Clayton, A., Jones, B., Keever, J., Perritt. R., Smith, D., Whitaker, D. and Whitmore, R. (1991). Indoor Pollutant Concentrations and Exposure. RTI Final Report. ARB A833-156.

12. SUDAAN: Professional Software for SUrvey DAta ANalysis, Research Triangle Institute, 1989.

13. Wallace, Lance A. The Total Exposure Assessment Methodology (TEAM) Study: Summary and Analysis: Volume I (1987). EPA/600/6-87/002a, NTIS PB 88-100060.

14. Wallace, L.A., Pellizzari, E.D., Hartwell, T.D., Davis, V., Michael, L.C. and Whitmore, R.W. (1989), The influence of personal activities on exposure to volatile organic compounds. Environ. Res. 50, 37-55.

15. Wallace, L.A., Pellizzari, E., Hartwell, T., Perritt, K. and Ziegenfus, R. (1987). Exposures to benzene and other volatile organic compounds from active and passive smoking. Archives of Environ. Health 42, 272-279.

16. Wallace, L.A., Pellizzari, E.D., Hartwell, T.D., Whitmore, R., Perritt, R. and Sheldon, L. (1988). The California TEAM study: breath concentrations and personal exposures to 26 volatile compounds in air and drinking water of 188 residents of Los Angeles, Antioch, and Pittsburg, CA. Atmos. Environ. 22, 2141-63.

APPENDIX II-A

TYPE A ANALYSIS

APPENDIX II-A: TYPE A ANALYSIS

EXAMPLE A1

Analyses of Type A will normally require the sequential execution of three programs -- one for preparing the input files (e.g., using a SAS program), one for performing the statistical calculations (i.e., the SUDAAN analysis program), and one for post-processing and printing of the results. (If the amount of output is small, then one may wish to simply view the SUDAAN output directly and omit the post-processing step.) A listing of the program logs for Example A1 of Analysis Type A is provided on subsequent pages in Exhibit A-1.[7] Note that the logs of all three types of programs are included. The log includes both user-supplied statements, which can be identified by the presence of a line number, and notes or messages produced by SAS or SUDAAN as the program executed.

The SAS program for preparing the files consists of four main parts. The first involves input of the chemical data, and selection of the pertinent data records (those for Study 1 with personal air concentrations for the six specified compounds). Note that 661 of the 8797 records were included in the working file called CEXTRACT. The second part entails input of the questionnaire data related to tobacco smoking (not all of the variables read are used), the selection of the Study 1 data records, and the creation of a variable called SMOKE whose values indicate the status of each participant with respect to three domains of interest (1=active smoker, 2=passive smoke exposure, 3=not exposed). The resultant working file , QEXTRACT, is then merged with CEXTRACT in part three of the program. Several additional variables are created at this stage as well:

CPERSQ = squared concentration (This variable was used in calculating the population variance, as described later.)

---

[7] In this and other listings of programs, some minor modifications (repaging, deletion or addition of titles or header material, etc.) have been made to the outputs via wordprocessing in order to condense or clarify their presentation.

C = recoded compound numbers (This variable was created so that "compound" could be considered as a domain in the SUDAAN analysis, since SUDAAN requires that such subgroup variables be coded as 1, 2, 3, etc.)

CSTRAT = a recoded stratum identifier that combines compound IDs and stratum IDs into a single variable (Creation of this variable, along with C, allows a single call of a SUDDAN procedure to deal with multiple compounds: the SUDAAN NEST statement can use CSTRAT for stratum identification and ID5 for PSU identification, so long as all requests in the TABLE statement include C as a domain.)

Part four of the SAS program involves sorting of the file and generation of several output files for use by SUDAAN -- one overall file containing data on all compounds (ACHEMDAT.SSD) and six containing data for each compound separately (APCHEM1.SSD, APCHEM2.SSD,...,APCHEM6.SSD). The former is used for estimation of the means and geometric means while the latter files are needed for the estimation of the percentiles.[8]

The SUDAAN program is invoked by using a command like

SUDPROC AANLYSIS.SUD AANLYSIS.SLG -R,

where AANLYSIS.SUD is the file containing the SUDAAN program (i.e., those statements in the SUDAAN program listing having line numbers) and AANLYSIS.SLG is the file containing the SUDAAN log (i.e., the SUDAAN portion of Exhibit A-1). Note that the design type specification for this analysis (as well as for the other SUDAAN analyses applicable to the consolidated database) is given as DESIGN=WR, which indicates a design in which PSUs were selected with replacement within strata. The SUDAAN program consists of simply a series of calls of the DESCRIPT procedure. The first three calls of the procedure each use ACHEMDAT.SSD as the input file and produce results for the overall Study 1 population and for each of the three smoking domains, which are denoted in the output as ACTIVE, PASSIVE, and UNEXPOSED. (Overall population estimates are generated by specifying the NOMARG option and specifying C (or _ONE_) in the TABLES statement. An alternative approach is to omit the NOMARG option and use the marginal portion of the C*SMOKE (or SMOKE)

---

[8]The separate files could also be used for the means and geometric means but more post-processing of the results would be needed. In that case, creation of the C and CSTRAT variables would not have been necessary.

table to provide the overall population estimate. These two approaches will produce identical results only if the SMOKE variable has no missing values.) The first call produces estimates of population means and geometric means, the second call of PROC DESCRIPT furnishes pairwise comparisons of the means and geometric means, and the third call generates the estimated means of the squared concentration variable (CPERSQ). The results from executing these three procedures are stored in AM1CHEM.DBS, AM2CHEM.DBS, and ACHEMSQ.DBS, respectively.[9]

The next set of six calls of PROC DESCRIPT are used to produce the percentile estimates; the input file and output file for the first chemical are APCHEM1.SSD and APCHEM1.DBS, respectively, with similar names for the other five chemicals. These DESCRIPT procedures produce the estimated percentiles using the "UNGROUPED" option, which is appropriate when sample sizes are small. (All of the studies are considered to have small sample sizes, relative to sample sizes that might be used in a general statewide or nationwide survey, for example). With this option, the calculation of standard errors is made indirectly -- by first calculating approximate confidence intervals. The use of separate files for each compound is required so that SUDAAN will correctly calculate the degrees of freedom associated with the confidence intervals. It should be noted that sample sizes are in some cases insufficient for calculating the precision of the percentile estimates (usually for the more extreme percentiles) and the SUDAAN log contains a corresponding "DATA WARNING." Also note that the "UNGROUPED" option requires the user to specify an approximate maximum sample size, which here appears as the option "EST_NO=120" in each call of PROC DESCRIPT (see Exhibit A-1).

Post-processing of the SUDAAN output is accomplished using a SAS program (named AANLPOST.SAS in Exhibit A-1). It begins with a record-by-record merging of the AM1CHEM.DBS and the ACHEMSQ.DBS files (each file contains 24 records -- 6 chemicals times four categories [overall and three domains]). The estimate of the population standard deviation

---

[9]SUDAAN produces several ASCII output files that can be used by subsequent programs. Those with the DBS extension contain the estimation results. See the SUDAAN manual for details and formats.

(SD) for each case is then approximated by calculating the square root of

$$CSQM - (CM)^2$$

where CSQM is the estimated mean of CPERSQ (from ACHEMSQ.DBS), and CM is the estimated mean of CPER (from AM1CHEM.DBS). These SD results are then printed, along with the estimated means and GMs, using PROC PRINT. The next part of the program inputs the contrast information (file AM2CHEM.DBS) and prints those results. The final part of the program inputs the percentile results for each chemical (files APCHEM1.DBS, APCHEM2.DBS, etc.) and prints those results.

Outputs produced by the programs in Exhibit A-1 can be found in Exhibit 4-1 of the main text.

EXAMPLE A2

The programs used to perform this example are very similar to those described for the prior example. The main differences are that Study 6 indoor data and associated sampling weights (WMLA) are selected (rather than Study 1 personal air data with WPER as the sampling weight variable), compound 17 (p-dichlorobenzene) is used rather than compound 16 (m,p-dichlorobenzene), and the HG40 variable is used to directly define two house-related smoking categories (rather than the three participant-related smoking categories). File names have also been modified and the maximum sample size specifications for the percentile estimations have been changed to 130. The program logs are shown in Exhibit A-2. Outputs produced by the programs in Exhibit A-2 can be found in Exhibit 4-2 of the main text.

Exhibit A-1. Analysis Type A: Program Logs for Example A1

SAS PROGRAM FOR PREPARING FILES FOR SUDAAN ANALYSIS

```
250    * PGM STORED IN CARB\AANLYSIS.SAS;
251    FILENAME INHR24 'C:\CARB\HOUR24.DAT';
252    FILENAME INQUEST 'C:\CARB\QUEST.DAT';
253    LIBNAME OUT 'C:\CARB';
254    OPTIONS LS=95 PS=56 NOCENTER NODATE NONUMBER;
255
256    ******************************************************************;
257    * Extract pertinent 24-hour chemical data;
258    DATA CEXTRACT;
259    INFILE INHR24;
260    INPUT
261
262    /* FOLLOWING STATEMENTS WERE EXTRACTED FROM HOUR24.CBK */
263
264    @1    ID1    1.    /*STUDY NUMBER*/
265    @2    ID2    7.    /*HOUSEHOLD/PARTICIPANT ID*/
266    @9    CMPD   2.    /*COMPOUND ID*/
267    @11   CPER   9.2   /*CONCENTRATION FOR PERSONAL AIR*/
268    @47   MPER   1.    /*MEASURABILITY INDICATOR FOR PERSONAL AIR*/
269    @51   WPER  10.4   /*SAMPLING WEIGHT FOR PERSONAL AIR*/
270     ;
271
272    IF ID1=1;        /* SELECT RECORDS FOR STUDY 1 */
273    IF MPER >=1;     /* SELECT RECORDS WITH PERSONAL AIR DATA */
274    IF CMPD=4 OR CMPD=5 OR CMPD=11 OR CMPD=14 OR CMPD=21 OR CMPD=16;
275       /* SELECT COMPOUNDS (FOR STUDY 1, CMPD 17 IS NOT MEASURED) */
276    RUN;
NOTE: The infile INHR24 is file C:\CARB\HOUR24.DAT.
NOTE: 8797 records were read from the infile C:\CARB\HOUR24.DAT.
      The minimum record length was 80.
      The maximum record length was 80.
```

Exhibit A-1. (continued)

```
NOTE: The data set WORK.CEXTRACT has 661 observations and 6 variables.
NOTE: The DATA statement used 31.00 seconds.
  277
  278    ******************************************************************;
  279    * Extract pertinent questionnaire data;
  280    DATA QEXTRACT;
  281    INFILE INQUEST LRECL=291;
  282    INPUT
  283
  284    /* FOLLOWING STATEMENTS WERE EXTRACTED FROM QUEST.CBK */
  285
  286    @1   ID1   1.     /*Study ID */
  287    @2   ID2   7.     /*Household/Participant ID */
  288    @16  ID4   7.     /*Stratum Code */
  289    @23  ID5   3.     /*Primary Sampling Unit Code */
  290    @26  ID6  10.4    /*Personal Sampling Weight */
  291    @127 PG36  2.     /*Smoking Status (Current Smoker, Former Smoker, Never Smoked) */
  292    @129 PG37  1.     /*Avg. No. Cigarettes/Day (<10, 10-30, >30) */
  293    @130 HG37  1.     /*HH Members Smoke Cigarettes (Participant Only, Others Only,
  294                                                     Both, Neither) */
  295    @131 PG38  1.     /*Avg. Hrs/Day Enclosed w/Smoker-Home (0, 1-5, 6-10, >10) */
  296    @132 PG39  1.     /*Avg. Hrs/Day Enclosed w/Smoker-Work (0, 1-5, 6-10, >10) */
  297    @133 HX38  3.     /*No. Cigarettes Smoked in House */
  298    @136 HX39  3.     /*No. Cigarettes, Cigars, or Pipefuls */
  299    @139 HG40  1.     /*Tobacco Prod. Smoked in House */
  300    @140 PX40  3.     /*No. Cigarettes Smoked */
  301    @143 PD40  3.     /*No. Cigarettes Smoked - Daytime */
  302    @146 PN40  3.     /*No. Cigarettes Smoked - Nighttime */
  303    @149 PX41  1.     /*Smoked Tobacco Products */
  304    @150 PD41  1.     /*Smoked Tobacco Products - Day */
  305    @151 PN41  1.     /*Smoked Tobacco Products - Night */
  306    @152 PX42  1.     /*Enclosed Area w/Smoker(s) */
  307    @153 PD42  1.     /*Enclosed Area w/Smokers - Day */
  308    @154 PN42  1.     /*Enclosed Area w/Smokers - Night */
  309    @155 PX43  6.2    /*% Time Indoors at Home with Smokers */
  310    @161 PX44  6.2    /*% Time Indoors Not at Home w/ Smokers */
  311    @167 PX45  6.2    /*% Time Enclosed in Transit w/ Smokers */
  312    ;
  313
```

Exhibit A-1. (continued)

```
  314    IF ID1=1;        /* SELECT RECORDS FOR STUDY 1 */
  315
  316    * CREATE DOMAIN DEFINING VARIABLE NAMED SMOKE
  317              (1=SMOKER, 2=EXPOSED TO PASSIVE SMOKING, 3= UNEXPOSED);
  318    IF PX41=1 THEN SMOKE=1;ELSE IF PX41=2 AND PX42 NE . THEN SMOKE=PX42+1;
  319    KEEP ID2 ID4 ID5 ID6 SMOKE;
  320    RUN;
NOTE: The infile INQUEST is file C:\CARB\QUEST.DAT.
NOTE: 462 records were read from the infile C:\CARB\QUEST.DAT.
      The minimum record length was 291.
      The maximum record length was 291.
NOTE: The data set WORK.QEXTRACT has 117 observations and 5 variables.
NOTE: The DATA statement used 9.00 seconds.
  321
  322    ******************************************************************;
  323    *  Merge chemical and questionnaire data;
  324    *  Since HOUR24.DAT and QUEST.DAT are sorted by ID1 and ID2, no
  325       sort is necessary prior to merge;
  326
  327    DATA CEXTRACT;
  328    MERGE CEXTRACT(IN=CCCC) QEXTRACT(IN=QQQQ);BY ID2;
  329    IF CCCC AND QQQQ;
  330    * CREATE SQUARE OF CONCENTRATIONS (USED FOR ESTIMATION OF POPULATION
  331      VARIANCE);
  332    CPERSQ=CPER*CPER;
  333    * RECODE COMPOUND IDS TO USE AS SUBGROUP VARIABLE IN SUDAAN;
  334    IF CMPD=4 THEN C=1;
  335    ELSE IF CMPD=5 THEN C=2;
  336    ELSE IF CMPD=11 THEN C=3;
  337    ELSE IF CMPD=14 THEN C=4;
  338    ELSE IF CMPD=16 THEN C=5;
  339    ELSE IF CMPD=21 THEN C=6;
  340    * REDEFINE STRATA TO DISTINGUISH COMPOUNDS;
  341    CSTRAT=ID4-2021200+100*C;
  342    DROP CMPD;
  343    RUN;
NOTE: The data set WORK.CEXTRACT has 661 observations and 12 variables.
NOTE: The DATA statement used 7.00 seconds.
```

Exhibit A-1. (continued)

```
 344
 345    ******************************************************************;
 346    * Sort data set by stratum and PSU to use SUDAAN;
 347
 348    PROC SORT DATA=CEXTRACT OUT=OUT.ACHEMDAT;BY CSTRAT ID5 ID2;
 349    RUN;
NOTE: The data set OUT.ACHEMDAT has 661 observations and 12 variables.
NOTE: The PROCEDURE SORT used 5.00 seconds.
 350    DATA OUT.APCHEM1
 351         OUT.APCHEM2
 352         OUT.APCHEM3
 353         OUT.APCHEM4
 354         OUT.APCHEM5
 355         OUT.APCHEM6;
 356    SET OUT.ACHEMDAT;
 357    IF C=1 THEN OUTPUT OUT.APCHEM1;
 358    ELSE IF C=2 THEN OUTPUT OUT.APCHEM2;
 359    ELSE IF C=3 THEN OUTPUT OUT.APCHEM3;
 360    ELSE IF C=4 THEN OUTPUT OUT.APCHEM4;
 361    ELSE IF C=5 THEN OUTPUT OUT.APCHEM5;
 362    ELSE IF C=6 THEN OUTPUT OUT.APCHEM6;
 363    RUN;
NOTE: The data set OUT.APCHEM1 has 111 observations and 12 variables.
NOTE: The data set OUT.APCHEM2 has 111 observations and 12 variables.
NOTE: The data set OUT.APCHEM3 has 110 observations and 12 variables.
NOTE: The data set OUT.APCHEM4 has 110 observations and 12 variables.
NOTE: The data set OUT.APCHEM5 has 109 observations and 12 variables.
NOTE: The data set OUT.APCHEM6 has 110 observations and 12 variables.
NOTE: The DATA statement used 9.00 seconds.
 364    ******************************************************************;
```

Exhibit A-1. (continued)

SUDAAN PROGRAM FOR DATA ANALYSIS

```
                    SUDAAN  Survey Data Analysis Software
          Copyright      Research Triangle Institute      April 1993
                                Release 6.32

1   PROC DESCRIPT DATA="C:\\CARB\\ACHEMDAT" DESIGN=WR FILETYPE=SAS
    MEANS GEOMETRIC NOMARG;
2   WEIGHT WPER;
3   NEST CSTRAT ID5;
4   VAR CPER;
5   SUBGROUP C SMOKE;
6   LEVELS 6 3;
7   TABLES C C*SMOKE;
8   PRINT / FILENAME="AM1CHEM.OUT" REPLACE;
9   OUTPUT NSUM WSUM MEAN GEOMEAN SEMEAN SEGEOMEAN
          / FILENAME="AM1CHEM" REPLACE;


Number of observations read    :    661    Weighted count:  2156955
Denominator degrees of freedom :    114


10  PROC DESCRIPT DATA="C:\\CARB\\ACHEMDAT" DESIGN=WR FILETYPE=SAS
    MEANS GEOMETRIC NOMARG;
11  WEIGHT WPER;
12  NEST CSTRAT ID5;
13  VAR CPER;
14  SUBGROUP C SMOKE;
15  LEVELS 6 3;
16  TABLES C;
17  PAIRWISE SMOKE;
18  PRINT / FILENAME="AM2CHEM.OUT" REPLACE;
19  OUTPUT NSUM WSUM MEAN GEOMEAN SEMEAN SEGEOMEAN
          / FILENAME="AM2CHEM" REPLACE;
```

