DEMOGRAPHIC DATA FOR LEAD EXPOSURE ANALYSES

by

Roy A. Paul
PEI Associates, Inc.
505 South Duke Street, Suite 503
Durham, North Carolina 27701-3196

Contract No. 68-02-4309
Work Assignment No. 14
PN 3659-14

Project Officer

Richard B. Atherton

Work Assignment Manager

Jeff Cohen

STRATEGIES AND AIR STANDARDS DIVISION
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

January 1986

DISCLAIMER

This report was prepared by PEI Associates, Inc., Cincinnati, Ohio, under Contract No. 68-02-4309, Work Assignment No. 14. It has been reviewed by the Strategies and Air Standards Division of the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products is not intended to constitute endorsement or recommendation for use.

CONTENTS

Figures iv
Tables. v
Acknowledgment. vi

1. Introduction. 1

2. Geographic Units of Analysis. 3

3. Population Data 4

Bureau of census data. 4
Age groups 4
Lead in homes. 5
Secondary occupational exposure. 7
Pregnant women 8
Population projections 13

4. Outputs from the demographic program. 15

References. 24

Appendix

Source code for demographic program. 25

FIGURES

Number		Page
1	Sample of Air Quality Data Used by Program MATCH-AQ	16
2	Pregnancy Data Used in Program MATCH-AQ	17
3	User-Specified Inputs for Program MATCH-AQ.	18
4	Sample of Geographic Data Table for Census Tracts Around a Point Source.	20
5	Sample of Air Quality Data Table for Census Tracts Around a Point Source.	21
6	Sample of Demographic Data Table for One Age Group.	22
7	Sample of Output File Records	23

TABLES

Number		Page
1	Estimates of Percentage of Homes With Lead-Based Paint. . . .	6
2	Birth Rates by Age Group of Mother and by Race of Child, 1981.	9
3	Fetal Death Ratios by Age Group of Mother and Race of Child, 1978.	10
4	Number of Fetal Deaths by Specified Periods of Gestation for States Requiring Reporting of All Fetal Deaths, 1978. . . .	11
5	BEA Population Projections and Corresponding Projection Fractions	14

ACKNOWLEDGMENT

This report was prepared for the U.S. Environmental Protection Agency by PEI Associates, Inc., Cincinnati, Ohio. Mr. Jeff Cohen was the EPA Work Assignment Manager. Mr. David Dunbar served as the Project Director; Mr. Roy Paul was the Project Manager. The principal author was Mr. Roy Paul. The author would like to thank Mr. Jeff Cohen for his guidance and direction on this work assignment. The author would also like to acknowledge Mr. Ted Johnson and Ms. Alicia Ferdo for their assistance and cooperation.

SECTION 1

INTRODUCTION

Under the Clean Air Act, the U.S. Environmental Protection Agency (EPA) is responsible for establishing National Ambient Air Quality Standards (NAAQS) and for reviewing them periodically to determine their adequacy on the basis of recent experience and research. In line with these responsibilities, the Strategies and Air Standards Division (SASD) of the Office of Air Quality Planning and Standards (OAQPS) is exploring the use of quantitative methods for assessing health risks associated with proposed air quality standards.

An important aspect of a health risk assessment is the estimation of population exposure. For the past few years, SASD has been engaged in developing an exposure model suitable for evaluating alternative ambient air standards. The model is known as NEM, an acronym for NAAQS Exposure Model.

Although different versions of NEM have been developed for different pollutants, the various versions are similar in that each considers the effect of a person's activity patterns on his or her exposure. Activity patterns are usually represented as a sequence of assignments to a geographic location, to a microenvironment (e.g., indoors), and to an activity level.[1] The exposure at each point in time is determined from these assignments.

