NUREG/CR-2364 Vol. IV
ANL-81-59 Vol. IV

NUREG/CR-2364 Vol. IV
ANL-81-59 Vol. IV

PROJECTION MODELS FOR HEALTH EFFECTS ASSESSMENT IN POPULATIONS EXPOSED TO RADIOACTIVE AND NONRADIOACTIVE POLLUTANTS

Volume IV

SPAHR User's Guide

by

James J. Collins and Robert T. Lundy

ARGONNE NATIONAL LABORATORY

ARGONNE NATIONAL LABORATORY, ARGONNE, ILLINOIS

Prepared for the Office of Nuclear Regulatory Research
U. S. NUCLEAR REGULATORY COMMISSION
under Interagency Agreement DOE 40-550-75

Available from

GPO Sales Program
Division of Technical Information and Document Control
U. S. Nuclear Regulatory Commission
Washington, D.C. 20555

and

National Technical Information Service
Springfield, Virginia 22161

NUREG/CR-2364 Vol. IV
ANL-81-59 Vol. IV

(Distribution
Code: RH)

NUREG/CR--2364-Vol.4

DE83 001020

ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

PROJECTION MODELS FOR HEALTH-EFFECTS ASSESSMENT IN POPULATIONS EXPOSED TO RADIOACTIVE AND NONRADIOACTIVE POLLUTANTS

Volume IV

SPAHR User's Guide

MASTER

by

James J. Collins and Robert T. Lundy

Division of Biological and Medical Research

September 1982

Prepared for the Health Effects Branch
Division of Health, Siting and Waste Management
Office of Nuclear Regulatory Research
U. S. Nuclear Regulatory Commission
Washington, D. C. 20555
Under Interagency Agreement DOE 40-550-75

NRC FIN No. A2059

This is Volume IV of a five volume series entitled Projection Models for Health Effects Assessment in Populations Exposed to Radioactive and Nonradioactive Pollutants, NUREG/CR-2364, ANL-81-59. The series presents version 4.1 of the Simulation Package for the Analysis of Health Risk (SPAHR) computer package and model. The complete series of SPAHR documentation is contained in the following five volumes:

Volume I	Introduction to the SPAHR Demographic Model for Health Risk J. J. Collins, R. T. Lundy, D. Grahn, and M. E. Ginevan
Volume II	SPAHR Introductory Guide J. J. Collins and R. T. Lundy
Volume III	SPAHR Interactive Package Guide J. J. Collins
Volume IV	SPAHR User's Guide J. J. Collins and R. T. Lundy
Volume V	SPAHR Programmer's Guide J. J. Collins and R. T. Lundy

PROJECTION MODELS FOR HEALTH EFFECTS ASSESSMENT IN POPULATIONS EXPOSED TO RADIOACTIVE AND NONRADIOACTIVE POLLUTANTS

ABSTRACT

The Simulation Package for the Analysis of Health Risk (SPAHR) is a computer software package based upon a demographic model for health risk projections. The model extends several health risk projection models by making realistic assumptions about the population at risk, and thus represents a distinct improvement over previous models. Complete documentation for use of SPAHR is contained in this five-volume publication. The demographic model in SPAHR estimates population response to environmental toxic exposures. Latency of response, changing dose level over time, competing risks from other causes of death, and population structure can be incorporated into SPAHR to project health risks. Risks are measured by morbid years, number of deaths, and loss of life expectancy. Comparisons of estimates of excess deaths demonstrate that previous health risk projection models may have underestimated excess deaths by a factor of from 2 to 10, depending on the pollutant and the exposure scenario. The software supporting the use of the demographic model is designed to be user oriented. Complex risk projections are made by responding to a series of prompts generated by the package. The flexibility and ease of use of SPAHR make it an important contribution to existing models and software packages.

FIN #	Title
A2059	Projection models for health effects assessment in populations exposed to radioactive and nonradioactive pollutants

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

TABLE OF CONTENTS, VOLUME IV

TABLE OF CONTENTS

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Prediction of the health consequences to the general population of exposure to airborne and waterborne pollutants is becoming an important feature of environmental impact analyses. Such prediction requires not only knowledge of the dose term and the dose-response function, but also a model for projecting the health risk to some future population. Health risk projections entail considerable uncertainty about the measurement of the dosage that individuals receive and about the magnitude and nature of the biological response at a given population exposure. The uncertainties regarding the individual dose and the dose-response function have received much attention, but the uncertainty associated with the health risk projection model itself has not been fully addressed.