Exhibit A-1. (continued)

```
Number of observations read    :    661    Weighted count:  2156955
Denominator degrees of freedom :    114


20  PROC DESCRIPT DATA="C:\\CARB\\ACHEMDAT" DESIGN=WR FILETYPE=SAS
    MEANS NOMARG;
21  WEIGHT WPER;
22  NEST CSTRAT ID5;
23  VAR CPERSQ;
24  SUBGROUP C SMOKE;
25  LEVELS 6 3;
26  TABLES C C*SMOKE;
27  PRINT / FILENAME="ACHEMSQ.OUT" REPLACE;
28  OUTPUT NSUM WSUM MEAN
          / FILENAME="ACHEMSQ" REPLACE;


Number of observations read    :    661    Weighted count:  2156955
Denominator degrees of freedom :    114


29  PROC DESCRIPT DATA="C:\\CARB\\APCHEM1" DESIGN=WR FILETYPE=SAS
    EST_NO=120 NOMARG;
30  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
31  WEIGHT WPER;
32  NEST CSTRAT ID5;
33  VAR CPER;
34  SUBGROUP _ONE_ SMOKE;
35  LEVELS 1 3;
36  PRINT / FILENAME="APCHEM1.OUT" REPLACE;
37  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="APCHEM1" REPLACE;


Number of observations read    :    111    Weighted count:   359492
Denominator degrees of freedom :     19
```

Exhibit A-1. (continued)

```
DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.


38  PROC DESCRIPT DATA="C:\\CARB\\APCHEM2" DESIGN=WR FILETYPE=SAS
    EST_NO=120 NOMARG;
39  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
40  WEIGHT WPER;
41  NEST CSTRAT ID5;
42  VAR CPER;
43  SUBGROUP _ONE_ SMOKE;
44  LEVELS 1 3;
45  PRINT / FILENAME="APCHEM2.OUT" REPLACE;
46  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="APCHEM2" REPLACE;
```

Exhibit A-1. (continued)

```
Number of observations read    :    111    Weighted count:    359492
Denominator degrees of freedom :     19

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.

47  PROC DESCRIPT DATA="C:\\CARB\\APCHEM3" DESIGN=WR FILETYPE=SAS
    EST NO=120 NOMARG;
48  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
49  WEIGHT WPER;
50  NEST CSTRAT ID5;
51  VAR CPER;
52  SUBGROUP _ONE_ SMOKE;
53  LEVELS 1 3;
54  PRINT / FILENAME="APCHEM3.OUT" REPLACE;
55  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="APCHEM3" REPLACE;

Number of observations read    :    110    Weighted count:    359492
Denominator degrees of freedom :     19
```

Exhibit A-1. (continued)

```
DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.


56  PROC DESCRIPT DATA="C:\\CARB\\APCHEM4" DESIGN=WR FILETYPE=SAS
    EST_NO=120 NOMARG;
57  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
58  WEIGHT WPER;
59  NEST CSTRAT ID5;
60  VAR CPER;
61  SUBGROUP _ONE_ SMOKE;
62  LEVELS 1 3;
63  PRINT / FILENAME="APCHEM4.OUT" REPLACE;
64  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="APCHEM4" REPLACE;


Number of observations read    :    110    Weighted count:    359492
Denominator degrees of freedom :     19

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.
```

Exhibit A-1. (continued)

```
DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.


65  PROC DESCRIPT DATA="C:\\CARB\\APCHEM5" DESIGN=WR FILETYPE=SAS
    EST NO=120 NOMARG;
66  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
67  WEIGHT WPER;
68  NEST CSTRAT ID5;
69  VAR CPER;
70  SUBGROUP _ONE_ SMOKE;
71  LEVELS 1 3;
72  PRINT / FILENAME="APCHEM5.OUT" REPLACE;
73  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="APCHEM5" REPLACE;


Number of observations read    :    109    Weighted count:    359492
Denominator degrees of freedom :     19

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.
```

Exhibit A-1. (continued)

```
DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.


74  PROC DESCRIPT DATA="C:\\CARB\\APCHEM6" DESIGN=WR FILETYPE=SAS
    EST_NO=120 NOMARG;
75  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
76  WEIGHT WPER;
77  NEST CSTRAT ID5;
78  VAR CPER;
79  SUBGROUP _ONE_ SMOKE;
80  LEVELS 1 3;
81  PRINT / FILENAME="APCHEM6.OUT" REPLACE;
82  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="APCHEM6" REPLACE;


Number of observations read    :    110    Weighted count:    359492
Denominator degrees of freedom :     19

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.
```

Exhibit A-1. (continued)

SAS PROGRAM FOR POST-PROCESSING OF SUDAAN FILES

```
 1    * PGM STORED IN CARB\AANLPOST.SAS;
 2    OPTIONS LS=95 PS=56 NOCENTER NODATE NONUMBER MISSING=' ';
 3    PROC FORMAT;
 4    VALUE CMPDF
 5                  1='1,1,1-Trichloroethane'
 6                  2='Benzene'
 7                  3='Tetrachloroethylene'
 8                  4='Styrene'
 9                  5='m,p-Dichlorobenzene'
10                  6='m,p-Xylene' ;
NOTE: Format CMPDF has been output.
11    VALUE SMOKEF 1='ACTIVE' 2='PASSIVE'
12                 3='UNEXPOSED';
NOTE: Format SMOKEF has been output.
13    VALUE CONTF  1='ACTIVE-PASSIVE'
14                 2='ACTIVE-UNEXPOSED'
15                 3='PASSIVE-UNEXPOSED';
NOTE: Format CONTF has been output.
16    VALUE TABLEF 1='1:  OVERALL POPULATION'
17                 2='2:  BY SMOKING CATEGORY';
NOTE: Format TABLEF has been output.
18    RUN;
NOTE: The PROCEDURE FORMAT used 4.00 seconds.
19    DATA POPVAR1;
20    INFILE 'C:\CARB\AM1CHEM.DBS';
21    INPUT TABLE 2. NDIM 2. C 3. SMOKE 3. +4 N 9. WSUM 11.
22     CM 11. CGM 11. CMSE 11. CGMSE 11.;
23    RUN;
NOTE: The infile 'C:\CARB\AM1CHEM.DBS' is file C:\CARB\AM1CHEM.DBS.
NOTE: 24 records were read from the infile C:\CARB\AM1CHEM.DBS.
      The minimum record length was 78.
      The maximum record length was 78.
```

Exhibit A-1. (continued)

```
NOTE: The data set WORK.POPVAR1 has 24 observations and 10 variables.
NOTE: The DATA statement used 6.00 seconds.
   24
   25    DATA POPVAR2;
   26    INFILE 'C:\CARB\ACHEMSQ.DBS';
   27    INPUT TABLE 2. NDIM 2. C 3. SMOKE 3. +24 CSQM 11.;
   28    RUN;
NOTE: The infile 'C:\CARB\ACHEMSQ.DBS' is file C:\CARB\ACHEMSQ.DBS.
NOTE: 24 records were read from the infile C:\CARB\ACHEMSQ.DBS.
      The minimum record length was 45.
      The maximum record length was 45.
NOTE: The data set WORK.POPVAR2 has 24 observations and 5 variables.
NOTE: The DATA statement used 2.00 seconds.
   29
   30    DATA POPVAR3;MERGE POPVAR1 POPVAR2;
   31    POPSD=SQRT(CSQM-CM*CM);
   32    IF SMOKE<0 THEN SMOKE=.;
   33    RUN;
NOTE: The data set WORK.POPVAR3 has 24 observations and 12 variables.
NOTE: The DATA statement used 2.00 seconds.
   34
   35    PROC PRINT; BY TABLE;ID C SMOKE;
   36    VAR N WSUM CM CMSE CGM CGMSE POPSD;
   37    FORMAT C CMPDF. SMOKE SMOKEF. TABLE TABLEF.
   38    N 4. WSUM 11. CM 8.2 CMSE 8.2 CGM 8.2 CGMSE 8.2 POPSD 8.2;
   39    TITLE
   40    'POPULATION ESTIMATES:  MEANS, GEOMETRIC MEANS, AND STANDARD DEVIATIONS';
   41    RUN;
NOTE: The PROCEDURE PRINT used 3.00 seconds.
   42
   43    DATA CONTRST;
   44    INFILE 'C:\CARB\AM2CHEM.DBS';
   45    INPUT @5 C 3. +7 CONTRST 4. N 9. WSUM 11.
   46     CMDIF 11. CGMDIF 11. CMDIFSE 11. CGMDIFSE 11.;
   47    RUN;
NOTE: The infile 'C:\CARB\AM2CHEM.DBS' is file C:\CARB\AM2CHEM.DBS.
NOTE: 18 records were read from the infile C:\CARB\AM2CHEM.DBS.
      The minimum record length was 82.
      The maximum record length was 82.
```

Exhibit A-1. (continued)

```
NOTE: The data set WORK.CONTRST has 18 observations and 8 variables.
NOTE: The DATA statement used 4.00 seconds.
   48
   49    PROC PRINT; ID C CONTRST;
   50    VAR N WSUM CMDIF CMDIFSE CGMDIF CGMDIFSE;
   51    FORMAT C CMPDF. CONTRST CONTF.
   52    N 4. WSUM 11. CMDIF 8.2 CMDIFSE 8.2 CGMDIF 8.2 CGMDIFSE 8.2;
   53    TITLE
   54    'POPULATION ESTIMATES:  DIFFERENCES IN MEANS & GEOMETRIC MEANS';
   55    RUN;
NOTE: The PROCEDURE PRINT used 2.00 seconds.
   56
   57    DATA PCT;
   58    INFILE 'C:\CARB\APCHEM1.DBS';
   59    INPUT TABLE 2. NDIM 2. +3 SMOKE 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
   60    SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
   61    IF SMOKE<0 THEN SMOKE=.;
   62    C=1;
   63    RUN;
NOTE: The infile 'C:\CARB\APCHEM1.DBS' is file C:\CARB\APCHEM1.DBS.
NOTE: 20 records were read from the infile C:\CARB\APCHEM1.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 20 observations and 11 variables.
NOTE: The DATA statement used 2.00 seconds.
   64
   65    PROC PRINT; BY C TABLE;ID SMOKE;
   66    VAR PCTILE N WSUM QTILE SEQTILE LOWQTILE UPQTILE;
   67    FORMAT C CMPDF. TABLE TABLEF. SMOKE SMOKEF.
   68    N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
   69    TITLE
   70    'POPULATION ESTIMATES:  PERCENTILES';
   71    RUN;
NOTE: The PROCEDURE PRINT used 2.00 seconds.
   72
   73    DATA PCT;
   74    INFILE 'C:\CARB\APCHEM2.DBS';
   75    INPUT TABLE 2. NDIM 2. +3 SMOKE 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
   76    SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
```

Exhibit A-1. (continued)

```
   77     IF SMOKE<0 THEN SMOKE=.;
   78     C=2;
   79     RUN;
NOTE: The infile 'C:\CARB\APCHEM2.DBS' is file C:\CARB\APCHEM2.DBS.
NOTE: 20 records were read from the infile C:\CARB\APCHEM2.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 20 observations and 11 variables.
NOTE: The DATA statement used 1.00 seconds.
   80
   81     PROC PRINT; BY C TABLE;ID SMOKE;
   82     VAR PCTILE N WSUM QTILE SEQTILE LOWQTILE UPQTILE;
   83     FORMAT C CMPDF. TABLE TABLEF. SMOKE SMOKEF.
   84     N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
   85     TITLE
   86     'POPULATION ESTIMATES:  PERCENTILES';
   87     RUN;
NOTE: The PROCEDURE PRINT used 3.00 seconds.
   88
   89     DATA PCT;
   90     INFILE 'C:\CARB\APCHEM3.DBS';
   91     INPUT TABLE 2. NDIM 2. +3 SMOKE 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
   92     SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
   93     IF SMOKE<0 THEN SMOKE=.;
   94     C=3;
   95     RUN;
NOTE: The infile 'C:\CARB\APCHEM3.DBS' is file C:\CARB\APCHEM3.DBS.
NOTE: 20 records were read from the infile C:\CARB\APCHEM3.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 20 observations and 11 variables.
NOTE: The DATA statement used 1.00 seconds.
   96
   97     PROC PRINT; BY C TABLE;ID SMOKE;
   98     VAR PCTILE N WSUM QTILE SEQTILE LOWQTILE UPQTILE;
   99     FORMAT C CMPDF. TABLE TABLEF. SMOKE SMOKEF.
  100     N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
  101     TITLE
  102     'POPULATION ESTIMATES:  PERCENTILES';
  103     RUN;
NOTE: The PROCEDURE PRINT used 2.00 seconds.
```

Exhibit A-1. (continued)

```
  104
  105    DATA PCT;
  106    INFILE 'C:\CARB\APCHEM4.DBS';
  107    INPUT TABLE 2. NDIM 2. +3 SMOKE 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
  108    SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
  109    IF SMOKE<0 THEN SMOKE=.;
  110    C=4;
  111    RUN;
NOTE: The infile 'C:\CARB\APCHEM4.DBS' is file C:\CARB\APCHEM4.DBS.
NOTE: 20 records were read from the infile C:\CARB\APCHEM4.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 20 observations and 11 variables.
NOTE: The DATA statement used 2.00 seconds.
  112
  113    PROC PRINT; BY C TABLE;ID SMOKE;
  114    VAR PCTILE N WSUM QTILE SEQTILE LOWQTILE UPQTILE;
  115    FORMAT C CMPDF. TABLE TABLEF. SMOKE SMOKEF.
  116    N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
  117    TITLE
  118    'POPULATION ESTIMATES:  PERCENTILES';
  119    RUN;
NOTE: The PROCEDURE PRINT used 2.00 seconds.
  120
  121    DATA PCT;
  122    INFILE 'C:\CARB\APCHEM5.DBS';
  123    INPUT TABLE 2. NDIM 2. +3 SMOKE 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
  124    SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
  125    IF SMOKE<0 THEN SMOKE=.;
  126    C=5;
  127    RUN;
NOTE: The infile 'C:\CARB\APCHEM5.DBS' is file C:\CARB\APCHEM5.DBS.
NOTE: 20 records were read from the infile C:\CARB\APCHEM5.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 20 observations and 11 variables.
NOTE: The DATA statement used 1.00 seconds.
  128
  129    PROC PRINT; BY C TABLE;ID SMOKE;
  130    VAR PCTILE N WSUM QTILE SEQTILE LOWQT.   UPQTILE;
```

Exhibit A-1. (continued)

```
  131     FORMAT C CMPDF. TABLE TABLEF. SMOKE SMOKEF.
  132     N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
  133     TITLE
  134     'POPULATION ESTIMATES:  PERCENTILES';
  135     RUN;
NOTE: The PROCEDURE PRINT used 3.00 seconds.
  136
  137     DATA PCT;
  138     INFILE 'C:\CARB\APCHEM6.DBS';
  139     INPUT TABLE 2. NDIM 2. +3 SMOKE 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
  140     SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
  141     IF SMOKE<0 THEN SMOKE=.;
  142     C=6;
  143     RUN;
NOTE: The infile 'C:\CARB\APCHEM6.DBS' is file C:\CARB\APCHEM6.DBS.
NOTE: 20 records were read from the infile C:\CARB\APCHEM6.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 20 observations and 11 variables.
NOTE: The DATA statement used 2.00 seconds.
  144
  145     PROC PRINT; BY C TABLE;ID SMOKE;
  146     VAR PCTILE N WSUM QTILE SEQTILE LOWQTILE UPQTILE;
  147     FORMAT C CMPDF. TABLE TABLEF. SMOKE SMOKEF.
  148     N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
  149     TITLE
  150     'POPULATION ESTIMATES:  PERCENTILES';
  151     RUN;
NOTE: The PROCEDURE PRINT used 2.00 seconds.
```

Exhibit A-2. Analysis Type A: Program Logs for Example A2

SAS PROGRAM FOR PREPARING FILES FOR SUDAAN ANALYSIS

```
 1    * PGM STORED IN CARB\AANLYSI2.SAS;
 2    FILENAME INHR24 'C:\CARB\HOUR24.DAT';
 3    FILENAME INQUEST 'C:\CARB\QUEST.DAT';
 4    LIBNAME OUT 'C:\CARB';
 5    OPTIONS LS=95 PS=56 NOCENTER NODATE NONUMBER;
 6
 7    ******************************************************************;
 8    * Extract pertinent 24-hour chemical data;
 9    DATA CEXTRACT;
10    INFILE INHR24;
11    INPUT
12
13    /* FOLLOWING STATEMENTS WERE EXTRACTED FROM HOUR24.CBK */
14
15    @1    ID1    1.    /*STUDY NUMBER*/
16    @2    ID2    7.    /*HOUSEHOLD/PARTICIPANT ID*/
17    @9    CMPD   2.    /*COMPOUND ID*/
18    @29   CMLA   9.2   /*CONCENTRATION FOR MAIN LIVING AREA*/
19    @49   MMLA   1.    /*MEASURABILITY INDICATOR FOR MAIN LIVING AREA*/
20    @61   WMLA  10.4   /*SAMPLING WEIGHT FOR MAIN LIVING AREA*/
21     ;
22
23    IF ID1=6;        /* SELECT RECORDS FOR STUDY 6 */
24    IF MMLA >=1;     /* SELECT RECORDS WITH INDOOR AIR DATA */
25    IF CMPD=4 OR CMPD=5 OR CMPD=11 OR CMPD=14 OR CMPD=21 OR CMPD=17;
26       /* SELECT COMPOUNDS */
27      * FOR STUDY 6, WE ADOPT THE FOLLOWING RULE FOR EACH COMPOUND:
28      IF THERE ARE SOME NONQUANTIFIABLE VALUES THAT ARE POSITIVE AND
29      SOME THAT ARE ZERO, THEN SET ZERO VALUES TO 0.5*SMALLEST POSITIVE
30      VALUE.  IF ALL NONQUANTIFIABLE VALUES ARE ZERO, THEN SET THEM TO
31      0.5*(MEDIAN QL VALUE).  THIS RESULTS IN THE FOLLOWING;
32    IF CMLA=0 THEN DO;
33       IF CMPD=4 THEN CMLA=0.06;
```