The SASD is currently assessing health risks associated with alternative NAAQS for lead. Estimation of population exposure to lead is a four-step process involving 1) development of air quality data, 2) development of activity data, 3) development of demographic data, and 4) calculation of exposures. Analyses of lead exposures have been simplified somewhat because dispersion model estimates are used in lieu of air quality monitoring data and because activity patterns are simplified to the average times spent in various microenvironments. The bases for these simplifications are discussed in an OAQPS staff paper.[2]

The GCA Corporation provided PEI Associates, Inc. (PEI) with air quality estimates for selected geographic locations around point sources. The model

used for these estimates was the Industrial Source Complex - Short-Term (ISC-ST) dispersion model, a standard model frequently used to estimate levels of control for point-source emissions. The ISC was used to estimate monthly average lead concentrations at the centers of census tracts located near given point sources. Point sources were located within one of four Standard Metropolitan Statistical Areas (SMSA's): Chicago, Dallas, Nashville, or Tampa-St. Petersburg.

Given air quality estimates at selected locations, PEI developed a computer program (MATCH-AQ) to calculate the number of children and pregnant women that may be exposed to estimated lead concentrations at each location. The number of children in each census tract was further subdivided by year of age. Age groups were further subdivided into three housing types: 1) lead-based-painted houses (LPH's) with deteriorated or peeling paint, 2) LPH's with sound paint, and 3) non-LPH homes (no lead-based paints). Housing-type groups were further subdivided into two groups representing high and low probability of experiencing secondary occupational exposure. These data are used by a second computer program that calculates uptake of lead into children's bodies. The uptake program is covered in another report,[3] which describes the population program, the geographic units used, the demographic data used, and the data generated by the program.

SECTION 2

GEOGRAPHIC UNITS OF ANALYSIS

Census tracts are relatively small units of geography that can be used effectively in population exposure analyses. Census tracts are statistical subdivisions of a county that generally have stable boundaries. When established, they are designed to be relatively homogeneous with regard to population characteristics, economic status, and living conditions. Although the land area of census tracts may vary, they usually have between 2500 and 8000 residents.[4] All standard metropolitan statistical areas (SMSA's) recognized for the 1980 census can be defined as groups of census tracts. Consequently, census tracts are particularly useful geographic units for population exposure analyses. PEI and EPA have developed a library of detailed census tract data from the 1980 Census of Population and Housing.

The location of the population within a census tract may conveniently be referenced by a single point, the population centroid, i.e., the average latitude and longitude of all the people in the census tract. Population centroids are usually estimated by the population-weighted average of all geographic centers of smaller units within the census tract, such as block groups. If the population can be assumed to be distributed homogeneously within the census tract, the geographic center of the census tract also may be used to estimate the population centroid.

For this study, geographic centers of census tracts were readily available on a special computer tape prepared by Geographic Data Technology, Inc. (GDT) in 1981. This special file was developed as a side product of digitizing latitude and longitude values representing boundaries of census tracts in SMSA's. Geographic centers were defined by the average longitude and latitude of tract boundaries.[5] In 1983, PEI provided GCA Technology, Inc., with lists of census tract centers for selected SMSA's for use in lead air quality modeling.

SECTION 3

POPULATION DATA

3.1 BUREAU OF CENSUS DATA

PEI and EPA have established a library of 1980 census data that resides on a series of 72 computer tapes at the National Computer Center (NCC). These data include summary tape file 3A (STF3A), one of a series of summary data files created by the Bureau of Census (BOC) for use in analyses requiring detailed information. STF3A contains summary computer records consisting of 12,096 characters (numbers and codes). One record exists for each one of a hierarchy of geographic units, such as State, county, minor civil divisions, places, census tracts, and block groups or enumeration districts. For example, there is one record for each one of the approximately 45,000 census tracts in the United States.[6]

Each STF3A record contains a series of identification codes and flags plus 150 tables relating to various demographic characteristics. These include age, sex, race, family status, types and ages of housing, places of work, commuting time and distance, school enrollment, disabilities, types of industries, types of employment, income and poverty status, household facilities, value of houses, and householder characteristics. In conducting population analyses, only selected tables are used for a particular geographic area.