The purpose of this publication is threefold. First, the uncertainties in various health risk projection models will be addressed, and the assumptions inherent in each model will be stated explicitly. Second, a new model that is an extension of earlier models will be introduced. It is argued that this new model, referred to as the demographic model, is superior to previous models because it makes fewer assumptions about the population at risk and the potential of the population to change over time. Third, a computer package referred to as the Simulation Package for Analysis of Health Risk (SPAHR) is presented which facilitates the application of this model for various pollutants and populations at risk.

The core of any risk assessment scheme is the exposure-response model. This is the quantitative relationship between the level of exposure to the hazard of interest and the deleterious effects resulting from that hazard. If the population exposed to the hazard is homogeneous with respect to its likelihood of suffering ill effects from the exposure, estimation of effects is straightforward; we need know only the total number of persons exposed to estimate the effects. However, if the population is heterogeneous (i.e., different persons have differing risks of suffering health effects from exposure to the hazard), then a reasonable assessment of population risk depends upon the distribution of persons by level of risk.

Research indicates that risk levels are often related to the age and sex characteristics of the exposed population. This is true for both radiation and air pollution exposures. When the risk level is a predictable function of age and sex or some other traceable component of the demographic structure of the population, the estimation of projected health effects becomes less straightforward. If one adds to this complexity the long latency periods between exposure and response, the competing risks from other causes of mortality, and the changing demographic structure of the population over time, the projection of health effects becomes even more complex.

Evaluation of the health consequences for populations exposed to pollutants has become an important issue because of the increasing number of known

or suspected carcinogens in the environment. To date, three projection methods have been used in health risk assessments: the single coefficient model, the multi-coefficient model, and the life table model. Each has its own shortcomings, as discussed in Volume I, Chapter 2. This document presents a fourth model that is more useful and realistic than the previous models because it incorporates age, fertility, and mortality structure, and can follow populations through time under changing levels of mortality, fertility, and pollution exposure. This model is referred to as the demographic model.

A sensitivity analysis of the demographic model indicates that population structure alone for a 100-year exposure to 1 rem may introduce more than a factor of 10 variation in the number of excess deaths. This finding substantiates the premise that the population structure may be more important in a health risk projection than the uncertainty inherent in the dose-response functions.

A comparison of the demographic model with the single coefficient model, the most widely used in health risk projections, is presented in Volume I, Chapter 7. It is concluded that the single coefficient model, even in a short-term projection, may seriously underestimate excess deaths since it is unable to accumulate exposure. For instance, comparison of the single coefficient model with the demographic model for continuous exposure to 0.87 ppb of benzene for 50 years yields widely different estimates of excess mortality. The single coefficient model estimates 2,250 deaths, while the demographic model estimates values from 6,386 to 17,568. In the years 2015-2020, the excess leukemia deaths projected by the demographic model are ten times as large as those of the single coefficient model.

The demographic model is also compared with the life table model used in the 1980 BEIR report to estimate excess cancer deaths from exposure to ionizing radiation. The life table model correctly estimates the increased individual probability of death associated with a given radiation scenario. However, the life table model yields misleading results in the estimation of excess deaths for a specific population. The results presented in the 1980 BEIR report underestimate excess deaths by 50% in some instances. For example, using the linear-quadratic, absolute risk model for a continuous exposure of 1 rad per year for 70 years, the life table model estimates 2459 excess male deaths per million while the demographic model estimates 3769 excess male deaths per million.

This document is divided into five volumes:

I. Introduction to the SPAHR Demographic Model for Health Risk

II. SPAHR Introductory Guide

III. SPAHR Interactive Package Guide

IV. SPAHR User's Guide

V. SPAHR Programmer's Guide

The first volume presents the theory behind the SPAHR health risk projection model and several applications of the model to actual pollution episodes. The elements required for an effective health risk projection model are specified, and the models that have been used to date in health risk projections are outlined. These are compared with the demographic model, whose formulation is described in detail. Examples of the application of air pollution and radiation dose-response functions are included in order to demonstrate the estimation of future mortality and morbidity levels and the range of variation in excess deaths that occurs when population structure is changed. Volumes II through V provide the potential user with detailed guidance and appropriate examples to aid in the interpretation of numerical demographic output from the application of the model to realistic circumstances.

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

1.0 INTRODUCTION

1.1 Purpose and Scope of this Volume

This volume is the fourth in a series of five that document the demographic health risk projection model and the accompanying computer software, the Simulation Package for the Analysis of Health Risk (SPAHR).