Exhibit A-2. (continued)

```
34        ELSE IF CMPD=5 THEN CMLA=0.035;
35        ELSE IF CMPD=11 OR CMPD=14 OR CMPD=17 THEN
36              CMLA=0.005;
37        ELSE IF CMPD=20 THEN CMLA=0.055;
38        ELSE IF CMPD=21 THEN CMLA=0.175;
39     END;
40     RUN;
NOTE: The infile INHR24 is file C:\CARB\HOUR24.DAT.
NOTE: 8797 records were read from the infile C:\CARB\HOUR24.DAT.
      The minimum record length was 80.
      The maximum record length was 80.
NOTE: The data set WORK.CEXTRACT has 736 observations and 6 variables.
NOTE: The DATA statement used 33.00 seconds.
41
42     ******************************************************************;
43     * Extract pertinent questionnaire data;
44     DATA QEXTRACT;
45     INFILE INQUEST LRECL=291;
46     INPUT
47
48     /* FOLLOWING STATEMENTS WERE EXTRACTED FROM QUEST.CBK */
49
50     @1   ID1   1.     /*Study ID */
51     @2   ID2   7.     /*Household/Participant ID */
52     @16  ID4   7.     /*Stratum Code */
53     @23  ID5   3.     /*Primary Sampling Unit Code */
54     @139 HG40  1.     /*Tobacco Prod. Smoked in House */
55     ;
56     IF ID1=6;       /* SELECT RECORDS FOR STUDY 6 */
57     KEEP ID2 ID4 ID5 HG40;
58     RUN;
NOTE: The infile INQUEST is file C:\CARB\QUEST.DAT.
NOTE: 462 records were read from the infile C:\CARB\QUEST.DAT.
      The minimum record length was 291.
      The maximum record length was 291.
NOTE: The data set WORK.QEXTRACT has 128 observations and 4 variables.
NOTE: The DATA statement used 4.00 seconds.
59
```

Exhibit A-2. (continued)

```
60    ************************************************************************;
61    *  Merge chemical and questionnaire data;
62    *  Since HOUR24.DAT and QUEST.DAT are sorted by ID1 and ID2, no
63       sort is necessary prior to merge;
64
65    DATA CEXTRACT;
66    MERGE CEXTRACT(IN=CCCC) QEXTRACT(IN=QQQQ);BY ID2;
67    IF CCCC AND QQQQ;
68    * CREATE SQUARE OF CONCENTRATIONS (USED FOR ESTIMATION OF POPULATION
69      VARIANCE);
70    CMLASQ=CMLA*CMLA;
71    * RECODE COMPOUND IDS TO USE AS SUBGROUP VARIABLE IN SUDAAN;
72    IF CMPD=4 THEN C=1;
73    ELSE IF CMPD=5 THEN C=2;
74    ELSE IF CMPD=11 THEN C=3;
75    ELSE IF CMPD=14 THEN C=4;
76    ELSE IF CMPD=17 THEN C=5;
77    ELSE IF CMPD=21 THEN C=6;
78    * REDEFINE STRATA TO DISTINGUISH COMPOUNDS;
79    CSTRAT=ID4+10*C;
80    DROP CMPD;
81    RUN;
NOTE: The data set WORK.CEXTRACT has 736 observations and 11 variables.
NOTE: The DATA statement used 7.00 seconds.
82
83    ************************************************************************;
84    * Sort data set by stratum and PSU to use SUDAAN;
85
86    PROC SORT DATA=CEXTRACT OUT=OUT.A6CHMDAT;BY CSTRAT ID5 ID2;
87    RUN;
NOTE: The data set OUT.A6CHMDAT has 736 observations and 11 variables.
NOTE: The PROCEDURE SORT used 7.00 seconds.
88    DATA OUT.AICHEM1
89         OUT.AICHEM2
90         OUT.AICHEM3
91         OUT.AICHEM4
92         OUT.AICHEM5
93         OUT.AICHEM6;
```

Exhibit A-2. (continued)

```
 94    SET OUT.A6CHMDAT;
 95    IF C=1 THEN OUTPUT OUT.AICHEM1;
 96    ELSE IF C=2 THEN OUTPUT OUT.AICHEM2;
 97    ELSE IF C=3 THEN OUTPUT OUT.AICHEM3;
 98    ELSE IF C=4 THEN OUTPUT OUT.AICHEM4;
 99    ELSE IF C=5 THEN OUTPUT OUT.AICHEM5;
100    ELSE IF C=6 THEN OUTPUT OUT.AICHEM6;
101    RUN;
NOTE: The data set OUT.AICHEM1 has 115 observations and 11 variables.
NOTE: The data set OUT.AICHEM2 has 124 observations and 11 variables.
NOTE: The data set OUT.AICHEM3 has 124 observations and 11 variables.
NOTE: The data set OUT.AICHEM4 has 123 observations and 11 variables.
NOTE: The data set OUT.AICHEM5 has 125 observations and 11 variables.
NOTE: The data set OUT.AICHEM6 has 125 observations and 11 variables.
NOTE: The DATA statement used 11.00 seconds.
215    ***********************************************************************;
```

SUDAAN PROGRAM FOR DATA ANALYSIS

```
                    SUDAAN  Survey Data Analysis Software
Copyright      Research Triangle Institute      June 1993
                         Release 6.34

1   PROC DESCRIPT DATA="C:\\CARB\\A6CHMDAT" DESIGN=WR FILETYPE=SAS
    MEANS GEOMETRIC NOMARG;
2   WEIGHT WMLA;
3   NEST CSTRAT ID5;
4   VAR CMLA;
5   SUBGROUP C HG40;
6   LEVELS 6 2;
7   TABLES C C*HG40;
8   PRINT / FILENAME="AIM1CHEM.OUT" REPLACE;
9   OUTPUT NSUM WSUM MEAN GEOMEAN SEMEAN SEGEOMEAN
          / FILENAME="AIM1CHEM" REPLACE;
```

Exhibit A-2. (continued)

```
Number of observations read    :    736    Weighted count:    86104
Denominator degrees of freedom :    143


10  PROC DESCRIPT DATA="C:\\CARB\\A6CHMDAT" DESIGN=WR FILETYPE=SAS
    MEANS GEOMETRIC NOMARG;
11  WEIGHT WMLA;
12  NEST CSTRAT ID5;
13  VAR CMLA;
14  SUBGROUP C HG40;
15  LEVELS 6 2;
16  TABLES C;
17  PAIRWISE HG40;
18  PRINT / FILENAME="AIM2CHEM.OUT" REPLACE;
19  OUTPUT NSUM WSUM MEAN GEOMEAN SEMEAN SEGEOMEAN
          / FILENAME="AIM2CHEM" REPLACE;


Number of observations read    :    736    Weighted count:    86104
Denominator degrees of freedom :    143


20  PROC DESCRIPT DATA="C:\\CARB\\A6CHMDAT" DESIGN=WR FILETYPE=SAS
    MEANS NOMARG;
21  WEIGHT WMLA;
22  NEST CSTRAT ID5;
23  VAR CMLASQ;
24  SUBGROUP C HG40;
25  LEVELS 6 2;
26  TABLES C C*HG40;
27  PRINT / FILENAME="AICHEMSQ.OUT" REPLACE;
28  OUTPUT NSUM WSUM MEAN
          / FILENAME="AICHEMSQ" REPLACE;


Number of observations read    :    736    Weighted count:    86104
Denominator degrees of freedom :    143
```

Exhibit A-2. (continued)

```
29  PROC DESCRIPT DATA="C:\\CARB\\AICHEM1" DESIGN=WR FILETYPE=SAS
    EST_NO=130 NOMARG;
30  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
31  WEIGHT WMLA;
32  NEST CSTRAT ID5;
33  VAR CMLA;
34  SUBGROUP _ONE_ HG40;
35  LEVELS 1 2;
36  PRINT / FILENAME="AICHEM1.OUT" REPLACE;
37  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="AICHEM1" REPLACE;


Number of observations read    :    115    Weighted count:    13346
Denominator degrees of freedom :     24

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.


38  PROC DESCRIPT DATA="C:\\CARB\\AICHEM2" DESIGN=WR FILETYPE=SAS
    EST_NO=130 NOMARG;
39  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
40  WEIGHT WMLA;
41  NEST CSTRAT ID5;
42  VAR CMLA;
43  SUBGROUP _ONE_ HG40;
44  LEVELS 1 2;
45  PRINT / FILENAME="AICHEM2.OUT" REPLACE;
46  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="AICHEM2" REPLACE;
```

Exhibit A-2. (continued)

```
Number of observations read    :    124    Weighted count:    14534
Denominator degrees of freedom :     24

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.


47  PROC DESCRIPT DATA="C:\\CARB\\AICHEM3" DESIGN=WR FILETYPE=SAS
    EST_NO=130 NOMARG;
48  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
49  WEIGHT WMLA;
50  NEST CSTRAT ID5;
51  VAR CMLA;
52  SUBGROUP _ONE_ HG40;
53  LEVELS 1 2;
54  PRINT / FILENAME="AICHEM3.OUT" REPLACE;
55  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="AICHEM3" REPLACE;


Number of observations read    :    124    Weighted count:    14541
Denominator degrees of freedom :     23

DATA WARNING:
Cannot extrapolate to compute the 10.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute the 10.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 75.00th percentile.
Generating a missing value.
```

Exhibit A-2. (continued)

```
DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.


56  PROC DESCRIPT DATA="C:\\CARB\\AICHEM4" DESIGN=WR FILETYPE=SAS
    EST NO=130 NOMARG;
57  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
58  WEIGHT WMLA;
59  NEST CSTRAT ID5;
60  VAR CMLA;
61  SUBGROUP _ONE_ HG40;
62  LEVELS 1 2;
63  PRINT / FILENAME="AICHEM4.OUT" REPLACE;
64  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="AICHEM4" REPLACE;


Number of observations read    :    123    Weighted count:    14435
Denominator degrees of freedom :     24

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.
```

Exhibit A-2. (continued)

```
DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.

DATA WARNING:
Cannot extrapolate to compute confidence limit for 10.00th percentile.
Generating a missing value.


65  PROC DESCRIPT DATA="C:\\CARB\\AICHEM5" DESIGN=WR FILETYPE=SAS
    EST_NO=130 NOMARG;
66  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
67  WEIGHT WMLA;
68  NEST CSTRAT ID5;
69  VAR CMLA;
70  SUBGROUP _ONE_ HG40;
71  LEVELS 1 2;
72  PRINT / FILENAME="AICHEM5.OUT" REPLACE;
73  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="AICHEM5" REPLACE;


Number of observations read    :    125    Weighted count:    14624
Denominator degrees of freedom :     24

DATA WARNING:
Cannot extrapolate to compute confidence limit for 90.00th percentile.
Generating a missing value.


74  PROC DESCRIPT DATA="C:\\CARB\\AICHEM6" DESIGN=WR FILETYPE=SAS
    EST_NO=130 NOMARG;
75  PERCENTILES 10 90 / MEDIAN QUARTILES UNGROUPED;
76  WEIGHT WMLA;
77  NEST CSTRAT ID5;
78  VAR CMLA;
79  SUBGROUP _ONE_ HG40;
80  LEVELS 1 2;
```

Exhibit A-2. (continued)

```
81  PRINT / FILENAME="AICHEM6.OUT" REPLACE;
82  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="AICHEM6" REPLACE;
```

```
Number of observations read    :    125    Weighted count:    14624
Denominator degrees of freedom :     24
```

SAS PROGRAM FOR POST-PROCESSING OF SUDAAN FILES

```
    1    * PGM STORED IN CARB\AANLPOS2.SAS;
    2    OPTIONS LS=95 PS=56 NOCENTER NODATE NONUMBER MISSING=' ';
    3    PROC FORMAT;
    4    VALUE CMPDF
    5                  1='1,1,1-Trichloroethane'
    6                  2='Benzene'
    7                  3='Tetrachloroethylene'
    8                  4='Styrene'
    9                  5='p-Dichlorobenzene'
   10                  6='m,p-Xylene' ;
NOTE: Format CMPDF has been output.
   11    VALUE HG40F  1='SMOKE' 2='NONSMOKE';
NOTE: Format HG40F has been output.
   12    VALUE CONTF  1='SMOKE-NONSMOKE';
NOTE: Format CONTF has been output.
   13    VALUE TABLEF 1='1:  OVERALL POPULATION'
   14                 2='2:  BY SMOKING CATEGORY';
NOTE: Format TABLEF has been output.
   15    RUN;
NOTE: The PROCEDURE FORMAT used 4.00 seconds.
   16    DATA POPVAR1;
   17    INFILE 'C:\CARB\AIM1CHEM.DBS';
   18    INPUT TABLE 2. NDIM 2. C 3. HG40 3. +4 N 9. WSUM 11.
   19     CM 11. CGM 11. CMSE 11. CGMSE 11.;
   20    RUN;
```

Exhibit A-2. (continued)

```
NOTE: The infile 'C:\CARB\AIM1CHEM.DBS' is file C:\CARB\AIM1CHEM.DBS.
NOTE: 18 records were read from the infile C:\CARB\AIM1CHEM.DBS.
      The minimum record length was 78.
      The maximum record length was 78.
NOTE: The data set WORK.POPVAR1 has 18 observations and 10 variables.
NOTE: The DATA statement used 2.00 seconds.
   21
   22    DATA POPVAR2;
   23    INFILE 'C:\CARB\AICHEMSQ.DBS';
   24    INPUT TABLE 2. NDIM 2. C 3. HG40 3. +24 CSQM 11.;
   25    RUN;
NOTE: The infile 'C:\CARB\AICHEMSQ.DBS' is file C:\CARB\AICHEMSQ.DBS.
NOTE: 18 records were read from the infile C:\CARB\AICHEMSQ.DBS.
      The minimum record length was 45.
      The maximum record length was 45.
NOTE: The data set WORK.POPVAR2 has 18 observations and 5 variables.
NOTE: The DATA statement used 1.00 seconds.
   26
   27    DATA POPVAR3;MERGE POPVAR1 POPVAR2;
   28    POPSD=SQRT(CSQM-CM*CM);
   29    IF HG40<0 THEN HG40=.;
   30    RUN;
NOTE: The data set WORK.POPVAR3 has 18 observations and 12 variables.
NOTE: The DATA statement used 2.00 seconds.
   31
   32    PROC PRINT; BY TABLE;ID C HG40;
   33    VAR N WSUM CM CMSE CGM CGMSE POPSD;
   34    FORMAT C CMPDF. HG40 HG40F. TABLE TABLEF.
   35    N 4. WSUM 11. CM 8.2 CMSE 8.2 CGM 8.2 CGMSE 8.2 POPSD 8.2;
   36    TITLE
   37    'POPULATION ESTIMATES:  MEANS, GEOMETRIC MEANS, AND STANDARD DEVIATIONS';
   38    RUN;
NOTE: The PROCEDURE PRINT used 3.00 seconds.
   39
   40    DATA CONTRST;
   41    INFILE 'C:\CARB\AIM2CHEM.DBS';
   42    INPUT @5 C 3. +7 CONTRST 4. N 9. WSUM 11.
   43     CMDIF 11. CGMDIF 11. CMDIFSE 11. CGMDIFSE 11.;
   44    RUN;
```

Exhibit A-2. (continued)

```
NOTE: The infile 'C:\CARB\AIM2CHEM.DBS' is file C:\CARB\AIM2CHEM.DBS.
NOTE: 6 records were read from the infile C:\CARB\AIM2CHEM.DBS.
      The minimum record length was 82.
      The maximum record length was 82.
NOTE: The data set WORK.CONTRST has 6 observations and 8 variables.
NOTE: The DATA statement used 1.00 seconds.
   45
   46    PROC PRINT; ID C CONTRST;
   47    VAR N WSUM CMDIF CMDIFSE CGMDIF CGMDIFSE;
   48    FORMAT C CMPDF. CONTRST CONTF.
   49    N 4. WSUM 11. CMDIF 8.2 CMDIFSE 8.2 CGMDIF 8.2 CGMDIFSE 8.2;
   50    TITLE
   51    'POPULATION ESTIMATES:  DIFFERENCES IN MEANS & GEOMETRIC MEANS';
   52    RUN;
NOTE: The PROCEDURE PRINT used 2.00 seconds.
   53
   54    DATA PCT;
   55    INFILE 'C:\CARB\AICHEM1.DBS';
   56    INPUT TABLE 2. NDIM 2. +3 HG40 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
   57    SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
   58    IF HG40<0 THEN HG40=.;
   59    C=1;
   60    RUN;
NOTE: The infile 'C:\CARB\AICHEM1.DBS' is file C:\CARB\AICHEM1.DBS.
NOTE: 15 records were read from the infile C:\CARB\AICHEM1.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 15 observations and 11 variables.
NOTE: The DATA statement used 1.00 seconds.
   61
   62    PROC PRINT; BY C TABLE;ID HG40;
   63    VAR PCTILE N WSUM QTILE SEQTILE LOWQTILE UPQTILE;
   64    FORMAT C CMPDF. TABLE TABLEF. HG40 HG40F.
   65    N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
   66    TITLE
   67    'POPULATION ESTIMATES:  PERCENTILES';
   68    RUN;
NOTE: The PROCEDURE PRINT used 2.00 seconds.
   69
```

Exhibit A-2. (continued)

```
   70    DATA PCT;
   71    INFILE 'C:\CARB\AICHEM2.DBS';
   72    INPUT TABLE 2. NDIM 2. +3 HG40 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
   73    SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
   74    IF HG40<0 THEN HG40=.;
   75    C=2;
   76    RUN;
NOTE: The infile 'C:\CARB\AICHEM2.DBS' is file C:\CARB\AICHEM2.DBS.
NOTE: 15 records were read from the infile C:\CARB\AICHEM2.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 15 observations and 11 variables.
NOTE: The DATA statement used 2.00 seconds.
   77
   78    PROC PRINT; BY C TABLE;ID HG40;
   79    VAR PCTILE N WSUM QTILE SEQTILE LOWQTILE UPQTILE;
   80    FORMAT C CMPDF. TABLE TABLEF. HG40 HG40F.
   81    N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
   82    TITLE
   83    'POPULATION ESTIMATES:  PERCENTILES';
   84    RUN;
NOTE: The PROCEDURE PRINT used 4.00 seconds.
   85
   86    DATA PCT;
   87    INFILE 'C:\CARB\AICHEM3.DBS';
   88    INPUT TABLE 2. NDIM 2. +3 HG40 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
   89    SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
   90    IF HG40<0 THEN HG40=.;
   91    C=3;
   92    RUN;
NOTE: The infile 'C:\CARB\AICHEM3.DBS' is file C:\CARB\AICHEM3.DBS.
NOTE: 15 records were read from the infile C:\CARB\AICHEM3.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 15 observations and 11 variables.
NOTE: The DATA statement used 2.00 seconds.
   93
   94    PROC PRINT; BY C TABLE;ID HG40;
   95    VAR PCTILE N WSUM QTILE SEQTILE LOWQT[illegible] UPQTILE;
   96    FORMAT C CMPDF. TABLE TABLEF. HG40 H[illegible]
```