3.2 AGE GROUPS

The EPA wishes to analyze lead exposures of seven age groups of children, age 0 through 6. The principal population data for this analysis is from STF3A Table 15, Sex by Age. The age categories listed in Table 15 for children under 7 are 1) under 1 year, 2) 1 and 2 years, 3) 3 and 4 years, 4) 5 years, and 5) 6 years. Each age year is determined by whether a person has reached the birthday indicated when the census is taken.

Because BOC age categories did not correspond exactly to the needed age groups, they were estimated when necessary. Child age group 1 corresponds exactly to BOC Category 1. Child age group 2 (1 to 1.99 years) was estimated by dividing BOC group 2 (age 1 and 2 years) by 2. To avoid rounding error, age group 3 (2 to 2.99 years) was estimated as BOC group 2 minus age group 2. Likewise, age group 4 was estimated by dividing BOC Category 3 by 2; age group 5 was estimated as BOC Category 3 minus age group 4. Age groups 6 and 7 correspond to BOC Categories 4 and 5, respectively.

Other age data are used in estimating numbers of pregnant women. The female age groups are 15 to 19, 20 to 34, and 35 to 44 years. The BOC age categories include exact subdivisions of these groups; therefore, the BOC categories were simply aggregated to the larger age groups.

3.3 LEAD IN HOMES

PEI developed a model for estimating the number of children 6 years and under who live in LPH's. The model also estimates the number of homes that have chalking, flaking, or peeling paint (referred to as unsound homes). Both of these models were incorporated into the demographic program to estimate LPH-unsound, LPH-sound, and non-LPH homes in each census tract.

Pope[7] reviewed literature on lead-based paint to select estimators of LPH. The most reliable independent variable was found to be the age of the home. Table 1 presents a range of estimates (from Pope) of percentages of homes built in various years that are LPH's. Table 109 in census tape STF3A lists the numbers of homes in each census tract that were built in each decade after 1940. These decade data were then aggregated into the same age categories of homes shown in Table 1. The number of homes in each age category times the fraction of LPH's in each category (PEI's best estimate) yields an estimate of the number of LPH's in each census tract.

Assuming that LPH's were distributed evenly across all age groups, the fraction of children living in LPH's was assumed to be the same as the fraction of all housing units in the census tract that are LPH's.

TABLE 1. ESTIMATES OF PERCENTAGE OF HOMES WITH LEAD-BASED PAINT[7]

Study	Amount of lead in painted surface	Pre-1940	1940 to 1959	1960 to 1974	1975 to 1980
EPA Criteria Document	>1.5 mg/cm[a]			20	
OAQPS Draft Lead Staff Paper	>1.5 mg/cm^2			22	
Billick and Gray	>1.5 mg/cm^2	75	50	20	
Lin-Fu	>2.0 mg/cm^2		43[a]	12[a]	
Shier and Hall	>1.0 mg/cm^2	77	26	15	
Arizona Health Depart-ment	>1.0 mg/cm^2	100	90		
Minneapolis	>1.0 mg/cm^2	100			
Gilsinn	>1%	100	50	0	
PEI's Best Estimate	>0.7 mg/cm^2	99	70	20	0
PEI's Range	>0.7 mg/cm^2	75-100	26-90	0-22	

[a]Estimate for 1975.

3.4 SECONDARY OCCUPATIONAL EXPOSURE

PEI developed a method for estimating the secondary occupational exposure (SOE) of children 6 years or younger. SOE occurs when a member of a family works in a production facility involving lead, for example, a lead smelter or lead battery plant. Particles of lead may be embedded in the clothing, skin, or hair of the employee and transferred to other members of the employee's family by personal contact.