This volume gives the more advanced user of the SPAHR computer package the information required to create "tailor-made" programs for addressing specific issues not covered by the three interactive packages. It assumes that the user is familiar with the concepts and terms relating to demography and health risk assessment. These topics are covered in detail in Volume I of this series.

1.2 Overview

The Simulation Package for Analysis of Health Risk (SPAHR) is a computer program developed to permit the evaluation of health impacts based on the simultaneous projection of the altered level of risk associated with a particular pollution scenario and the population exposed to that risk. Given an initial population and a set of vital rates, the observable health impact of a given exposure history can be estimated over time using a variety of measurements including:

1) Excess deaths from appropriate causes, defined as deaths occurring in excess of the expected number per unit time.

2) Life shortening and other actuarial measures of individual risk.

3) Excess incidence (as opposed to deaths) from the various causes, taking into account the recovery rates defined in the document Cancer Patient Survival (NCI 1976).

4) Alterations in the age structure of the exposed population.

The SPAHR program permits a number of options and variations in the projection procedure:

1) Projections may be either deterministic or Monte Carlo. A deterministic projection generates expected values for all quantities, while a Monte Carlo projection generates a large number of randomly perturbed projections and prints out the means and standard deviations of items of interest.

2) Projections may be performed in either period or cohort mode. A period mode projection estimates the age distribution and related parameters for 5-year intervals over the course of the projection and estimates values such as total deaths, deaths by cause, births, and life expectancies for the intermediate intervals. A cohort mode projection takes a selected age group in

the initial population and follows its survival from the current year forward, estimating the same parameters as does a period projection.

3) Projections of the change in mortality risks over time, independent of the population, may also be generated. When this is done, detailed multiple decrement life table analyses based on the dose response model for the effluent of interest are printed out in place of the population analysis. This option allows the user to make detailed studies of risk to individuals and emulates the procedure developed by Cook et al. (1978). However, the Cook code is limited to life table manipulations. While it does correctly address competing risk issues for individuals, it neither manipulates changes in mortality resulting from outside factors nor deals with problems of population dynamics.

4) The dose response model, relating increased risk of death to exposure, is of key importance in this projection scheme. Two basic models with several variations are provided. Air pollutants are treated with a cigarette-based dose response function adapted to several studies of air pollution health effects (Lundy and Grahn, 1977). Radiation effects are treated by implementation of the BEIR I and the BEIR III radiation risk models (NAS, 1972; NAS, 1980). Relative and absolute risk variations are available for both models. Because no predefined set of models can cover all cases, a facility is included that permits the user to specify his own set of coefficients and parameters for each model in place of those included in the original code. Because background vital rates fluctuate, a facility is provided to allow the baseline birth and death rates to be interpolated between any values over any time interval.

With these options, the SPAHR program can be used for a variety of purposes. For example:

1) Deterministic period projections may be useful if the effect of variation in characteristics of a given situation is to be analyzed. Such an analysis might assess the effect of different populations on expected results in a surveillance program around a particular facility, or the effects that different facilities might have on a particular population.

2) A Monte Carlo period projection might be appropriate if the upper and lower limits of the effects in a particular scenario are to be estimated.

3) Cohort projections, or period projections of selected population subsets with the births rates set to zero, can be used to derive the expected results in a reanalysis of an existing epidemiological study or to plan a new study. The Monte Carlo option may then be used to estimate the effect of such parameters as population size and dose level on the power and sensitivity of such a study.

4) Projections of life tables instead of populations may be appropriate if the goal of the analysis is to determine the effect of an individual's exposure on possible insurability.

As can be seen from the preceding discussion, the SPAHR program has a relatively large number of features and options, not all of which are likely to be used in any given application, and some of which are mutually incompatible. Therefore, a control system has been developed that would permit a user who is not a programmer to indicate the desired sequence of operations in order to achieve the desired output.

The program control language has free-format, data-directed input as the primary means of communication. However, input in the older, rigid-format tradition of computer programming is also supported. The control system includes facilities for looping many times through a single set of instructions, thereby minimizing the amount of repetitive typing that must be done. Conditional branches may also be made, thus permitting variations in such factors as population size, calculated life expectancies, or specific identification numbers to be programmed into large analyses.

1.3 Command Summary

This section lists the commands available in version 4.1 of SPAHR and gives a brief description of their use. The commands are described in detail in Chapter 4.