Exhibit A-2. (continued)

```
   97    N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
   98    TITLE
   99    'POPULATION ESTIMATES:  PERCENTILES';
  100    RUN;
NOTE: The PROCEDURE PRINT used 4.00 seconds.
  101
  102    DATA PCT;
  103    INFILE 'C:\CARB\AICHEM4.DBS';
  104    INPUT TABLE 2. NDIM 2. +3 HG40 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
  105    SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
  106    IF HG40<0 THEN HG40=.;
  107    C=4;
  108    RUN;
NOTE: The infile 'C:\CARB\AICHEM4.DBS' is file C:\CARB\AICHEM4.DBS.
NOTE: 15 records were read from the infile C:\CARB\AICHEM4.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 15 observations and 11 variables.
NOTE: The DATA statement used 2.00 seconds.
  109
  110    PROC PRINT; BY C TABLE;ID HG40;
  111    VAR PCTILE N WSUM QTILE SEQTILE LOWQTILE UPQTILE;
  112    FORMAT C CMPDF. TABLE TABLEF. HG40 HG40F.
  113    N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
  114    TITLE
  115    'POPULATION ESTIMATES:  PERCENTILES';
  116    RUN;
NOTE: The PROCEDURE PRINT used 2.00 seconds.
  117
  118    DATA PCT;
  119    INFILE 'C:\CARB\AICHEM5.DBS';
  120    INPUT TABLE 2. NDIM 2. +3 HG40 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
  121    SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
  122    IF HG40<0 THEN HG40=.;
  123    C=5;
  124    RUN;
NOTE: The infile 'C:\CARB\AICHEM5.DBS' is file C:\CARB\AICHEM5.DBS.
NOTE: 15 records were read from the infile C:\CARB\AICHEM5.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
```

Exhibit A-2. (continued)

```
NOTE: The data set WORK.PCT has 15 observations and 11 variables.
NOTE: The DATA statement used 2.00 seconds.
  125
  126     PROC PRINT; BY C TABLE;ID HG40;
  127     VAR PCTILE N WSUM QTILE SEQTILE LOWQTILE UPQTILE;
  128     FORMAT C CMPDF. TABLE TABLEF. HG40 HG40F.
  129     N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
  130     TITLE
  131     'POPULATION ESTIMATES:  PERCENTILES';
  132     RUN;
NOTE: The PROCEDURE PRINT used 2.00 seconds.
  133
  134     DATA PCT;
  135     INFILE 'C:\CARB\AICHEM6.DBS';
  136     INPUT TABLE 2. NDIM 2. +3 HG40 3. +4 PCTILE 6. N 9. WSUM 11. QTILE 11.
  137     SEQTILE 11. LOWQTILE 11. UPQTILE 11.;
  138     IF HG40<0 THEN HG40=.;
  139     C=6;
  140     RUN;
NOTE: The infile 'C:\CARB\AICHEM6.DBS' is file C:\CARB\AICHEM6.DBS.
NOTE: 15 records were read from the infile C:\CARB\AICHEM6.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.PCT has 15 observations and 11 variables.
NOTE: The DATA statement used 3.00 seconds.
  141
  142     PROC PRINT; BY C TABLE;ID HG40;
  143     VAR PCTILE N WSUM QTILE SEQTILE LOWQTILE UPQTILE;
  144     FORMAT C CMPDF. TABLE TABLEF. HG40 HG40F.
  145     N 4. WSUM 11. PCTILE 3. QTILE SEQTILE LOWQTILE UPQTILE 8.2;
  146     TITLE
  147     'POPULATION ESTIMATES:  PERCENTILES';
  148     RUN;
NOTE: The PROCEDURE PRINT used 2.00 seconds.
```

APPENDIX II-B

TYPE B ANALYSIS

APPENDIX II-B: TYPE B ANALYSIS

Though the programs for file preparation and analysis (see Exhibit B-1) are similar to those for Type A (Appendix II-A), there are some differences. For instance, concentration data for compound 16 (m,p-dichlorobenzene), available for Studies 1 and 3, were combined with data for compound 17, available for Study 6 (i.e., a single variable was created). Study ID (variable ID1) was recoded to a variable called STUDY having values of 1, 2, and 3 so that it could be used as a subgroup identifier in the SUDAAN programs. Formatted outputs were also provided for this variable. In the post processing of the SUDAAN output files, the results associated with the estimated medians were combined with the results for the means and GMs. Outputs produced by the programs in Exhibit B-1 can be found in Exhibit 5-1 of the main text.

Exhibit B-1. Analysis Type B: Program Logs

SAS PROGRAM FOR PREPARING FILES FOR SUDAAN ANALYSIS

```
 1    * PGM STORED IN CARB\BANLYSIS.SAS;
 2    FILENAME INHR24 'C:\CARB\HOUR24.DAT';
 3    FILENAME INQUEST 'C:\CARB\QUEST.DAT';
 4    LIBNAME OUT 'C:\CARB';
 5    OPTIONS LS=95 PS=56 NOCENTER NODATE NONUMBER;
 6
 7    ******************************************************************;
 8    * Extract pertinent 24-hour chemical data;
 9    DATA CEXTRACT;
10    INFILE INHR24;
11    INPUT
12
13    /* FOLLOWING STATEMENTS WERE EXTRACTED FROM HOUR24.CBK */
14
15    @1    ID1    1.    /*STUDY NUMBER*/
16    @2    ID2    7.    /*HOUSEHOLD/PARTICIPANT ID*/
17    @9    CMPD   2.    /*COMPOUND ID*/
18    @11   CPER   9.2   /*CONCENTRATION FOR PERSONAL AIR*/
19    @47   MPER   1.    /*MEASURABILITY INDICATOR FOR PERSONAL AIR*/
20    @51   WPER  10.4   /*SAMPLING WEIGHT FOR PERSONAL AIR*/
21     ;
22
23    IF ID1=1 OR ID1=3 OR ID1=6; /* SELECT RECORDS FOR STUDIES 1,3,6 */
24    IF MPER >=1;     /* SELECT RECORDS WITH PERSONAL AIR DATA */
25    IF CMPD=4 OR CMPD=5 OR CMPD=11 OR CMPD=14 OR CMPD=21 OR CMPD=16
26       OR CMPD=17;IF CMPD=17 THEN CMPD=16;
27       /* SELECT COMPOUNDS & CONVERT CMPD 17 TO CMPD 16) */
28    RUN;
NOTE: The infile INHR24 is file C:\CARB\HOUR24.DAT.
NOTE: 8797 records were read from the infile C:\CARB\HOUR24.DAT.
      The minimum record length was 80.
      The maximum record length was 80.
```

Exhibit B-1. (continued)

```
NOTE: The data set WORK.CEXTRACT has 1619 observations and 6 variables.
NOTE: The DATA statement used 42.00 seconds.
   29
   30    ******************************************************************;
   31    * Extract pertinent questionnaire data;
   32    DATA QEXTRACT;
   33    INFILE INQUEST LRECL=291;
   34    INPUT
   35
   36    /* FOLLOWING STATEMENTS WERE EXTRACTED FROM QUEST.CBK */
   37
   38    @1   ID1   1.     /*Study ID */
   39    @2   ID2   7.     /*Household/Participant ID */
   40    @16  ID4   7.     /*Stratum Code */
   41    @23  ID5   3.     /*Primary Sampling Unit Code */
   42    @127 PG36  2.     /*Smoking Status (Current Smoker, Former Smoker, Never Smoked) */
   43    @129 PG37  1.     /*Avg. No. Cigarettes/Day (<10, 10-30, >30) */
   44    @130 HG37  1.     /*HH Members Smoke Cigarettes (Participant Only, Others Only,
   45                                                     Both, Neither) */
   46    @131 PG38  1.     /*Avg. Hrs/Day Enclosed w/Smoker-Home (0, 1-5, 6-10, >10) */
   47    @132 PG39  1.     /*Avg. Hrs/Day Enclosed w/Smoker-Work (0, 1-5, 6-10, >10) */
   48    @133 HX38  3.     /*No. Cigarettes Smoked in House */
   49    @136 HX39  3.     /*No. Cigarettes, Cigars, or Pipefuls */
   50    @139 HG40  1.     /*Tobacco Prod. Smoked in House */
   51    @140 PX40  3.     /*No. Cigarettes Smoked */
   52    @143 PD40  3.     /*No. Cigarettes Smoked - Daytime */
   53    @146 PN40  3.     /*No. Cigarettes Smoked - Nighttime */
   54    @149 PX41  1.     /*Smoked Tobacco Products */
   55    @150 PD41  1.     /*Smoked Tobacco Products - Day */
   56    @151 PN41  1.     /*Smoked Tobacco Products - Night */
   57    @152 PX42  1.     /*Enclosed Area w/Smoker(s) */
   58    @153 PD42  1.     /*Enclosed Area w/Smokers - Day */
   59    @154 PN42  1.     /*Enclosed Area w/Smokers - Night */
   60    @155 PX43  6.2    /*% Time Indoors at Home with Smokers */
   61    @161 PX44  6.2    /*% Time Indoors Not at Home w/ Smokers */
   62    @167 PX45  6.2    /*% Time Enclosed in Transit w/ Smokers */
   63    ;
```

Exhibit B-1. (continued)

```
   64
   65    IF ID1=1 OR ID1=3 OR ID1=6; /* SELECT RECORDS FOR STUDIES 1,3,6 */
   66
   67    * CREATE DOMAIN DEFINING VARIABLE NAMED SMOKE
   68              (1=SMOKER, 2=EXPOSED TO PASSIVE SMOKING, 3= UNEXPOSED);
   69    IF PX41=1 THEN SMOKE=1;ELSE IF PX41=2 AND PX42 NE . THEN SMOKE=PX42+1;
   70    KEEP ID1 ID2 ID4 ID5 SMOKE;
   71    RUN;
NOTE: The infile INQUEST is file C:\CARB\QUEST.DAT.
NOTE: 462 records were read from the infile C:\CARB\QUEST.DAT.
      The minimum record length was 291.
      The maximum record length was 291.
NOTE: The data set WORK.QEXTRACT has 316 observations and 5 variables.
NOTE: The DATA statement used 12.00 seconds.
   72
   73    ********************************************************************;
   74    *  Merge chemical and questionnaire data;
   75    *  Since HOUR24.DAT and QUEST.DAT are sorted by ID1 and ID2, no
   76       sort is necessary prior to merge;
   77
   78    DATA CEXTRACT;
   79    MERGE CEXTRACT(IN=CCCC) QEXTRACT(IN=QQQQ);BY ID1 ID2;
   80    IF CCCC AND QQQQ;
   81    * RECODE COMPOUND IDS TO USE AS SUBGROUP VARIABLE IN SUDAAN;
   82    IF CMPD=4 THEN C=1;
   83    ELSE IF CMPD=5 THEN C=2;
   84    ELSE IF CMPD=11 THEN C=3;
   85    ELSE IF CMPD=14 THEN C=4;
   86    ELSE IF CMPD=16 THEN C=5;
   87    ELSE IF CMPD=21 THEN C=6;
   88    * RECODE STUDY IDS TO USE AS SUBGROUP VARIABLE IN SUDAAN;
   89    IF ID1=1 THEN STUDY=1;
   90    ELSE IF ID1=3 THEN STUDY=2;
   91    ELSE IF ID1=6 THEN STUDY=3;
   92    * REDEFINE STRATA TO DISTINGUISH COMPOUNDS AND STUDIES;
   93    IF ID1=1 THEN CSTRAT=ID4-2021200+1000*C+100*ID1;
```

Exhibit B-1. (continued)

```
   94    ELSE IF ID1=3 THEN CSTRAT=ID4-2023100+1000*C+100*ID1;
   95    ELSE IF ID1=6 THEN CSTRAT=ID4+1000*C+100*ID1;
   96    DROP CMPD ID1;
   97    RUN;
NOTE: The data set WORK.CEXTRACT has 1619 observations and 10 variables.
NOTE: The DATA statement used 14.00 seconds.
   98
   99    ********************************************************************;
  100    * Sort data set by stratum and PSU to use SUDAAN;
  101
  102    PROC SORT DATA=CEXTRACT OUT=OUT.BCHEMDAT;BY CSTRAT ID5 ID2;
  103    RUN;
NOTE: The data set OUT.BCHEMDAT has 1619 observations and 10 variables.
NOTE: The PROCEDURE SORT used 9.00 seconds.
  104    DATA OUT.BPCHEM1
  105         OUT.BPCHEM2
  106         OUT.BPCHEM3
  107         OUT.BPCHEM4
  108         OUT.BPCHEM5
  109         OUT.BPCHEM6;
  110    SET OUT.BCHEMDAT;
  111    IF C=1 THEN OUTPUT OUT.BPCHEM1;
  112    ELSE IF C=2 THEN OUTPUT OUT.BPCHEM2;
  113    ELSE IF C=3 THEN OUTPUT OUT.BPCHEM3;
  114    ELSE IF C=4 THEN OUTPUT OUT.BPCHEM4;
  115    ELSE IF C=5 THEN OUTPUT OUT.BPCHEM5;
  116    ELSE IF C=6 THEN OUTPUT OUT.BPCHEM6;
  117    RUN;
NOTE: The data set OUT.BPCHEM1 has 270 observations and 10 variables.
NOTE: The data set OUT.BPCHEM2 has 271 observations and 10 variables.
NOTE: The data set OUT.BPCHEM3 has 270 observations and 10 variables.
NOTE: The data set OUT.BPCHEM4 has 269 observations and 10 variables.
NOTE: The data set OUT.BPCHEM5 has 269 observations and 10 variables.
NOTE: The data set OUT.BPCHEM6 has 270 observations and 10 variables.
NOTE: The DATA statement used 15.00 seconds.
```

Exhibit B-1. (continued)

SUDAAN PROGRAM FOR DATA ANALYSIS

```
                    SUDAAN  Survey Data Analysis Software
          Copyright      Research Triangle Institute      April 1993
                                Release 6.32

1   PROC DESCRIPT DATA="C:\\CARB\\BCHEMDAT" DESIGN=WR FILETYPE=SAS
    MEANS GEOMETRIC NOMARG;
2   WEIGHT WPER;
3   NEST CSTRAT ID5;
4   VAR CPER;
5   SUBGROUP C STUDY SMOKE;
6   LEVELS 6 3 3;
7   TABLES C*STUDY C*STUDY*SMOKE;
8   PRINT / FILENAME="BM1CHEM.OUT" REPLACE;
9   OUTPUT NSUM WSUM MEAN GEOMEAN SEMEAN SEGEOMEAN
          / FILENAME="BM1CHEM" REPLACE;


Number of observations read    :   1619    Weighted count:  2889701
Denominator degrees of freedom :    282

10  PROC DESCRIPT DATA="C:\\CARB\\BCHEMDAT" DESIGN=WR FILETYPE=SAS
    MEANS GEOMETRIC NOMARG;
11  WEIGHT WPER;
12  NEST CSTRAT ID5;
13  VAR CPER;
14  SUBGROUP C STUDY;
15  LEVELS 6 3;
16  TABLES C;
17  PAIRWISE STUDY;
18  PRINT / FILENAME="BM2CHEM.OUT" REPLACE;
19  OUTPUT NSUM WSUM MEAN GEOMEAN SEMEAN SEGEOMEAN
          / FILENAME="BM2CHEM" REPLACE;

Number of observations read    :   1619    Weig' 'd count:  2889701
Denominator degrees of freedom :    282
```

Exhibit B-1. (continued)

```
20  PROC DESCRIPT DATA="C:\\CARB\\BCHEMDAT" DESIGN=WR FILETYPE=SAS
    MEANS GEOMETRIC NOMARG;
21  WEIGHT WPER;
22  NEST CSTRAT ID5;
23  VAR CPER;
24  SUBGROUP C STUDY SMOKE;
25  LEVELS 6 3 3;
26  TABLES C*SMOKE;
27  PAIRWISE STUDY;
28  PRINT / FILENAME="BM3CHEM.OUT" REPLACE;
29  OUTPUT NSUM WSUM MEAN GEOMEAN SEMEAN SEGEOMEAN
          / FILENAME="BM3CHEM" REPLACE;


Number of observations read    :   1619    Weighted count:  2889701
Denominator degrees of freedom :    282


30  PROC DESCRIPT DATA="C:\\CARB\\BPCHEM1" DESIGN=WR FILETYPE=SAS
    EST_NO=290 NOMARG;
31  PERCENTILES / MEDIAN UNGROUPED;
32  WEIGHT WPER;
33  NEST CSTRAT ID5;
34  VAR CPER;
35  SUBGROUP STUDY SMOKE;
36  LEVELS 3 3;
37  TABLES STUDY STUDY*SMOKE;
38  PRINT / FILENAME="BPCHEM1.OUT" REPLACE;
39  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="BPCHEM1" REPLACE;


Number of observations read    :    270    Weighted count:   481410
Denominator degrees of freedom :     47
```

Exhibit B-1. (continued)

```
40  PROC DESCRIPT DATA="C:\\CARB\\BPCHEM2" DESIGN=WR FILETYPE=SAS
    EST_NO=290 NOMARG;
41  PERCENTILES / MEDIAN UNGROUPED;
42  WEIGHT WPER;
43  NEST CSTRAT ID5;
44  VAR CPER;
45  SUBGROUP STUDY SMOKE;
46  LEVELS 3 3;
47  TABLES STUDY STUDY*SMOKE;
48  PRINT / FILENAME="BPCHEM2.OUT" REPLACE;
49  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="BPCHEM2" REPLACE;


Number of observations read    :    271    Weighted count:    481658
Denominator degrees of freedom :     47


50  PROC DESCRIPT DATA="C:\\CARB\\BPCHEM3" DESIGN=WR FILETYPE=SAS
    EST_NO=290 NOMARG;
51  PERCENTILES / MEDIAN UNGROUPED;
52  WEIGHT WPER;
53  NEST CSTRAT ID5;
54  VAR CPER;
55  SUBGROUP STUDY SMOKE;
56  LEVELS 3 3;
57  TABLES STUDY STUDY*SMOKE;
58  PRINT / FILENAME="BPCHEM3.OUT" REPLACE;
59  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="BPCHEM3" REPLACE;


Number of observations read    :    270    Weighted count:    481658
Denominator degrees of freedom :     47
```