The number of employees occupationally exposed to lead was estimated as the number of lead production workers at the point source of lead being analyzed. These data were determined by contacting either the company or the local chamber of commerce (if the company did not provide data). The distribution of the homes of these workers could not be determined directly from census data, so it was estimated. Table 66 of Census tape STF3A lists numbers of employed persons by broad categories of industry. Durable goods manufacturing (DGM) is an industrial category that includes various primary and secondary smelters and production facilities, including smelters and lead battery production. Thus, lead-exposed employees were assumed to be geographically distributed around the point source in the same manner as DGM employees, as follows:

$$L_t = (L_T)(DGM_t/DGM_T) \qquad (1)$$

where L_t = Lead-exposed employees in census tract t

L_T = Total lead-exposed employees around point source

DGM_t = Employees in durable goods manufacturing in census tract t

DGM_T = Total employees in durable goods manufacturing.

The number of children with SOE may differ from the number of exposed employees because the number of children per employee will vary. To find the number of SOE children, the child/employee ratio was calculated for each census tract:

$$CE_t = \sum_{y=0}^{6} C_{ty}/E_t \qquad (2)$$

where CE_t = Child-employee ratio for census tract t

C_{ty} = Number of children age y in census tract t

E_t = Total employees in census tract t.

Thus, the fraction of all children age 0 to 6 with SOE could be estimated as follows:

$$F_t = (L_t)(CE_t)/C_t \quad (3)$$

where F_t = Fraction of children with SOE in census tract t

C_t = Total children age 0 to 6 in census tract t

L_t = Lead-exposed employees in census tract t

CE_t = Child-employee ratio for census tract t.

Assuming all age groups and housing groups of children had the same probability of SOE, each group was multiplied by the same fraction (F_t) to estimate the number of children in each group with a high probability of having SOE. At the point sources under study, only a few children received SOE. When these were distributed over a large number of census tracts (e.g., 100), seven age groups, and three housing categories, only a small fraction ($<<0.5$) of a child was estimated for most demographic subgroups. Rounded off, it was usually zero. Thus, it was necessary to adjust the distributions of SOE children around the point source by multiplying the number of children in each SOE subgroup by an empirically determined adjustment fraction. The subgroup populations were repeatedly adjusted until the total of rounded numbers of all children in SOE subgroups was exactly equal to the estimated number of SOE children around the point source.

3.5 PREGNANT WOMEN

The number of pregnant women in a census tract may be estimated by 1) the specific race and age distribution of the census tract, and 2) national race and age-specific birth rates and fetal death rates. Pregnancy may lead to either a live birth or a fetal death, thus both rates are required. Methodology for these estimates was developed by PEI in 1983.[8]

A key source of statistics about birth rates is Vital Statistics of the U.S., produced by the National Center for Health Statistics.[9,10] Data are accumulated from birth certificates from every State in the Nation. As a result of birth registration laws that have been in effect since 1915, data on births are better than 90 percent complete in all States and average 99.6 percent complete. Complete natality statistics were available for the year 1978, however, an advance report[11] for 1980 provided sufficient data needed for estimating cohort populations. For example, birth rates can be obtained by race of child and by age of mother, as shown in Table 2. Note that birth rates are substantially higher for blacks than for nonblacks in all age categories.

TABLE 2. BIRTH RATES BY AGE GROUP OF MOTHER AND BY RACE OF CHILD, 1981

Age group	Race	Number of women	Number of live births	Birth rate[a]
15-19	Black	1,503,530	150,353	100.0
	Nonblack	8,914,602	401,808	45.1
20-34	Black	3,699,630	409,645	110.7
	Nonblack	29,360,935	2,475,200	84.3
35-44	Black	1,489,196	23,621	15.9
	Nonblack	11,542,075	140,262	12.2

[a]Number of births per 1000 women in race and age group.

Another key source of data is fetal death certificates, a standard form used since 1978 to record any kind of pregnancy termination.[11] Fetal deaths are consistently reported from all States for interrupted pregnancies with a gestation period greater than 20 weeks in length. Statistics on fetal death ratios can be obtained by specific age and race groups, as shown in Table 3. Fetal death ratios (based on reported fetal deaths) vary significantly for different races. For example, in 1978 the fetal death ratio was 8.5 per 1000 live births for "white" babies and 14.7 per 1000 for "all others." These fetal death ratios can be used to estimate the number of pregnancies that result in fetal death.