Procedure Command Statements

Command[1] Name	Function
OPTIONS	Sets the default values defining files to be used for various purposes, numbers of sex and age groups, etc.
DATA	Enters primary demographic data into the main data blocks.
LIFETAB	Calculates a life table.
MULDEC	Calculates multiple and associated single decrement life tables for several causes of death.
PROJECT	Projects the population forward through time. Has the ability to project death rates as a function of exposure to toxic substances using the ADJUST subcommand.
ANALYSIS	Produces a summary analysis of the mortality and fertility rates in the current population. Indicates long-term growth by several indicators including stable population age structure and mean generation length.
ROOT[2]	Calculates the intrinsic rate of natural increase by three different methods.
ZEROS[2]	Calculates all roots, both real and complex, of the Leslie matrix.
LOTKA[2]	Fits the net maternity function to the normal distribution curve.
STABLE	Performs a stable population analysis based on the current life table and generates model stable populations for a range of intrinsic rates of natural increase.

[1] Command names are defined in Chapter 2.

[2] Adaptation of the Keyfitz-Flieger (1971) portable routine of the same name.

Language Statements	
Command Name	Functional Description
DO	Defines groups of commands that are to be associated with each other for conditional or repetitive execution.
END	Terminates a DO-group or terminates execution of the SPAHR program.
IF	Permits conditional execution of a statement or groups of statements depending on whether certain conditions have been met.
SET	Defines SPAHR variables.
PRINT	Prints out the current values of SPAHR variables.
TITLE	Defines the string of title information to be printed at the top of every page.

2.0 THE SPAHR FILE STRUCTURE

2.1 Overview of the SPAHR File Structure

All instructions to SPAHR are entered in a computer file called the command file. This file contains specific SPAHR procedure statements selected by the user to simulate various assessment models. The particular population for which the analysis is being performed is located in a file referred as the data file. The command file (i.e., the SPAHR procedure statements) and the data file (i.e., the demographic characteristics of the population being analyzed) are defined separately for two reasons.

1) There is a strong logical distinction between the population being projected and the projection itself. The projection assumptions contained in the command file define a specific set of conditions. These conditions, however, must be applied to the age and sex structure of a specific population contained in the data file. The same conditions applied to different populations will yield different results. The user may, therefore, wish to apply the same analysis (contained in the command file) to several different populations (contained in several data files). This can be done by simply switching the data files rather than by reconstructing the command file each time.

2) With these two file structures, the user may easily simulate various mortality scenarios without providing a data file. This is done by using one of the population structures provided by SPAHR. The default population is the U.S. total population in 1969-71. The U.S. Black population in 1970[1] and various populations from other countries are also supplied. These populations have been selected to cover a wide range of age and sex structures and mortality levels.

2.2 The Command File Structure

The SPAHR procedural commands are composed of keywords, delimiters, and sometimes numbers. A command always begins with a keyword and ends with a delimiter. The keywords in SPAHR are straightforward mnemonics. For instance, LIFETAB constructs a life table, while PROJECT performs a population projection. The delimiters are used to divide the commands into sections and to separate individual commands.

---

[1] The U.S. population is supplied in its two major components because in analyzing a heterogeneous population, a pooled analysis generally yields incorrect results. These incorrect results occur because fertility and mortality levels differ markedly between the White and Black populations in the United States. In order to project the total U.S. population, the user should project each population independently using the ACCUM option of the PROJECT command.

Some commands in SPAHR (e.g. PROJECT) contain several subcommands. In addition, a number (i.e., a scalar parameter) or series of numbers (i.e., a data array or arrays) can be included in a command or subcommand. For instance, the command PROJECT with the further specification YEARS = 100 produces a 100-year population projection. Exposition of the commands and associated options is presented in Chapter 4. A command may contain one or all of the following:

1) Subcommands. Subcommands are logically separate parts of commands. They may contain their own groups of scalar parameters and arrays. For example, in the PROJECT command, the scalars and arrays that are an intrinsic part of the projection process make up the main body of the command. These parameters control such factors as the duration of the projection, the size and structure of the population being projected, and whether the projection will be in period or cohort mode.

2) Scalar Parameters. The scalar parameter section follows immediately after a command. It is composed of keywords (e.g., BACKGROUND) or <keyword> = <value> assignments (e.g., REM = 0.1).

3) Arrays. Data arrays consist of a keyword giving the name of the array, followed by a set of numbers. Some data arrays are multidimensional and consist of groups of arrays, each with its own section name preceding the group names (e.g., DOSE LUNG 25*3.2, 15*1.5, 5*1.0 $$ BONE 30*2.0, 15*1.5;).