Exhibit B-1. (continued)

```
60  PROC DESCRIPT DATA="C:\\CARB\\BPCHEM4" DESIGN=WR FILETYPE=SAS
    EST_NO=290 NOMARG;
61  PERCENTILES / MEDIAN UNGROUPED;
62  WEIGHT WPER;
63  NEST CSTRAT ID5;
64  VAR CPER;
65  SUBGROUP STUDY SMOKE;
66  LEVELS 3 3;
67  TABLES STUDY STUDY*SMOKE;
68  PRINT / FILENAME="BPCHEM4.OUT" REPLACE;
69  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="BPCHEM4" REPLACE;


Number of observations read    :    269    Weighted count:    481658
Denominator degrees of freedom :     47


70  PROC DESCRIPT DATA="C:\\CARB\\BPCHEM5" DESIGN=WR FILETYPE=SAS
    EST_NO=290 NOMARG;
71  PERCENTILES / MEDIAN UNGROUPED;
72  WEIGHT WPER;
73  NEST CSTRAT ID5;
74  VAR CPER;
75  SUBGROUP STUDY SMOKE;
76  LEVELS 3 3;
77  TABLES STUDY STUDY*SMOKE;
78  PRINT / FILENAME="BPCHEM5.OUT" REPLACE;
79  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="BPCHEM5" REPLACE;


Number of observations read    :    269    Weighted count:    481658
Denominator degrees of freedom :     47
```

Exhibit B-1. (continued)

```
80  PROC DESCRIPT DATA="C:\\CARB\\BPCHEM6" DESIGN=WR FILETYPE=SAS
    EST_NO=290 NOMARG;
81  PERCENTILES / MEDIAN UNGROUPED;
82  WEIGHT WPER;
83  NEST CSTRAT ID5;
84  VAR CPER;
85  SUBGROUP STUDY SMOKE;
86  LEVELS 3 3;
87  TABLES STUDY STUDY*SMOKE;
88  PRINT / FILENAME="BPCHEM6.OUT" REPLACE;
89  OUTPUT NSUM WSUM QTILE SEQTILE LOWQTILE UPQTILE
          / FILENAME="BPCHEM6" REPLACE;


Number of observations read    :    270    Weighted count:    481658
Denominator degrees of freedom :     47
```

SAS PROGRAM FOR POST-PROCESSING OF SUDAAN FILES

```
 420    * PGM STORED IN CARB\BANLPOST.SAS;
 421    OPTIONS LS=118 PS=46 NOCENTER NODATE NONUMBER MISSING=' ';
 422    PROC FORMAT;
 423    VALUE CMPDF
 424                 1='1,1,1-Trichloroethane'
 425                 2='Benzene'
 426                 3='Tetrachloroethylene'
 427                 4='Styrene'
 428                 5='m,p-Dichlorobenzene'
 429                 6='m,p-Xylene' ;
NOTE: Format CMPDF has been output.
 430    VALUE SMOKEF 1='ACTIVE' 2='PASSIVE'
 431                 3='UNEXPOSED';
NOTE: Format SMOKEF has been output.
 432    VALUE CONTF  1='ACTIVE-PASSIVE'
 433                 2='ACTIVE-UNEXPOSED'
 434                 3='PASSIVE-UNEXPOSED';
```

Exhibit B-1. (continued)

```
NOTE: Format CONTF has been output.
  435    VALUE TABLEF 1='1:  BY STUDY'
  436                 2='2:  BY STUDY AND SMOKING CATEGORY';
NOTE: Format TABLEF has been output.
  437    VALUE STUDYF 1='LA,WINTER84' 2='P/A,JUNE84' 3='WL,JUNE90';
NOTE: Format STUDYF has been output.
  438    VALUE CONTS  1='LA,WINTER84-P/A,JUNE84'
  439                 2='LA,WINTER84-WL,JUNE90'
  440                 3='P/A,JUNE84-WL,JUNE90';
NOTE: Format CONTS has been output.
  441    RUN;
NOTE: The PROCEDURE FORMAT used 5.00 seconds.
  442    DATA POPSTATS;
  443    INFILE 'C:\CARB\BM1CHEM.DBS';
  444    INPUT TABLE 2. +2 C 3. STUDY 3. SMOKE 3. +4 N 9. WSUM 11.
  445     CM 11. CGM 11. CMSE 11. CGMSE 11.;
  446    IF SMOKE<0 THEN SMOKE=.;
  447    RUN;
NOTE: The infile 'C:\CARB\BM1CHEM.DBS' is file C:\CARB\BM1CHEM.DBS.
NOTE: 72 records were read from the infile C:\CARB\BM1CHEM.DBS.
      The minimum record length was 81.
      The maximum record length was 81.
NOTE: The data set WORK.POPSTATS has 72 observations and 10 variables.
NOTE: The DATA statement used 2.00 seconds.
  448
  449    DATA POPMEDN;
  450    INFILE 'C:\CARB\BPCHEM1.DBS' EOF=CH2;C=1;GO TO INP;
  451    CH2: INFILE 'C:\CARB\BPCHEM2.DBS' EOF=CH3;C=2;GO TO INP;
  452    CH3: INFILE 'C:\CARB\BPCHEM3.DBS' EOF=CH4;C=3;GO TO INP;
  453    CH4: INFILE 'C:\CARB\BPCHEM4.DBS' EOF=CH5;C=4;GO TO INP;
  454    CH5: INFILE 'C:\CARB\BPCHEM5.DBS' EOF=CH6;C=5;GO TO INP;
  455    CH6: INFILE 'C:\CARB\BPCHEM6.DBS';C=6;
  456    INP: INPUT TABLE 2. +2 STUDY 3. SMOKE 3. +30 CMED 11.
  457    CMEDSE 11. LOWCMED 11. UPCMED 11.;
  458    IF SMOKE<0 THEN SMOKE=.;
  459    RUN;
NOTE: The infile 'C:\CARB\BPCHEM1.DBS' is file C:\CARB\BPCHEM1.DBS.
NOTE: 12 records were read from the infile C:\CARB\BPCHEM1.DBS.
```

Exhibit B-1. (continued)

```
  471     TITLE
  472     'POPULATION ESTIMATES:  MEANS, GEOMETRIC MEANS, AND MEDIANS';
  473     RUN;
NOTE: The PROCEDURE PRINT used 4.00 seconds.
  474
  475     DATA CONTRST1;
  476     INFILE 'C:\CARB\BM2CHEM.DBS';
  477     INPUT @5 C 3. +7 CONTRST 4. N 9. WSUM 11.
  478      CMDIF 11. CGMDIF 11. CMDIFSE 11. CGMDIFSE 11.;
  479     RUN;
NOTE: The infile 'C:\CARB\BM2CHEM.DBS' is file C:\CARB\BM2CHEM.DBS.
NOTE: 18 records were read from the infile C:\CARB\BM2CHEM.DBS.
      The minimum record length was 82.
      The maximum record length was 82.
NOTE: The data set WORK.CONTRST1 has 18 observations and 8 variables.
NOTE: The DATA statement used 2.00 seconds.
  480     PROC PRINT; ID C CONTRST;
  481     VAR N WSUM CMDIF CMDIFSE CGMDIF CGMDIFSE;
  482     FORMAT C CMPDF. CONTRST CONTS.
  483     N 4. WSUM 10. CMDIF CMDIFSE CGMDIF CGMDIFSE 7.2;
  484     TITLE
  485     'POPULATION ESTIMATES:  DIFFERENCES IN MEANS & GEOMETRIC MEANS';
  486     RUN;
NOTE: The PROCEDURE PRINT used 1.00 seconds.
  487
  488     DATA CONTRST2;
  489     INFILE 'C:\CARB\BM3CHEM.DBS';
  490     INPUT @5 C 3. +3 SMOKE 3. +4 CONTRST 4. N 9. WSUM 11.
  491      CMDIF 11. CGMDIF 11. CMDIFSE 11. CGMDIFSE 11.;
  492     RUN;
NOTE: The infile 'C:\CARB\BM3CHEM.DBS' is file C:\CARB\BM3CHEM.DBS.
NOTE: 54 records were read from the infile C:\CARB\BM3CHEM.DBS.
      The minimum record length was 85.
      The maximum record length was 85.
NOTE: The data set WORK.CONTRST2 has 54 observations and 9 variables.
NOTE: The DATA statement used 2.00 seconds.
```

Exhibit B-1. (continued)

```
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The infile 'C:\CARB\BPCHEM2.DBS' is file C:\CARB\BPCHEM2.DBS.
NOTE: 12 records were read from the infile C:\CARB\BPCHEM2.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The infile 'C:\CARB\BPCHEM3.DBS' is file C:\CARB\BPCHEM3.DBS.
NOTE: 12 records were read from the infile C:\CARB\BPCHEM3.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The infile 'C:\CARB\BPCHEM4.DBS' is file C:\CARB\BPCHEM4.DBS.
NOTE: 12 records were read from the infile C:\CARB\BPCHEM4.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The infile 'C:\CARB\BPCHEM5.DBS' is file C:\CARB\BPCHEM5.DBS.
NOTE: 12 records were read from the infile C:\CARB\BPCHEM5.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The infile 'C:\CARB\BPCHEM6.DBS' is file C:\CARB\BPCHEM6.DBS.
NOTE: 12 records were read from the infile C:\CARB\BPCHEM6.DBS.
      The minimum record length was 84.
      The maximum record length was 84.
NOTE: The data set WORK.POPMEDN has 72 observations and 8 variables.
NOTE: The DATA statement used 3.00 seconds.
  460    PROC SORT;BY TABLE C STUDY SMOKE;
  461    RUN;
NOTE: The data set WORK.POPMEDN has 72 observations and 8 variables.
NOTE: The PROCEDURE SORT used 1.00 seconds.
  462    DATA POPSTATS;MERGE POPSTATS POPMEDN;
  463    BY TABLE C STUDY SMOKE;
  464    IF SMOKE<0 THEN SMOKE=.;
  465    RUN;
NOTE: The data set WORK.POPSTATS has 72 observations and 14 variables.
NOTE: The DATA statement used 2.00 seconds.
  466
  467    PROC PRINT; BY TABLE;ID C STUDY SMOKE;
  468    VAR N WSUM CM CMSE CGM CGMSE CMED CMEDSE LOWCMED UPCMED;
  469    FORMAT C CMPDF. SMOKE SMOKEF. TABLE TABLEF. STUDY STUDYF.
  470    N 3. WSUM 8. CM CMSE CGM CGMSE CMED CMEDSE LOWCMED UPCMED 6.2;
```

Exhibit B-1. (continued)

```
PROC PRINT; ID C SMOKE CONTRST;
VAR N WSUM CMDIF CMDIFSE CGMDIF CGMDIFSE;
FORMAT C CMPDF. CONTRST CONTS. SMOKE SMOKEF.
N 4. WSUM 10. CMDIF CMDIFSE CGMDIF CGMDIFSE 7.2;
TITLE
'POPULATION ESTIMATES, BY SMOKING CATEGORY:  DIFFERENCES IN MEANS & GEOMETRIC MEANS';
RUN;
```

NOTE: The PROCEDURE PRINT used 3.00 seconds.

APPENDIX II-C

TYPE C ANALYSIS

APPENDIX II-C: TYPE C ANALYSIS

Exhibit C-1 provides the log of the SAS program used to create the working files and to perform the analysis. Two files, EXTRACT4 and EXTRACT5, were initially created, one for each study. Then these were merged by compound and participant ID.[10] The working file for the analysis (called EXTRACT) was then created that contained the following newly created variables:

TYPE = type of concentration variable (1 = personal, 2 = indoor, 3 = outdoor, and 4 = I/O ratio)

Y1 = Study 4 concentration or indoor/outdoor ratio

Y2 = Study 5 concentration or indoor/outdoor ratio

LY1 = ln(Y1)

LY2 = ln(Y2)

DIFF = Y1-Y2

RATIO = Y1/Y2

LRATIO = ln(RATIO) = ln(Y1) - ln(Y2)

The file was structured to contain a separate record for each combination of compound and TYPE. Both members of the pair were required to have nonmissing values for Y1 and Y2 in order to be included on the file. The analysis portion of the program consisted simply of a call to the SAS MEANS procedure, by compound and TYPE.

Outputs produced by the program in Exhibit C-1 can be found in Exhibit 6-1 of the main text.

---

[10] For the indoor and outdoor data, one could also choose to merge by household ID (i.e., the first six digits of ID1); this would produce somewhat different results, since that approach would require a match only for the house (not both the house and the participant).

Exhibit C-1. Analysis Type C: Program Log

SAS PROGRAM FOR PREPARING FILES AND CONDUCTING ANALYSIS

```
367     * PGM STORED IN CARB\CANLYSIS.SAS;
368     FILENAME IN 'C:\CARB\HOUR24.DAT';
369     OPTIONS LS=85 PS=60 NOCENTER NODATE NONUMBER;
370
371     PROC FORMAT;
372     /* FOLLOWING STATEMENTS WERE EXTRACTED FROM CHEM.FMT */
373     VALUE CMPDF   2='Chloroform'
374                   3='1,2-Dichloroethane'
375                   4='1,1,1-Trichloroethane'
376                   5='Benzene'
377                   6='Carbon Tetrachloride'
378                   7='Trichloroethylene'
379                   8='Bromodichloromethane'
380                   9='Dibromochloromethane'
381                  11='Tetrachloroethylene'
382                  12='Chlorobenzene'
383                  13='Bromoform'
384                  14='Styrene'
385                  15='m-Dichlorobenzene'
386                  16='p-Dichlorobenzene'
387                  17='p-Dichlorobenzene'
388                  18='o-Dichlorobenzene'
389                  19='Ethylbenzene'
390                  20='o-Xylene'
391                  21='m,p-Xylene'
392                  23='n-Decane'
393                  24='n-Dodecane'
394                  25='1,4-Dioxane'
395                  26='1,1,1,2-Tetrachloroethane'
396                  27='1,2-Dibromoethane'
397                  28='n-Octane'
398                  29='n-Undecane'
399                  30='1,1,2,2-Tetrachloroethane'
400                  31='a-Pinene'
401                  32='Limonene'
402                  33='n-Nonane';
NOTE: Format CMPDF has been output.
403
404       VALUE TYPEF 1='PERSONAL'
405                   2='INDOOR'
406                   3='OUTDOOR'
407                   4='I/O RATIO';
NOTE: Format TYPEF has been output.
408       RUN;
NOTE: The PROCEDURE FORMAT used 3.00 seconds.
409
```

Exhibit C-1. (continued)

```
410     * Extract pertinent 24-hr chemical data and create data sets for
411       Study 4 and for Study 5;
412
413     DATA EXTRACT4(RENAME=(CKIT=CIND1 CPER=CPER1 COUT=COUT1 IO=IO1))
414          EXTRACT5(RENAME=(CKIT=CIND2 CPER=CPER2 COUT=COUT2 IO=IO2));
415     INFILE IN;
416     INPUT
417
418     /* FOLLOWING STATEMENTS WERE EXTRACTED FROM HOUR24.CBK */
419
420     @1    ID1    1.    /*STUDY NUMBER*/
421     @2    ID2    7.    /*HOUSEHOLD/PARTICIPANT ID*/
422     @9    CMPD   2.    /*COMPOUND ID*/
423     @11   CPER   9.2   /*CONCENTRATION FOR PERSONAL AIR*/
424     @20   CKIT   9.2   /*CONCENTRATION FOR KITCHEN*/
425     @38   COUT   9.2   /*CONCENTRATION FOR OUTDOORS*/
426      ;
427     IF ID1=4 OR ID1=5;
428       /* SELECT RECORDS FOR STUDIES 4 AND 5 */
429     IF CMPD=4 OR CMPD=5 OR CMPD=11 OR CMPD=14 OR CMPD=21 OR
430          CMPD=16 OR CMPD=17;
431     IF CMPD=17 THEN CMPD=16;
432         /* SELECT COMPOUNDS AND CONVERT CMPD 17 TO CMPD 16 */
433     IF MAX(CPER,CKIT,COUT)=. THEN DELETE;
434     IO=CKIT/COUT;
435        /* CREATE INDOOR/OUTDOOR RATIO */
436     IF ID1=4 THEN OUTPUT EXTRACT4;
437     ELSE IF ID1=5 THEN OUTPUT EXTRACT5;
438     KEEP ID2 CMPD CKIT CPER COUT IO;
439     RUN;
NOTE: The infile IN is file C:\CARB\HOUR24.DAT.
NOTE: 8797 records were read from the infile C:\CARB\HOUR24.DAT.
      The minimum record length was 80.
      The maximum record length was 80.
NOTE: Missing values were generated as a result of performing an operation on
      missing values.
      Each place is given by: (Number of times) at (Line):(Column).
      182 at 434:4
NOTE: The data set WORK.EXTRACT4 has 287 observations and 6 variables.
NOTE: The data set WORK.EXTRACT5 has 249 observations and 6 variables.
NOTE: The DATA statement used 33.00 seconds.
440
441     * Sort data sets to allow merge by CMPD, ID2;
442
443     PROC SORT DATA=EXTRACT4;BY CMPD ID2;
444     RUN;
NOTE: The data set WORK.EXTRACT4 has 287 observations and 6 variables.
NOTE: The PROCEDURE SORT used 2.00 seconds.
445     PROC SORT DATA=EXTRACT5;BY CMPD ID2;
446     RUN;
NOTE: The data set WORK.EXTRACT5 has 249 observations and 6 variables.
NOTE: The PROCEDURE SORT used 4.00 seconds.
447
```

Exhibit C-1. (continued)

```
448     * Merge data sets and create ratios and differences;
449
450     DATA EXTRACT;MERGE EXTRACT4(IN=D1) EXTRACT5(IN=D2);BY CMPD ID2;
451     IF D1 AND D2;
452     IF CPER1 NE . AND CPER2 NE . THEN DO;
453     Y1=CPER1;Y2=CPER2;
454     LY1=LOG(Y1);LY2=LOG(Y2);
455     TYPE=1;DIFF=CPER1-CPER2;RATIO=CPER1/CPER2;LRATIO=LOG(RATIO);OUTPUT;
456     END;
457     IF CIND1 NE . AND CIND2 NE . THEN DO;
458     Y1=CIND1;Y2=CIND2;
459     LY1=LOG(Y1);LY2=LOG(Y2);
460     TYPE=2;DIFF=CIND1-CIND2;RATIO=CIND1/CIND2;LRATIO=LOG(RATIO);OUTPUT;
461     END;
462     IF COUT1 NE . AND COUT2 NE . THEN DO;
463     Y1=COUT1;Y2=COUT2;
464     LY1=LOG(Y1);LY2=LOG(Y2);
465     TYPE=3;DIFF=COUT1-COUT2;RATIO=COUT1/COUT2;LRATIO=LOG(RATIO);OUTPUT;
466     IF CIND1 NE . AND CIND2 NE . THEN DO;
467     Y1=IO1;Y2=IO2;
468     LY1=LOG(Y1);LY2=LOG(Y2);
469     TYPE=4;DIFF=IO1-IO2;RATIO=IO1/IO2;LRATIO=LOG(RATIO);OUTPUT;
470     END;
471     END;
472     KEEP ID2 CMPD TYPE Y1 Y2 LY1 LY2 DIFF RATIO LRATIO;
473     RUN;
NOTE: The data set WORK.EXTRACT has 530 observations and 10 variables.
NOTE: The DATA statement used 8.00 seconds.
474
475     PROC SORT DATA=EXTRACT;BY CMPD TYPE ID2;
476     RUN;
NOTE: The data set WORK.EXTRACT has 530 observations and 10 variables.
NOTE: The PROCEDURE SORT used 4.00 seconds.
477     /*  THE FOLLOWING STATEMENTS WILL PROVIDE A LISTING OF THE DATA
478     PROC PRINT;BY CMPD TYPE;ID ID2;
479     VAR Y1 Y2 LY1 LY2 DIFF RATIO LRATIO;
480     FORMAT CMPD CMPDF. TYPE TYPEF.
481     Y1 Y2 DIFF RATIO 8.2 LY1 LY2 LRATIO 8.4;
482     TITLE 'STUDY 4 VS. STUDY 5 COMPARISON';
483     RUN;
484     */
485     PROC MEANS MIN MAX MEAN STD STDERR T PRT FW=7 MAXDEC=3;BY CMPD TYPE;
486     VAR Y1 Y2 LY1 LY2 DIFF RATIO LRATIO;
487     FORMAT CMPD CMPDF. TYPE TYPEF.;
488     RUN;
NOTE: The PROCEDURE MEANS used 14.00 seconds.
```

APPENDIX II-D

TYPE D ANALYSIS

APPENDIX II-D: TYPE D ANALYSIS

Exhibit D-1 contains the program log. The first step in preparing the files for analysis involved selecting the data records associated with the compounds of interest (2661 out of the 8797 records on the 24-h concentration data file). Concentration data for compounds 16 and 17 were then combined. A single indoor-air concentration variate was created (i.e., the location within the house [kitchen or main living area] was ignored). The file was then restructured using SAS programming statements to produce a working file called FORCORR. This file contained 447 records -- one data record per participant (for those participants having some concentration data on one or more of the selected compounds).

The SAS procedure CORR was then executed three times -- once for the personal air data, once for the indoor air data, and once for the outdoor air data.

A new working file, called LFORCORR, was then created in which the original concentration data values were replaced with their logarithms. This was straightforward for all studies except Study 6, since no zero concentration values had been retained except possibly for Study 6. For Study 6, the following rule was adopted to deal with this: If for a given compound, there were some nonquantifiable values other than zeros, then zeros were to be replaced by a value equal to 1/2 of the smallest such value. If all of the nonquantifiable values were zeros (among the six compounds, this occurred only for m,p-Xylene), then these zeros would be replaced by 1/2 of the median QL value for Study 6 (see Sheldon et al., 1991, Tables 9-3, 9-4, and 9-5).

The SAS procedure CORR was then executed three additional times -- once for the logs of the personal air data, once for the logs of the indoor air data, and once for the logs of the outdoor air data.

Outputs produced by the program in Exhibit D-1 can be found in Exhibit 7-1 of the main text.

Exhibit D-1. Analysis Type D: Program Log

SAS PROGRAM FOR PREPARING FILES AND CONDUCTING ANALYSIS

```
 1    * PGM STORED IN CARB/DANLYSIS.SAS;
 2    FILENAME IN 'C:\CARB\HOUR24.DAT';
 3    OPTIONS LS=95 PS=56 NOCENTER NODATE NONUMBER;
 4
 5    * Extract pertinent 24-hr chemical data;
 6
 7    DATA EXTRACT;
 8    INFILE IN;
 9    INPUT
10
11    /* FOLLOWING STATEMENTS WERE EXTRACTED FROM HOUR24.CBK */
12
13    @1    ID1    1.    /*STUDY NUMBER*/
14    @2    ID2    7.    /*HOUSEHOLD/PARTICIPANT ID*/
15    @9    CMPD   2.    /*COMPOUND ID*/
16    @11   CPER   9.2   /*CONCENTRATION FOR PERSONAL AIR*/
17    @20   CKIT   9.2   /*CONCENTRATION FOR KITCHEN*/
18    @29   CMLA   9.2   /*CONCENTRATION FOR MAIN LIVING AREA*/
19    @38   COUT   9.2   /*CONCENTRATION FOR OUTDOORS*/
20     ;
21    IF CMPD=4 OR CMPD=5 OR CMPD=11 OR CMPD=14 OR CMPD=21 OR
22         CMPD=16 OR CMPD=17;
23    IF CMPD=17 THEN CMPD=16;
24        /* SELECT COMPOUNDS AND CONVERT CMPD 17 TO CMPD 16 */
25    IF CMLA NE . THEN CIND=CMLA;
26    ELSE IF CKIT NE . THEN CIND=CKIT;
27       /* CREATE SINGLE INDOOR CONCENTRATION VARIABLE */
28    IF MAX(CPER,CIND,COUT)=. THEN DELETE;
29    DROP CKIT CMLA;
30    RUN;
NOTE: The infile IN is file C:\CARB\HOUR24.DAT.
NOTE: 8797 records were read from the infile C:\CARB\HOUR24.DAT.
      The minimum record length was 80.
      The maximum record length was 80.
NOTE: The data set WORK.EXTRACT has 2661 observations and 6 variables.
NOTE: The DATA statement used 49.00 seconds.
31
32    * Restructure file to place data for all selected compounds into
33      one data record;
34
35    DATA FORCORR;SET EXTRACT;BY ID1 ID2;
36    RETAIN CP4 CP5 CP11 CP14 CP16 CP21
37           CI4 CI5 CI11 CI14 CI16 CI21
38           CO4 CO5 CO11 CO14 CO16 CO21;
39    ARRAY CP{6} CP4 CP5 CP11 CP14 CP16 CP21;
40    ARRAY CI{6} CI4 CI5 CI11 CI14 CI16 CI21;
```

xhibit D-1. (continued)

```
41    ARRAY CO{6} CO4 CO5 CO11 CO14 CO16 CO21;
42    IF FIRST.ID2 THEN DO I=1 TO 6;
43       CP{I}=.;CI{I}=.;CO{I}=.;
44    END;
45    IF CMPD=4 THEN I=1;
46    ELSE IF CMPD=5 THEN I=2;
47    ELSE IF CMPD=11 THEN I=3;
48    ELSE IF CMPD=14 THEN I=4;
49    ELSE IF CMPD=16 THEN I=5;
50    ELSE IF CMPD=21 THEN I=6;
51    CP{I}=CPER;CI{I}=CIND;CO{I}=COUT;
52    IF LAST.ID2 THEN OUTPUT;
53    DROP CMPD I CPER CIND COUT ID2;
54    LABEL        CP4='1,1,1-Trichloroethane'
55                 CP5='Benzene'
56                CP11='Tetrachloroethylene'
57                CP14='Styrene'
58                CP16='m,p- or p-Dichlorobenzene'
59                CP21='m,p-Xylene'
60                 CO4='1,1,1-Trichloroethane'
61                 CO5='Benzene'
62                CO11='Tetrachloroethylene'
63                CO14='Styrene'
64                CO16='m,p- or p-Dichlorobenzene'
65                CO21='m,p-Xylene'
66                 CI4='1,1,1-Trichloroethane'
67                 CI5='Benzene'
68                CI11='Tetrachloroethylene'
69                CI14='Styrene'
70                CI16='m,p- or p-Dichlorobenzene'
71                CI21='m,p-Xylene'
72                ;
73    RUN;
NOTE: The data set WORK.FORCORR has 447 observations and 19 variables.
NOTE: The DATA statement used 13.00 seconds.
74
75    * Perform correlation analyses;
76
77    PROC CORR;
78    VAR CP4 CP5 CP11 CP14 CP16 CP21;
79    TITLE '24-HR PERSONAL AIR CONCENTRATION DATA';
80    RUN;
NOTE: The PROCEDURE CORR used 3.00 seconds.
81    PROC CORR;
82    VAR CI4 CI5 CI11 CI14 CI16 CI21;
83    TITLE '24-HR INDOOR AIR CONCENTRATION DATA';
84    RUN;
NOTE: The PROCEDURE CORR used 2.00 seconds.
85    PROC CORR;
86    VAR CO4 CO5 CO11 CO14 CO16 CO21;
87    TITLE '24-HR OUTDOOR AIR CONCENTRATION DATA';
88    RUN;
```

Exhibit D-1. (continued)

```
NOTE: The PROCEDURE CORR used 2.00 seconds.
   89
   90     * Convert concentrations to log concentrations;
   91
   92     DATA LFORCORR;
   93     SET FORCORR;
   94     ARRAY X{18} CP4 CP5 CP11 CP14 CP16 CP21
   95                 CI4 CI5 CI11 CI14 CI16 CI21
   96                 CO4 CO5 CO11 CO14 CO16 CO21;
   97     DO I=1 TO 18;
   98     IF X{I} NE . THEN DO;
   99     IF 1<=ID1<=5 OR (ID1=6 AND X{I} NE 0) THEN X{I}=LOG(X{I});
  100     * FOR STUDIES 1-5, NONQUANTIFIABLE VALUES HAVE BEEN PREVIOUSLY
ASSIGNED.
  101       FOR STUDY 6, WE ADOPT THE FOLLOWING RULE FOR EACH COMPOUND (PERSONAL,
  102       OUTDOOR AND INDOOR DATA ARE COMBINED FOR THIS PURPOSE):
  103       IF THERE ARE SOME NONQUANTIFIABLE VALUES THAT ARE POSITIVE AND
  104       SOME THAT ARE ZERO, THEN SET ZERO VALUES TO 0.5*SMALLEST POSITIVE
  105       VALUE.  IF ALL NONQUANTIFIABLE VALUES ARE ZERO, THEN SET THEM TO
  106       0.5*(MEDIAN QL VALUE).  THIS RESULTS IN THE FOLLOWING;
  107     ELSE DO;
  108     J=MOD(I,6);IF J=1 THEN X{I}=LOG(0.06);
  109     ELSE IF J=2 THEN X{I}=LOG(0.035);
  110     ELSE IF 3<=J<=5 THEN X{I}=LOG(0.005);
  111     ELSE IF J=0 THEN X{I}=LOG(0.175);
  112     END;
  113     END;
  114     END;
  115     DROP I J;
  116     RUN;
NOTE: The data set WORK.LFORCORR has 447 observations and 19 variables.
NOTE: The DATA statement used 14.00 seconds.
  117
  118     * Perform correlation analyses on log-scale data;
  119
  120     PROC CORR;
  121     VAR CP4 CP5 CP11 CP14 CP16 CP21;
  122     TITLE 'LN(24-HR PERSONAL AIR CONCENTRATION DATA)';
  123     RUN;
NOTE: The PROCEDURE CORR used 3.00 seconds.
  124     PROC CORR;
  125     VAR CI4 CI5 CI11 CI14 CI16 CI21;
  126     TITLE 'LN(24-HR INDOOR AIR CONCENTRATION DATA)';
  127     RUN;
NOTE: The PROCEDURE CORR used 5.00 seconds.
  128     PROC CORR;
  129     VAR CO4 CO5 CO11 CO14 CO16 CO21;
  130     TITLE 'LN(24-HR OUTDOOR AIR CONCENTRATION DATA)';
  131     RUN;
NOTE: The PROCEDURE CORR used 2.00 seconds.
```

APPENDIX II-E

TYPE E ANALYSIS

APPENDIX II-E: TYPE E ANALYSIS

Exhibit E-1 provides the program logs for

- file preparation
- preliminary regressions (for each of six cases described in Section 8.1)
- stepwise regressions (selection based on prior results)
- final regressions (based on the stepwise results).

File preparation entailed: extracting the pertinent chemical data (1699 out of 8797 data records); constructing a single indoor air concentration variate (i.e., ignoring whether measurements were made in the kitchen or the main living area); extracting the relevant questionnaire data (222 observations) and recoding of the variate values to produce the X variables shown in Table 8-1; merging of the chemical and questionnaire data; revising zero concentrations reported for Study 6 in the manner described in Appendix II-D; creating the variates Y1, Y2, Y3, Y4, COUT, and LCOUT (defined in Table 8-1); and sorting the working file (CEXTRACT) by compound and study number (ID1).

The preliminary regressions involved fitting regression models, by compound, using the SAS PROC REG procedure for each of the six cases described in Section 8.1. Case 4 (i.e., dependent variable Y3 with LCOUT as a covariate) was selected as the case to carry forward with the stepwise regressions.

The stepwise regressions were also performed using PROC REG. Significance levels for both entry and retention of candidate regressors of the stepwise models were set at 0.10. Exhibit E-2 shows the complete output of the stepwise procedure for one of the compounds and Exhibit E-3 shows the final portion of the stepwise output for the remaining compounds. This portion of the output indicates the final model selected and indicates the order of variable selections and deletions.

The final results, which are given in Exhibit 8-1 of the main text, were obtained by refitting of the models selected by the stepwise procedure. This was done to utilize the maximum number of observations for each compound's model, since the stepwise procedure requires data values on all candidate regressors while these refit models require data values only for those regressors actually in the chosen model.

Exhibit E-1. Analysis Type E: Program Logs

SAS PROGRAM FOR PREPARING FILES AND CONDUCTING ANALYSIS

```
375     * PGM STORED IN CARB/EANLYSIS.SAS;
376     FILENAME INHR24 'C:\CARB\HOUR24.DAT';
377     FILENAME INQUEST 'C:\CARB\QUEST.DAT';
378     LIBNAME OUT 'C:\CARB';
379     OPTIONS LS=80 PS=60 NOCENTER NODATE NONUMBER MISSING=' ';
380     ******************************************************************;
381     /* FOLLOWING FORMATS WERE EXTRACTED FROM CHEM.FMT AND QUEST.FMT */
382     PROC FORMAT;
383     VALUE CMPDF   2='Chloroform'
384                   3='1,2-Dichloroethane'
385                   4='1,1,1-Trichloroethane'
386                   5='Benzene'
387                   6='Carbon Tetrachloride'
388                   7='Trichloroethylene'
389                   8='Bromodichloromethane'
390                   9='Dibromochloromethane'
391                  11='Tetrachloroethylene'
392                  12='Chlorobenzene'
393                  13='Bromoform'
394                  14='Styrene'
395                  15='m-Dichlorobenzene'
396                  16='m,p-Dichlorobenzene'
397                  17='p-Dichlorobenzene'
398                  18='o-Dichlorobenzene'
399                  19='Ethylbenzene'
400                  20='o-Xylene'
401                  21='m,p-Xylene'
402                  23='n-Decane'
403                  24='n-Dodecane'
404                  25='1,4-Dioxane'
405                  26='1,1,1,2-Tetrachloroethane'
406                  27='1,2-Dibromoethane'
407                  28='n-Octane'
408                  29='n-Undecane'
409                  30='1,1,2,2-Tetrachloroethane'
410                  31='a-Pinene'
411                  32='Limonene'
412                  33='n-Nonane';
NOTE: Format CMPDF has been output.
413
414       VALUE QUANTF  1='QUANTIFIABLE'
415                     2='BELOW THE QL';
NOTE: Format QUANTF has been output.
416       VALUE YESNOF 1='YES' 2='NO';
NOTE: Format YESNOF has been output.
417       VALUE HG4F  1='Steam' 2='Prof. Dry Clean' 3='Other' 4='None';
NOTE: Format HG4F has been output.
418       ******************************************************************;
419
```

Exhibit E-1. (continued)

```
 420    * Extract pertinent 24-hour chemical data;
 421    DATA CEXTRACT;
NOTE: The PROCEDURE FORMAT used 3.00 seconds.
 422    INFILE INHR24;
 423    INPUT
 424
 425    /* FOLLOWING STATEMENTS WERE EXTRACTED FROM HOUR24.CBK */
 426
 427    @1    ID1    1.    /*STUDY NUMBER*/
 428    @2    ID2    7.    /*HOUSEHOLD/PARTICIPANT ID*/
 429    @9    CMPD   2.    /*COMPOUND ID*/
 430    @20   CKIT   9.2   /*CONCENTRATION FOR KITCHEN*/
 431    @29   CMLA   9.2   /*CONCENTRATION FOR MAIN LIVING AREA*/
 432    @38   COUT   9.2   /*CONCENTRATION FOR OUTDOORS*/
 433    @48   MKIT   1.    /*MEASURABILITY INDICATOR FOR KITCHEN*/
 434    @49   MMLA   1.    /*MEASURABILITY INDICATOR FOR MAIN LIVING AREA*/
 435    @50   MOUT   1.    /*MEASURABILITY INDICATOR FOR OUTDOORS*/
 436     ;
 437
 438    IF ID1=4 OR ID1=5 OR ID1=6; /* SELECT RECORDS FOR STUDIES 4,5,6 */
 439    IF MKIT >=1 OR MMLA >=1; /* SELECT RECORDS WITH INDOOR AIR DATA */
 440    IF CMPD=4 OR CMPD=5 OR CMPD=11 OR CMPD=14 OR CMPD=21
 441       OR CMPD=17 OR CMPD=7 OR CMPD=12 OR CMPD=20;
 442       /* SELECT COMPOUNDS */
 443    IF MKIT>=1 THEN DO;CIND=CKIT;MIND=MKIT;END;
 444    ELSE IF MMLA>=1 THEN DO;CIND=CMLA;MIND=MMLA;END;
 445       /* CREATE SINGLE INDOOR CONCENTRATION VARIABLE */
 446    DROP CKIT CMLA MKIT MMLA;
 447    RUN;
NOTE: The infile INHR24 is file C:\CARB\HOUR24.DAT.
NOTE: 8797 records were read from the infile C:\CARB\HOUR24.DAT.
      The minimum record length was 80.
      The maximum record length was 80.
NOTE: The data set WORK.CEXTRACT has 1699 observations and 7 variables.
NOTE: The DATA statement used 47.00 seconds.
 448
 449    ********************************************************************;
 450    * Extract pertinent questionnaire data;
 451    DATA QEXTRACT;
 452    INFILE INQUEST LRECL=291;
 453    INPUT
 454
 455    /* FOLLOWING STATEMENTS WERE EXTRACTED FROM QUEST.CBK */
 456
 457    @1   ID1   1.     /*Study ID */
 458    @2   ID2   7.     /*Household/Participant ID */
 459    @59  HG4   1.     /*Carpet Cleaning in Past Year (Steam, Prof. Dry
Clean, Other, None) */
 460    @60  HG5   1.     /*Comm. Clean (Drapes/Carpet/Furn.) */
 461    @61  HG6   1.     /*Attached Garage */
 462    @68  HX7   1.     /*Central Air Conditioning */
 463    @69  HX8   1.     /*Window Air Conditioning */
 464    @70  HX9   1.     /*Circulation Fans */
 465    @71  HX10  1.     /*Exhaust Fans */
```