SPAHR also utilizes four special symbols to separate commands and sections of commands. These four special symbols, referred to as delimiters, are defined as follows:

1) The single semicolon (;) is used to terminate an entire command. SPAHR interprets the single semicolon to mean "the command is now complete; go do it!" (e.g., PROJECT YEARS = 100;).

2) The double semicolon (;;) is used to terminate a subcommand. It is interpreted to mean that the subcommand is complete; however, another subcommand or command level array follows (e.g., PROJECT YEARS = 100 $ ADJUST MODEL = 5;; ADJUST MODEL = 6;).

3) The single dollar sign ($) is used to terminate a complete array or the group of scalar parameters following a command. It is interpreted by SPAHR to mean that the array is complete, but another array or subcommand follows (e.g., PROJECT YEARS = 100 $).

4) The double dollar sign ($$) is used to terminate a section of a multidimensional array (e.g., dose levels for several disease groupings used in an assessment). It is interpreted to mean that this section of the array is complete, but another section follows immediately (e.g., DOSE LEUKEMIA 20*0.5 10*0.2 $$ LUNG 30*0.8;).

The examples in Chapter 2.4 will clarify the use of these delimiters.

2.3 Constructing Command Files

A SPAHR job has three major steps:

1) The Data Entry step defines the base population for the analysis. It is accomplished by using the DATA command.

2) The Preliminary Demographic Analysis step uses the base population defined in the Data Entry step to calculate life tables and other measures required for all subsequent analyses. It uses the LIFETAB and sometimes the MULDEC commands.

3) The actual Population at Risk and Health Effects Projection is the final step. Here data from the first two steps together with exposure data for the hazard(s) of interest are used to generate estimated excess mortality and related measures. It uses the PROJECT command and its associated ADJUST subcommand.

Most of the options and parameters relating to risk assessment are found in the ADJUST subcommand, which adjusts mortality rates as the consequence of exposure to various hazards. Briefly, the ADJUST subcommand has the following features:

1) A variety of models relating mortality and hazard exposure are available through ADJUST. They are selected by using the MODEL parameter. Relative risk versions of the models may be specified by using the REL parameter. The default period of risk (30 years for most cancers except leukemia) can be reset to lifetime risk by using the LPLAT parameter.

2) Exposure to the hazards of interest may be defined in two ways: If the exposure level is constant over the period of exposure, a single dose rate may be entered by assigning the dose rate to a single-valued parameter named REM (or SO_2 or whichever dose keyword is appropriate for the model). On the other hand, if the exposure rate varies over time, an annual exposure for each year is entered by using a separate array named DOSE.

3) New causes of death can be defined or coefficients for causes of death already defined may be changed with the use of the COEFS, LATENT, and PLATEAU arrays.

4) As many as five ADJUST subcommands may be defined for each PROJECT command. Thus the effects of a variety of hazards operating simultaneously may be assessed as competing risks.

5) ADJUST relates causes of death, risk model coefficients, and, in the case of radiation models only, hazard exposures through the use of a cause name of 1 to 8 characters. If we wish to project radiogenic lung cancer using

the BEIR I relative risk model, the ADJUST subcommand will define a new cause of death named LUNG (the name of the new cause will later be changed to make it unique), and exposure levels will be read in from the DOSE array following the keyword LUNG. To change the predefined coefficients of the model relating lung exposure to mortality, we would likewise specify the keyword LUNG within the COEFS array. SPAHR will use the LUNG death rates entered in the DATA step as the basis for its relative risk calculations.

2.4 The Data File

Users may create their own population data sets if the data sets provided in SPAHR are not appropriate, although creating data sets in SPAHR is tedious because of the amount of information required. Errors in transcribing the numbers can occur. Construction of data sets should only be undertaken as a last resort.

The SPAHR program is oriented toward producing analyses of specific populations. A population can be completely characterized for SPAHR by the following parameters:

1) The year associated with the population.

2) The number of people in the population by sex and age group.

3) The number of deaths by sex and age group.

4) The birth rates by age group of mother.

5) The number of deaths by age, sex, and cause of death for each cause of death to be analyzed.

These are referred to as raw demographic data. They are entered into a data file in free format. This means that they can be put anywhere on the card image, and as many card images as needed can be used. Each item must be identified with appropriate keywords and terminated with appropriate delimiters. The keywords are English language mnemonics, and with the exception of DDC (which stands for deaths by cause) their interpretations are clear. POP, for example, refers to the living population, DEATHS indicates deaths from all causes combined, and BIRTHR refers to birth rates. The names for causes of death that follow the DDC keyword are compromises between the names assigned to these causes in BEIR I report (NAS 1972) and the requirement of an IBM computer for names of 8 characters or less. These names are also used when referring to causes of death in the DOSE and COEFS arrays of the ADJUST subcommands, as illustrated in Chapter 4.12.3.

Delimiters are hierarchically defined. The lowest level of delimiter is the double dollar sign ($$), which is used only to indicate the end of a subsection of a multidimensional array. The single dollar sign ($) is the next highest and signifies the end of an entire array. Either may be used to indicate the end of a set of single-valued parameters. The double semicolon (;;) denotes the end of a subcommand. The highest level delimiter is the semicolon (;), which indicates the end of an entire command. If more than one of the above delimiters appears to be appropriate in a particular situation (e.g., when an array is the last item in a command and both the $ and the ; would seem appropriate), the higher level delimiter should be used.

Age groups in SPAHR are defined as follows: 0-1, 1-4, 5-9, 10-14, and so on in 5-year age groups to 85+ for a total of 19 age groups. If more than 19 numbers are placed in an age-specific array, the last one is assumed to be age unknown, and all those (if any) between the 19th and the last are assumed to be 5-year age groups terminating in an open-ended group, which are therefore added into the 19th by SPAHR to form the 85+ age group. A comprehensive discussion of SPAHR data files and their interactions with the DATA command is given in Chapter 4. Read this section carefully after becoming familiar with the basic data set shown in the following example:

```
PARMS DATE=1970 $ NAME UNITED STATES (WHITE) $ POP MALE 1501250 5873083
8633093 9033725 8291270 6940820
 5849792 4925069 4784375 5194497 5257619 4832555 4310921 3647243
 2807974 2107552 1437628 805564 486957 $$
       FEMALE 1433839 5614968 8264333 8647392 8079090 7341007
 5962122 5042368 4936494 5412335 5587023 5169302 4695581 4157467
 3491080 2874531 2114943 1314258 889855 $ DEATHS MALE 31725 4910 4099
4382 12200 13812 9897 9130 12459 21819
 35992 53092 76502 98781 113614 122829 124979 101556 90339 320 $$
     FEMALE 23151 3714 2646 2410 4672 4826 4360 4899 7447 12557
   20873 28920 39009 50841 67187 90091 113145 116567 142201 143 $
BIRTHR 3*0 0.0011 0.0674 0.1506 0.1348 0.0676 0.0287 0.0071 0.0004$
CAUSES LEUKEMIA LUNG STOMACH ALIMENRY PANCREAS BREAST BONE THYROID
  OTHER CANCER ='CANCER'= LEUKEMIA + LUNG + STOMACH +
   ALIMENRY + PANCREAS + BREAST + BONE + THYROID + OTHER $ DDC LEUKEMIA
MALE 25 247 344 229 211 155 120 117 153 182 298 368
       535 695 829 941 928 659 318 89 10 1 0 $$
      FEMALE 36 195 283 163 153 117 82 95 127 166 224 300 369 432
              550 687 761 649 348 109 21 2 0 $$
   LUNG FEMALE 1 2 3 2 3 5 13 47 187 410 885 1358 1645 1648 1477
      1256 1093 711 335 95 18 2 1$$
        MALE 3 2 0 2 7 15 31 100 349 1119 2314 4073 6556 8312 8616
     7359 5136 2417 775 173 20 4 5 $$
    STOMACH FEMALE 0 0 0 0 0 4 5 15 47 83 167 232 375 484 668 823
             985 863 504 145 25 5 0 $$
         MALE 0 0 0 0 3 4 14 24 72 142 266 461 758 1027 1247 1337
              1292 978 450 133 21 0 0 $$
    ALIMENRY FEMALE 7 7 6 6 11 22 40 84 158 372 831 1477 2228 2847
      3714 4379 4607 3728 2080 706 129 11 3 $$
      MALE 7 18 3 2 27 32 51 80 164 403 859 1606 2762 3762 4209
           4431 4167 2908 1433 383 58 6 2 $$
```

```
   PANCREAS MALE 0 1 0 0 1 5 8 7 36 106 218 398 571 820 999 1120
           1193 891 504 138 22 1 1 $$
      FEMALE 0 0 0 0 1 3 5 27 78 153 349 657 1014 1314 1506 1430
    1292 770 350 83 12 0 1 $$
   BREAST FEMALE 0 0 0 0 0 12 92 311 680 1408 2525 3181 3657 3360
      3220 2884 2626 1819 997 354 81 9 0 $$
       MALE 6*0 1 1 1 4 12 24 22 25 39 28 36 24 10 2 2 1 0 $$
  BONE FEMALE 2 6 18 54 57 18 21 5 16 18 25 37 65 59 73 79 73 63
    35 18 0 0 $$
  MALE 0 1 14 54 121 54 15 15 14 19 39 54 70 88 106 90 85 68 28 11
     0 0 0 $$
 THYROID FEMALE 0 0 1 0 3 3 2 4 8 7 18 22 44 74 76 125 117 84 44
     16 1 0 0 $$
     MALE 3*0 1 1 3 6 5 6 12 14 29 42 40 51 44 40 32 10 2 3*0$$
 OTHER FEMALE 31 178 193 191 246 302 426 393 940 1695 3075 4100
         4686 6046 6526 6654 6230 4491 2332 786 7 14 9 $$
     MALE 30 233 347 238 371 601 557 572 771 1358 2312 3621 5374
            7241 8088 8657 8492 3802 1369 188 55 3 1 ;
```