Exhibit E-1. (continued)

```
@72  HX11  1.     /*Gas Furnace */
@73  HX12  1.     /*Fireplace */
@74  HX13  1.     /*Window Fan */
@76  HX15  1.     /*Doors or Windows Open */
@77  HX16  1.     /*Gas or Kerosene Space Heater */
@78  HX17  1.     /*Woodstove */
@80  HG19  1.     /*Gas Water Heater */
@83  HX22  1.     /*Gas Stove */
@87  HG23  1.     /*Paints/Varnishes/Thinners/Removers */
@88  HG24  1.     /*Gasoline/Petroleum Products */
@89  HG25  1.     /*Auto, Motorcycle, or Gas Lawnmower */
@90  HG26  1.     /*-Cides or Garden/Lawn Chemicals */
@91  HG27  1.     /*Cleaning Supplies */
@121 HG31  1.     /*Painting */
@122 HG32  1.     /*Furniture Refinishing */
@123 HG33  1.     /*Model Building */
@139 HG40  1.     /*Tobacco Prod. Smoked in House */
@173 HG41  1.     /*Pesticides in Home */
@174 HG42  1.     /*Drycleaned Clothes in Home */
@176 HX44  1.     /*Shower/Bath */
@180 PX49  1.     /*Pumped Gas */
@187 PX52  1.     /*Paints/Solvents */
@189 PX54  1.     /*Mothballs/Crystals, Fresh/Deodorizers */
@190 PX55  1.     /*Odorous, vaporizing glues/adhesives */
@193 PX58  1.     /*Cleaning Solutions */
@202 PX61  1.     /*Aerosol Personal Care Products */
;
IF ID1=4 OR ID1=5 OR ID1=6; /* SELECT RECORDS FOR STUDIES 4,5,6 */
* CONSTRUCT OTHER DUMMY VARIABLES FOR REGRESSIONS;
LOC1=(ID1=4);
LOC2=(ID1=6);
IF ID1=6 THEN HG5R=(HG4=1 OR HG4=2);ELSE HG5R=2-HG5;
IF HX7=. AND HX8=. THEN HX7R=.;ELSE HX7R=(HX7=1 OR HX8=1);
IF HX12=. AND HX17=. THEN HX12R=.;ELSE HX12R=(HX12=1 OR HX17=1);
IF HG31=. AND HG32=. AND HG33=. THEN HG31R=.;
  ELSE HG31R=(HG31=1 OR HG32=1 OR HG33=1);
HG6R=2-HG6;
HX9R=2-HX9;
HX10R=2-HX10;
HX11R=2-HX11;
HX15R=2-HX15;
HX16R=2-HX16;
HG19R=2-HG19;
HX22R=2-HX22;
HG23R=2-HG23;
HG24R=2-HG24;
HG25R=2-HG25;
HG26R=2-HG26;
HG27R=2-HG27;
HG40R=2-HG40;
HG41R=2-HG41;
HG42R=2-HG42;
```

Exhibit E-1. (continued)

```
  518    HX44R=2-HX44;
  519    PX61R=2-PX61;
  520    RUN;
NOTE: The infile INQUEST is file C:\CARB\QUEST.DAT.
NOTE: 462 records were read from the infile C:\CARB\QUEST.DAT.
      The minimum record length was 291.
      The maximum record length was 291.
NOTE: Missing values were generated as a result of performing an operation on
      missing values.
      Each place is given by: (Number of times) at (Line):(Column).
      1 at 502:6
      10 at 503:6
      10 at 504:7
      10 at 506:7
      5 at 511:7
      5 at 513:7
      7 at 515:7
      2 at 512:7
      2 at 505:7
      2 at 507:7
      2 at 509:7
      2 at 517:7
      1 at 518:7
      1 at 519:7
      3 at 508:7
NOTE: The data set WORK.QEXTRACT has 222 observations and 59 variables.
NOTE: The DATA statement used 19.00 seconds.
  521    ***********************************************************************;
  522    *  Merge chemical and questionnaire data;
  523    *  Since HOUR24.DAT and QUEST.DAT are sorted by ID1 and ID2, no
  524       sort is necessary prior to merge;
  525
  526    DATA CEXTRACT;
  527    MERGE CEXTRACT(IN=CCCC) QEXTRACT(IN=QQQQ);BY ID1 ID2;
  528    IF CCCC AND QQQQ;
  529    * FOR STUDIES 4 AND 5, NONQUANTIFIABLE VALUES HAVE BEEN PREVIOUSLY
  530      ASSIGNED.  FOR STUDY 6, WE ADOPT THE FOLLOWING RULE FOR EACH
  531      COMPOUND (OUTDOOR AND INDOOR DATA ARE COMBINED FOR THIS PURPOSE):
  532      IF THERE ARE SOME NONQUANTIFIABLE VALUES THAT ARE POSITIVE AND
  533      SOME THAT ARE ZERO, THEN SET ZERO VALUES TO 0.5*SMALLEST POSITIVE
  534      VALUE.  IF ALL NONQUANTIFIABLE VALUES ARE ZERO, THEN SET THEM TO
  535      0.5*(MEDIAN QL VALUE).  THIS RESULTS IN THE FOLLOWING;
  536    IF ID1=6 AND (CIND=0 OR COUT=0) THEN DO;
  537       IF CMPD=4 THEN CONST=0.06;
  538       ELSE IF CMPD=5 THEN CONST=0.035;
  539       ELSE IF CMPD=7 OR CMPD=11 OR CMPD=12 OR CMPD=14 OR CMPD=17 THEN
  540             CONST=0.005;
  541       ELSE IF CMPD=20 THEN CONST=0.055;
  542       ELSE IF CMPD=21 THEN CONST=0.175;
  543       IF COUT=0 THEN COUT=CONST;
  544       IF CIND=0 THEN CIND=CONST;
  545    END;
```

Exhibit E-1. (continued)

```
  546     Y1=CIND;
  547     Y2=CIND/COUT;
  548     LCOUT=LOG(COUT);
  549     Y3=LOG(Y1);
  550     Y4=LOG(Y2);
  551     KEEP Y1-Y4 CMPD COUT LCOUT ID1
  552     LOC1 LOC2 HG5R HX7R HG6R HX9R HX10R HX11R HX12R HX15R HX16R HG19R
  553     HX22R HG23R HG24R HG25R HG26R HG27R HG31R HG40R HG41R HG42R HX44R
PX61R;
  554     RUN;
NOTE: Missing values were generated as a result of performing an operation on
      missing values.
      Each place is given by: (Number of times) at (Line):(Column).
      768 at 547:4
      768 at 548:11
      768 at 550:8
NOTE: The data set WORK.CEXTRACT has 1699 observations and 32 variables.
NOTE: The DATA statement used 30.00 seconds.
  555
  556     ********************************************************************;
  557     * SORT BY COMPOUND;
  558     PROC SORT;BY CMPD ID1;
  559     RUN;
NOTE: The data set WORK.CEXTRACT has 1699 observations and 32 variables.
NOTE: The PROCEDURE SORT used 56.00 seconds.
  560
  561     ********************************************************************;
  562     * PRELIMINARY REGRESSIONS TO DECIDE ON DEPENDENT VARIABLE AND TO
  563       IDENFIFY PROBLEMS WITH INDEPENDENT VARIABLES;
  564     PROC REG SIMPLE;BY CMPD;
  565     M1: MODEL Y1 Y3 =
  566     LOC1 LOC2 HG5R HX7R HG6R HX9R HX10R HX11R HX12R HX15R HX16R HG19R
  567     HX22R HG23R HG24R HG25R HG26R HG27R HG31R HG40R HG41R HG42R HX44R
PX61R;
  568     FORMAT CMPD CMPDF.;
  569     TITLE 'PRELIMINARY REGRESSION RESULTS';
  570     RUN;
NOTE: 188 observations read.
      21 observations have missing values.
      167 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=1,1,1-Trichloroethane
NOTE: 192 observations read.
      23 observations have missing values.
      169 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Benzene
NOTE: 197 observations read.
      23 observations have missing values.
      174 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Trichloroethylene
```

Exhibit E-1. (continued)

```
NOTE: 190 observations read.
      23 observations have missing values.
      167 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Tetrachloroethylene
NOTE: 157 observations read.
      23 observations have missing values.
      134 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Chlorobenzene
NOTE: 192 observations read.
      23 observations have missing values.
      169 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Styrene
NOTE: 192 observations read.
      23 observations have missing values.
      169 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=p-Dichlorobenzene
NOTE: 195 observations read.
      23 observations have missing values.
      172 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=o-Xylene
NOTE: 196 observations read.
      23 observations have missing values.
      173 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=m,p-Xylene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 34.00 seconds.
  571     PROC REG SIMPLE;BY CMPD;
  572     M2: MODEL Y1 = COUT
  573     LOC1 LOC2 HG5R HX7R HG6R HX9R HX10R HX11R HX12R HX15R HX16R HG19R
  574     HX22R HG23R HG24R HG25R HG26R HG27R HG31R HG40R HG41R HG42R HX44R
PX61R;
  575     M3: MODEL Y3 = LCOUT
  576     LOC1 LOC2 HG5R HX7R HG6R HX9R HX10R HX11R HX12R HX15R HX16R HG19R
  577     HX22R HG23R HG24R HG25R HG26R HG27R HG31R HG40R HG41R HG42R HX44R
PX61R;
  578     M4: MODEL Y2 Y4 =
  579     LOC1 LOC2 HG5R HX7R HG6R HX9R HX10R HX11R HX12R HX15R HX16R HG19R
  580     HX22R HG23R HG24R HG25R HG26R HG27R HG31R HG40R HG41R HG42R HX44R
PX61R;
  581     FORMAT CMPD CMPDF.;
  582     TITLE 'PRELIMINARY REGRESSION RESULTS';
  583     RUN;
NOTE: 188 observations read.
      84 observations have missing values.
      104 observations used in computations.
NOTE: The above message was for the following by-group:
```

Exhibit E-1. (continued)

```
      CMPD=1,1,1-Trichloroethane
NOTE: 192 observations read.
      93 observations have missing values.
      99 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Benzene
NOTE: 197 observations read.
      96 observations have missing values.
      101 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Trichloroethylene
NOTE: 190 observations read.
      91 observations have missing values.
      99 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Tetrachloroethylene
NOTE: 157 observations read.
      87 observations have missing values.
      70 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Chlorobenzene
NOTE: 192 observations read.
      90 observations have missing values.
      102 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Styrene
NOTE: 192 observations read.
      93 observations have missing values.
      99 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=p-Dichlorobenzene
NOTE: 195 observations read.
      93 observations have missing values.
      102 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=o-Xylene
NOTE: 196 observations read.
      94 observations have missing values.
      102 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=m,p-Xylene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 1.13 minutes.
  584
  585    ********************************************************************;
  586    * FOLLOWING STATEMENTS WERE ADDED AFTER REVIEWING THE PRELIMINARY
  587      REGRESSION RESULTS;
  588    PROC REG;BY CMPD;
  589    M3: MODEL Y3 = LCOUT
  590    LOC1 LOC2 HG5R HX7R HG6R HX9R HX10R HX11R HX12R HX15R HX16R HG19R
  591    HX22R HG23R HG24R HG25R HG26R HG27R HG31R HG40R HG41R HG42R HX44R PX61R
  592        /SELECTION=STEPWISE SLENTRY=.10 SLSTAY=.10;
```

Exhibit E-1. (continued)

```
  593     FORMAT CMPD CMPDF.;
  594     TITLE 'STEPWISE REGRESSION RESULTS';
  595     RUN;
NOTE: 188 observations read.
      84 observations have missing values.
      104 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=1,1,1-Trichloroethane
NOTE: 192 observations read.
      93 observations have missing values.
      99 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Benzene
NOTE: 197 observations read.
      96 observations have missing values.
      101 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Trichloroethylene
NOTE: 190 observations read.
      91 observations have missing values.
      99 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Tetrachloroethylene
NOTE: 157 observations read.
      87 observations have missing values.
      70 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Chlorobenzene
NOTE: 192 observations read.
      90 observations have missing values.
      102 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Styrene
NOTE: 192 observations read.
      93 observations have missing values.
      99 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=p-Dichlorobenzene
NOTE: 195 observations read.
      93 observations have missing values.
      102 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=o-Xylene
NOTE: 196 observations read.
      94 observations have missing values.
      102 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=m,p-Xylene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 30.00 seconds.
  596
  597    ******************************************************************;
```

Exhibit E-1. (continued)

```
598     * FOLLOWING STATEMENTS WERE ADDED AFTER REVIEWING THE STEPWISE
599       REGRESSION RESULTS;
600     DATA CC4 CC5 CC7 CC11 CC12 CC14 CC17 CC20 CC21;
601     SET CEXTRACT;
602     IF CMPD=4 THEN OUTPUT CC4;
603     ELSE IF CMPD=5 THEN OUTPUT CC5;
604     ELSE IF CMPD=7 THEN OUTPUT CC7;
605     ELSE IF CMPD=11 THEN OUTPUT CC11;
606     ELSE IF CMPD=12 THEN OUTPUT CC12;
607     ELSE IF CMPD=14 THEN OUTPUT CC14;
608     ELSE IF CMPD=17 THEN OUTPUT CC17;
609     ELSE IF CMPD=20 THEN OUTPUT CC20;
610     ELSE IF CMPD=21 THEN OUTPUT CC21;
611     RUN;
NOTE: The data set WORK.CC4 has 188 observations and 32 variables.
NOTE: The data set WORK.CC5 has 192 observations and 32 variables.
NOTE: The data set WORK.CC7 has 197 observations and 32 variables.
NOTE: The data set WORK.CC11 has 190 observations and 32 variables.
NOTE: The data set WORK.CC12 has 157 observations and 32 variables.
NOTE: The data set WORK.CC14 has 192 observations and 32 variables.
NOTE: The data set WORK.CC17 has 192 observations and 32 variables.
NOTE: The data set WORK.CC20 has 195 observations and 32 variables.
NOTE: The data set WORK.CC21 has 196 observations and 32 variables.
NOTE: The DATA statement used 40.00 seconds.
612     PROC REG SIMPLE DATA=CC4;BY CMPD;
613     MFINAL: MODEL Y3 = LCOUT LOC2 HG6R HX11R HX12R HG26R HG42R;
614     FORMAT CMPD CMPDF.;
615     TITLE 'FINAL REGRESSION RESULTS';
616     RUN;
NOTE: 188 observations read.
      82 observations have missing values.
      106 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=1,1,1-Trichloroethane
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 5.00 seconds.
617     PROC REG SIMPLE DATA=CC5;BY CMPD;
618     MFINAL: MODEL Y3 = LCOUT HX7R HX9R HG25R HG40R;
619     FORMAT CMPD CMPDF.;
620     TITLE 'FINAL REGRESSION RESULTS';
621     RUN;
NOTE: 192 observations read.
      91 observations have missing values.
      101 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Benzene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 4.00 seconds.
622     PROC REG SIMPLE DATA=CC7;BY CMPD;
623     MFINAL: MODEL Y3 = LCOUT HX7R HG6R HG42R;
624     FORMAT CMPD CMPDF.;
625     TITLE 'FINAL REGRESSION RESULTS';
```

Exhibit E-1. (continued)

```
  626     RUN;
NOTE: 197 observations read.
      94 observations have missing values.
      103 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Trichloroethylene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 5.00 seconds.
  627     PROC REG SIMPLE DATA=CC11;BY CMPD;
  628     MFINAL: MODEL Y3 = LCOUT HG5R HG6R HX9R HX22R HG23R HG24R HG42R;
  629     FORMAT CMPD CMPDF.;
  630     TITLE 'FINAL REGRESSION RESULTS';
  631     RUN;
NOTE: 190 observations read.
      89 observations have missing values.
      101 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Tetrachloroethylene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 5.00 seconds.
  632     PROC REG SIMPLE DATA=CC12;BY CMPD;
  633     MFINAL: MODEL Y3 = LCOUT HX7R HG41R;
  634     FORMAT CMPD CMPDF.;
  635     TITLE 'FINAL REGRESSION RESULTS';
  636     RUN;
NOTE: 157 observations read.
      84 observations have missing values.
      73 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Chlorobenzene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 5.00 seconds.
  637     PROC REG SIMPLE DATA=CC14;BY CMPD;
  638     MFINAL: MODEL Y3 = LCOUT HX7R HX9R HG23R HG25R;
  639     FORMAT CMPD CMPDF.;
  640     TITLE 'FINAL REGRESSION RESULTS';
  641     RUN;
NOTE: 192 observations read.
      88 observations have missing values.
      104 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=Styrene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 5.00 seconds.
  642     PROC REG SIMPLE DATA=CC17;BY CMPD;
  643     MFINAL: MODEL Y3 = LCOUT LOC2 HG5R HX12R HX15R;
  644     FORMAT CMPD CMPDF.;
  645     TITLE 'FINAL REGRESSION RESULTS';
  646     RUN;
NOTE: 192 observations read.
      90 observations have missing values.
      102 observations used in computations.
```