This example is reproduced from Chapter 4.3.3.2, where a detailed description and interpretation of each item may be found.

The data in this example were taken either from the U. S. Census for 1970 or the Vital Statistics of the United States for 1970. When entering the data into a SPAHR data file, remember that SPAHR can only read the first 72 characters (including the blanks) in each card.

The WYLBUR text editor stores files in two ways: compressed and uncompressed (CARD). Unless it is told to store in CARD format, it will use the compressed format. Unfortunately, WYLBUR is the only computer program that can read compressed format. Consequently, when manipulating the data file using WYLBUR, always remember to SAVE the data file using the CARD option !! This is not a concern with the command file, because when this file is run, it is decompressed automatically.

2.5 Other SPAHR Files

In addition to the COMMAND and the DATA files, four other types of files are available in SPAHR. However, for most applications the user need not be acquainted with the others, which are automatically allocated by the computer system. A brief description of each file type and the computer unit designation follows:

Command File

This file contains all SPAHR command statements. It also contains data for a number of special procedures. These statements are described in detail in Chapter 4. Only one command file may be defined in a SPAHR run.

Data File

A data file contains raw demographic data for the population being studied. It may either be free-format or in one of the predefined SPAHR formats described in Chapter 4. More than one data file may be used. By default, SPAHR will assume a single data file on unit 4. Any others must be defined explicitly as required by the operating system.

Log File

The log file, located on unit 6, is a very short output file designed to give the user a fast overview of the course of the program. Typically, each command executed will put out one or two lines on this file to indicate the most interesting summary statistics and whether the command encountered an error. An example of a typical log file is shown in Chapter 5.

Print File(s)

Print files contain the tables that are the usual output of a SPAHR run. They are usually routed to a line printer. Unit 3 is assumed by default for the single predefined print file. Other print files may be allocated on other unused units. Most SPAHR commands have a parameter called PRINT that determines whether output is to be placed on this file and what unit number it will have.

Punch (Card Image) Files

Punch (card image) files permit SPAHR to write out information in a form that other computer programs can read. By default, one such file is defined on unit 7. Others may be defined on unallocated units. Most SPAHR commands will generate some machine-readable output if directed to do so in the PUNCH parameter.

Auxilliary Storage Files

Some routines in SPAHR require more temporary storage space than can be accommodated in the computer's memory. These routines write out intermediate results and data onto a temporary file that is later rewound and reread. In the current version of SPAHR (version 4.1), only one of these files, on unit 1, is in use at any time. Future releases of SPAHR, however, may use more temporary files and consequently reserve other units as well.

Summary of File Unit Allocations

File Number	Allocation
1	Auxilliary Storage File (sometimes called a scratch file).
2	Reserved for future use.
3	Default Print File. Others may be specified on units 8-99 as desired.
4	Default Data File. Others may be specified on other unoccupied units as needed.
5	Command File.
6	Log File.
7	Default Punch File. Others may be specified on unoccupied units.
8-99	Undefined and available for allocation as desired.

3.0 THE SPAHR LANGUAGE STRUCTURE

3.1 Overview of the Language Structure

SPAHR is a collection of more or less independent computer programs tied together with a language processor (a program that reads and interprets the control statements) and a set of data blocks (special sections of computer memory) in which information necessary for the execution of many of the independent programs is stored.