Exhibit E-1. (continued)

```
NOTE: The above message was for the following by-group:
      CMPD=p-Dichlorobenzene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 5.00 seconds.
  647    PROC REG SIMPLE DATA=CC20;BY CMPD;
  648    MFINAL: MODEL Y3 = LCOUT LOC1 LOC2 HX7R HG23R HG25R;
  649    FORMAT CMPD CMPDF.;
  650    TITLE 'FINAL REGRESSION RESULTS';
  651    RUN;
NOTE: 195 observations read.
      91 observations have missing values.
      104 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=o-Xylene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 7.00 seconds.
  652    PROC REG SIMPLE DATA=CC21;BY CMPD;
  653    MFINAL: MODEL Y3 = LCOUT LOC1 HX7R HG6R HX11R HG26R HX44R;
  654    FORMAT CMPD CMPDF.;
  655    TITLE 'FINAL REGRESSION RESULTS';
  656    RUN;
NOTE: 196 observations read.
      92 observations have missing values.
      104 observations used in computations.
NOTE: The above message was for the following by-group:
      CMPD=m,p-Xylene
NOTE: Interactivity disabled with BY processing.
NOTE: The PROCEDURE REG used 5.00 seconds.
  657    ********************************************************************
```

Exhibit E-2. Analysis Type E: Complete Stepwise Regression Results for CMPD=1,1,1-Trichloroethane

Stepwise Procedure for Dependent Variable Y3

Step 1 Variable LCOUT Entered R-square = 0.36454451 C(p) = 19.90651702

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	1	38.91860349	38.91860349	58.51	0.0001
Error	102	67.84093582	0.66510721		
Total	103	106.75953931			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	1.30906752	0.12179415	76.83567848	115.52	0.0001
LCOUT	0.57471249	0.07513077	38.91860349	58.51	0.0001

Bounds on condition number: 1, 1

Step 2 Variable HG42R Entered R-square = 0.42023823 C(p) = 11.39745685

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	2	44.86443941	22.43221971	36.60	0.0001
Error	101	61.89509990	0.61282277		
Total	103	106.75953931			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	1.24658324	0.11861756	67.68298730	110.44	0.0001
LCOUT	0.55838338	0.07230758	36.54535297	59.63	0.0001
HG42R	0.77952440	0.25025957	5.94583592	9.70	0.0024

Bounds on condition number: 1.005284, 4.021136

Step 3 Variable HX11R Entered R-square = 0.45086539 C(p) = 7.61830004

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	3	48.13418167	16.04472722	27.37	0.0001
Error	100	58.62535764	0.58625358		
Total	103	106.75953931			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	1.26189200	0.11619866	69.13972729	117.93	0.0001
LCOUT	0.59810958	0.07269573	39.68526482	67.69	0.0001
HX11R	-0.53117556	0.22491793	3.26974226	5.58	0.0201
HG42R	0.89988358	0.25002372	7.59443979	12.95	0.0005

Bounds on condition number: 1.10767, 9.656071

Exhibit E-2. (continued)

Step 4 Variable HG26R Entered R-square = 0.47876627 C(p) = 4.35357828

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	4	51.11286590	12.77821648	22.73	0.0001
Error	99	55.64667341	0.56208761		
Total	103	106.75953931			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	1.06512028	0.14230946	31.48718227	56.02	0.0001
LCOUT	0.64942195	0.07459006	42.60847764	75.80	0.0001
HX11R	-0.56095696	0.22061313	3.63412954	6.47	0.0125
HG26R	0.36912235	0.16034673	2.97868423	5.30	0.0234
HG42R	0.93256230	0.24522759	8.12869911	14.46	0.0002

Bounds on condition number: 1.16631, 17.73313

Step 5 Variable HG6R Entered R-square = 0.49383204 C(p) = 3.51076061

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	5	52.72128066	10.54425613	19.12	0.0001
Error	98	54.03825865	0.55141080		
Total	103	106.75953931			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	1.16291951	0.15213924	32.21752087	58.43	0.0001
LCOUT	0.65589821	0.07397550	43.34832749	78.61	0.0001
HG6R	-0.25659776	0.15024204	1.60841476	2.92	0.0908
HX11R	-0.57222591	0.21860741	3.77816168	6.85	0.0103
HG26R	0.43678789	0.16368378	3.92649609	7.12	0.0089
HG42R	0.92767035	0.24290428	8.04252268	14.59	0.0002

Bounds on condition number: 1.171737, 27.85252

Step 6 Variable LOC2 Entered R-square = 0.51023264 C(p) = 2.41606771

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	6	54.47220186	9.07870031	16.84	0.0001
Error	97	52.28733746	0.53904472		
Total	103	106.75953931			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	1.48426111	0.23327532	21.82266668	40.48	0.0001
LCOUT	0.52322494	0.10377246	13.70368858	25.42	0.000[illegible]
LOC2	-0.42838712	0.23769251	1.75092120	3.25	0.0746
HG6R	-0.30813643	0.15127527	2.23653150	4.15	0.0444

Exhibit E-2. (continued)

HX11R	-0.52773352	0.21754748	3.17209395	5.88	0.0171
HG26R	0.54942899	0.17348690	5.40647975	10.03	0.0021
HG42R	0.85273285	0.24373783	6.59788750	12.24	0.0007

Bounds on condition number: 2.624302, 57.83772

Step 7 Variable HX12R Entered R-square = 0.52465732 C(p) = 1.69422021

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	7	56.01217388	8.00173913	15.14	0.0001
Error	96	50.74736544	0.52861839		
Total	103	106.75953931			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	1.49076247	0.23103967	22.00827709	41.63	0.0001
LCOUT	0.52045697	0.10277676	13.55570509	25.64	0.0001
LOC2	-0.44709053	0.23563747	1.90302503	3.60	0.0608
HG6R	-0.29893144	0.14990217	2.10217852	3.98	0.0490
HX11R	-0.76000522	0.25481501	4.70246314	8.90	0.0036
HX12R	0.68298635	0.40015362	1.53997202	2.91	0.0911
HG26R	0.54490993	0.17182129	5.31664601	10.06	0.0020
HG42R	0.87305873	0.24166270	6.89937798	13.05	0.0005

Bounds on condition number: 2.62999, 80.79201

All variables in the model are significant at the 0.1000 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y3

Step	Variable Entered	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	LCOUT		1	0.3645	0.3645	19.9065	58.5148	0.0001
2	HG42R		2	0.0557	0.4202	11.3975	9.7024	0.0024
3	HX11R		3	0.0306	0.4509	7.6183	5.5774	0.0201
4	HG26R		4	0.0279	0.4788	4.3536	5.2993	0.0234
5	HG6R		5	0.0151	0.4938	3.5108	2.9169	0.0908
6	LOC2		6	0.0164	0.5102	2.4161	3.2482	0.0746
7	HX12R		7	0.0144	0.5247	1.6942	2.9132	0.0911

Exhibit E-3. Analysis Type E: Summary of Stepwise Regression Results

CMPD=Benzene

Stepwise Procedure for Dependent Variable Y3

Step 5 Variable HX9R Entered R-square = 0.49282883 C(p) = -2.82914198

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	5	47.48209747	9.49641949	18.07	0.0001
Error	93	48.86392407	0.52541854		
Total	98	96.34602154			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	0.48453019	0.16166795	4.71953503	8.98	0.0035
LCOUT	0.69929266	0.09372929	29.24641689	55.66	0.0001
HX7R	0.67880531	0.21684759	5.14857036	9.80	0.0023
HX9R	-0.32814184	0.17557666	1.83524916	3.49	0.0648
HG25R	0.34936601	0.15564847	2.64714031	5.04	0.0272
HG40R	0.32018680	0.15860747	2.14123819	4.08	0.0464

Bounds on condition number: 1.377657, 30.41227

All variables in the model are significant at the 0.1000 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y3

Step	Variable Entered	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	LCOUT		1	0.3930	0.3930	5.7362	62.8061	0.0001
2	HX7R		2	0.0415	0.4345	0.8457	7.0487	0.0093
3	HG25R		3	0.0198	0.4543	-0.4352	3.4416	0.0667
4	HG40R		4	0.0195	0.4738	-1.6678	3.4794	0.0653
5	HX9R		5	0.0190	0.4928	-2.8291	3.4929	0.0648

Exhibit E-3. (continued)

CMPD=Trichloroethylene

Stepwise Procedure for Dependent Variable Y3

Step 4 Variable HG42R Entered R-square = 0.24938280 C(p) = 7.15226967

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	4	66.57317884	16.64329471	7.97	0.0001
Error	96	200.37858925	2.08727697		
Total	100	266.95176809			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	-0.22787937	0.36923820	0.79501784	0.38	0.5386
LCOUT	0.47541241	0.10197904	45.36273885	21.73	0.0001
HX7R	1.41075912	0.42510916	22.98713707	11.01	0.0013
HG6R	0.61146836	0.28826684	9.39158039	4.50	0.0365
HG42R	0.91604217	0.48870026	7.33374625	3.51	0.0639

Bounds on condition number: 1.280708, 18.33892

All variables in the model are significant at the 0.1000 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y3

Step	Variable Entered	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	LCOUT		1	0.1105	0.1105	19.3119	12.2994	0.0007
2	HX7R		2	0.0790	0.1895	10.9783	9.5556	0.0026
3	HG6R		3	0.0324	0.2219	8.7446	4.0363	0.0473
4	HG42R		4	0.0275	0.2494	7.1523	3.5135	0.0639

Exhibit E-3. (continued)

CMPD=Tetrachloroethylene

Stepwise Procedure for Dependent Variable Y3

Step14 Variable HG5R Entered R-square = 0.70287445 C(p) = 3.41007041

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	8	200.80925582	25.10115698	26.61	0.0001
Error	90	84.88793455	0.94319927		
Total	98	285.69719037			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	0.06309576	0.22911166	0.07153349	0.08	0.7836
LCOUT	0.64282220	0.06809864	84.04431938	89.11	0.0001
HG5R	0.39596907	0.22002898	3.05468354	3.24	0.0753
HG6R	0.97858990	0.23451491	16.42343493	17.41	0.0001
HX9R	-0.67004825	0.23662457	7.56303216	8.02	0.0057
HX22R	0.49977683	0.20956695	5.36426330	5.69	0.0192
HG23R	-0.54382269	0.22190942	5.66456610	6.01	0.0162
HG24R	-0.65249870	0.26997368	5.50959800	5.84	0.0177
HG42R	0.91999124	0.31397526	8.09804096	8.59	0.0043

Bounds on condition number: 1.517401, 79.22788

All variables in the model are significant at the 0.1000 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y3

Step	Variable Entered	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	LCOUT		1	0.5639	0.5639	28.9028	125.4053	0.0001
2	LOC2		2	0.0488	0.6127	17.0312	12.1026	0.0008
3	LOC1		3	0.0146	0.6273	14.8908	3.7146	0.0569
4	HG5R		4	0.0119	0.6392	13.5040	3.1058	0.0813
5	HG6R		5	0.0115	0.6507	12.2312	3.0673	0.0832
6	HX9R		6	0.0121	0.6628	10.8024	3.2927	0.0728
7	HG23R		7	0.0123	0.6751	9.3048	3.4481	0.0666
8	HG42R		8	0.0126	0.6877	7.7287	3.6274	0.0600
9		LOC2	7	0.0075	0.6801	7.8706	2.1726	0.1440
10		HG5R	6	0.0090	0.6712	8.4165	2.5495	0.1138
11		LOC1	5	0.0098	0.6614	9.2025	2.7438	0.1010
12	HG24R		6	0.0153	0.6767	6.8570	4.3523	0.0397
13	HX22R		7	0.0155	0.6922	4.4476	4.5886	0.0349
14	HG5R		8	0.0107	0.7029	3.4101	3.2386	0.0753

Exhibit E-3. (continued)

CMPD=Chlorobenzene

Stepwise Procedure for Dependent Variable Y3

Step 3 Variable HG41R Entered R-square = 0.12823959 C(p) = -9.12011505

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	3	6.47780383	2.15926794	3.24	0.0277
Error	66	44.03548616	0.66720434		
Total	69	50.51328999			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	-2.04059737	0.49576858	11.30356671	16.94	0.0001
LCOUT	0.28684627	0.14803743	2.50503721	3.75	0.0569
HX7R	0.51997726	0.23329552	3.31447516	4.97	0.0292
HG41R	-0.37280020	0.21682075	1.97246553	2.96	0.0902

Bounds on condition number: 1.090766, 9.54664

All variables in the model are significant at the 0.1000 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y3

Step	Variable Entered	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	LCOUT		1	0.0456	0.0456	-8.1089	3.2510	0.0758
2	HX7R		2	0.0436	0.0892	-8.7515	3.2046	0.0779
3	HG41R		3	0.0390	0.1282	-9.1201	2.9563	0.0902

Exhibit E-3. (continued)

CMPD=Styrene

Stepwise Procedure for Dependent Variable Y3

Step 5 Variable HG25R Entered R-square = 0.43133667 C(p) = -1.13121168

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	5	38.57253611	7.71450722	14.56	0.0001
Error	96	50.85305291	0.52971930		
Total	101	89.42558903			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	0.62622254	0.13331943	11.68735722	22.06	0.0001
LCOUT	0.35097866	0.05463846	21.85799080	41.26	0.0001
HX7R	0.77767008	0.21719174	6.79125406	12.82	0.0005
HX9R	-0.48757224	0.16751524	4.48760990	8.47	0.0045
HG23R	-0.60002281	0.16305614	7.17310158	13.54	0.0004
HG25R	0.43879261	0.16316777	3.83085449	7.23	0.0084

Bounds on condition number: 1.320056, 29.93497

All variables in the model are significant at the 0.1000 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y3

Step	Variable Entered Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	LCOUT	1	0.2786	0.2786	14.7330	38.6255	0.0001
2	HG23R	2	0.0336	0.3122	11.4865	4.8322	0.0303
3	HX7R	3	0.0315	0.3437	8.5652	4.7023	0.0325
4	HX9R	4	0.0448	0.3885	3.5634	7.1070	0.0090
5	HG25R	5	0.0428	0.4313	-1.1312	7.2319	0.0084

Exhibit E-3. (continued)

CMPD=p-Dichlorobenzene

Stepwise Procedure for Dependent Variable Y3

Step 5 Variable HX12R Entered R-square = 0.51123419 C(p) = 2.67937102

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	5	182.03670131	36.40734026	19.46	0.0001
Error	93	174.03631553	1.87135823		
Total	98	356.07301683			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	-0.89404458	0.79551715	2.36361197	1.26	0.2640
LCOUT	0.88613398	0.10180590	141.77830330	75.76	0.0001
LOC2	1.42339042	0.34784213	31.33578929	16.74	0.0001
HG5R	-1.18719760	0.31558158	26.48376163	14.15	0.0003
HX12R	1.11419838	0.62975155	5.85792069	3.13	0.0801
HX15R	2.22895566	0.80913159	14.20108335	7.59	0.0071

Bounds on condition number: 1.572638, 30.90195

All variables in the model are significant at the 0.1000 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y3

Step	Variable Entered	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	LCOUT		1	0.3330	0.3330	27.3844	48.4244	0.0001
2	LOC2		2	0.0601	0.3931	18.3592	9.5046	0.0027
3	HG5R		3	0.0599	0.4530	9.3668	10.4046	0.0017
4	HX15R		4	0.0418	0.4948	3.6979	7.7766	0.0064
5	HX12R		5	0.0165	0.5112	2.6794	3.1303	0.0801

Exhibit E-3. (continued)

CMPD=o-Xylene

Stepwise Procedure for Dependent Variable Y3

Step 6 Variable HG23R Entered R-square = 0.63526193 C(p) = 2.59259295

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	6	53.25206713	8.87534452	27.58	0.0001
Error	95	30.57487806	0.32184082		
Total	101	83.82694519			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	0.81846592	0.15618076	8.83867717	27.46	0.0001
LCOUT	0.48612259	0.08465185	10.61352998	32.98	0.0001
LOC1	0.37742271	0.16057793	1.77797432	5.52	0.0208
LOC2	-0.33063121	0.18410428	1.03800880	3.23	0.0757
HX7R	0.73789659	0.17940038	5.44484384	16.92	0.0001
HG23R	-0.26761025	0.12928266	1.37900572	4.28	0.0412
HG25R	0.34325288	0.13125769	2.20099982	6.84	0.0104

Bounds on condition number: 2.619431, 64.834

All variables in the model are significant at the 0.1000 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y3

Step	Variable Entered	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	LCOUT		1	0.5188	0.5188	21.5307	107.7936	0.0001
2	LOC1		2	0.0300	0.5488	16.0740	6.5868	0.0118
3	HX7R		3	0.0309	0.5796	10.4064	7.1971	0.0086
4	LOC2		4	0.0243	0.6039	6.3768	5.9452	0.0166
5	HG25R		5	0.0149	0.6188	4.6786	3.7499	0.0558
6	HG23R		6	0.0165	0.6353	2.5926	4.2847	0.0412

Exhibit E-3. (continued)

CMPD=m,p-Xylene

Stepwise Procedure for Dependent Variable Y3

Step 7 Variable HX44R Entered R-square = 0.68571431 C(p) = 1.37647278

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	7	64.85062746	9.26437535	29.30	0.0001
Error	94	29.72320080	0.31620426		
Total	101	94.57382826			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	0.41703146	0.34336954	0.46642501	1.48	0.2276
LCOUT	0.57993238	0.06969495	21.89373086	69.24	0.0001
LOC1	0.51468942	0.16822220	2.95999921	9.36	0.0029
HX7R	0.58241603	0.17082581	3.67559261	11.62	0.0010
HG6R	0.33332349	0.11668595	2.58025400	8.16	0.0053
HX11R	-0.32082419	0.17101022	1.11290428	3.52	0.0638
HG26R	-0.28749140	0.13370016	1.46202081	4.62	0.0341
HX44R	0.58925521	0.33523724	0.97694491	3.09	0.0820

Bounds on condition number: 2.043786, 69.94166

All variables in the model are significant at the 0.1000 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable Y3

Step	Variable Entered	Variable Removed	Number In	Partial R**2	Model R**2	C(p)	F	Prob>F
1	LCOUT		1	0.5800	0.5800	18.7718	138.0850	0.0001
2	LOC1		2	0.0244	0.6044	13.9782	6.1154	0.0151
3	HX7R		3	0.0254	0.6298	8.9100	6.7310	0.0109
4	HG6R		4	0.0193	0.6491	5.5501	5.3297	0.0231
5	HG26R		5	0.0151	0.6642	3.3590	4.3097	0.0406
6	HX11R		6	0.0112	0.6754	2.2484	3.2744	0.0735
7	HX44R		7	0.0103	0.6857	1.3765	3.0896	0.0820

CARB LIBRARY
05154