SPAHR is controlled by statements. A statement is a command that tells SPAHR what to do. Statements are written in the SPAHR language, which will be described in detail later. Statements are physically written on input records, which may be either computer cards or lines typed on a terminal.

Two major types of statements occur in SPAHR. The most important of these is the procedure command, which controls the independent program segments in which the user is most interested. The other type of statement is the language statement, which controls such things as repetition of groups of statements.

A series of SPAHR statements is called a program. For most simple jobs, a SPAHR program will need only procedure command statements to do all the work of data input and analysis. All procedure control statements have the same general structure, because each statement invokes a specialized control routine that initializes all necessary variables in the common data blocks and then calls the routines that perform the analysis. All procedure control statements must begin with a command name and end with a semicolon (;) and may span as many records as are needed. The command name is followed by a series of single-valued parameters. Following the parameters may come a set of array data, which consists of array names followed by sets of numbers, or subcommands, which are similar in structure to commands. Schematically, the structure of a procedure command statement (with optional items enclosed in brackets) is:

```
<command name> [<parameters> $] [<array data> $]
               [ <subcommands> ;; ] ;
```

The command name is a string of 1-8 characters that defines the procedure to be called. The parameters are a set of single-valued quantities that can define both single-valued items of raw data and execution specifications for the procedure. The array data are groups of raw data, usually age-specific, which are to be analyzed by the procedure. Subcommands are control statements for auxilliary procedures that are called by the main procedure.

Sections of a SPAHR command (e.g. arrays and subcommands) are set off from each other with delimiters. Delimiters and their use are discussed in Chapter 2.4.

3.2 Command Names

The command name is composed of 1-8 characters and is separated from the rest of the statement by a comma and/or one or more spaces. For example, to perform a projection on a population already defined in the current data block, taking all defaults, one would enter the control statement

PROJECT;

The semicolon (;) signals SPAHR that the command has ended and is now to be executed.

As entered above, the PROJECT procedure will be executed using all of its default values. If the user wishes to substitute other values, it will be necessary to enter parameters, array data, or subcommands as well. If array data or subcommands are to be entered following the command, the command should be delimited by a double dollar sign ($$) or a dollar sign ($), and the semicolon should be deferred until all arrays and subcommands have been entered.

Because a semicolon terminates reading of a command and causes its execution, several commands may be placed on a single line. For example, the sequence

LIFTAB ; LOTKA ; ANALYSIS ; PROJECT ;

will cause the execution of four separate procedures, each with all of its defaults.

3.3 Single-Valued Parameters

Single-valued parameters immediately follow the command name and are separated from it and from each other only by blanks or commas. Parameters consist either of switches or assigned values. A switch option is a word that can stand alone. An assigned value option, on the other hand, is associated with a particular value. It must be followed immediately on the same line by an equal sign (=) and a number. If, for example, the population were to be projected from 1970 to 2020, and pyramids were to be printed out for each projection interval, then the assigned values for START and STOP would be 1970 and 2020 respectively, and the switch name PYRAMID would be written. The statement would then have the form

PROJECT START=1970 STOP=2020 PYRAMID;

A switch is always turned on by its appearance in a command. It is turned off by assigning it a value of zero. The PRINT switch is by default set at on. If in some command the user does not wish to see the normal printed output, he may override the switch by entering PRINT=0 in the command.

Some scalar parameters may be used either as assigned values or as switches. Any parameter referring to a file unit number can be used as a switch. When this is done, the default file unit will be used. PRINT, DETAIL, PUNCH, and PYRAMID all have this property. In the example above, PYRAMID was mentioned as a switch. If the user wishes to place the population pyramids on a file other than the one used for regular printed output, then PYRAMID can be assigned the unit number of the alternate file unit. If we wished to suppress the standard printed output and print the projected pyramids on file number 8, our model PROJECT command would appear as:

```
PROJECT PYRAMID = 8 PRINT=0 ;
```

Delimiters between data items within command statements, unless specified otherwise, consist of a single comma and/or one or more blanks. For example,

```
PROJECT,START=1970,STOP=2020;
PROJECT START=1970 STOP=2020;
PROJECT START=1970, STOP=2020;
```

will all be interpreted identically.

3.4 Array Data

Array data statement sections supply SPAHR procedures with the data they need to perform their assigned tasks. The user may enter data in arrays without considering where the numbers are located on the input record. This is referred to as free-format input. It is a useful feature when only a small amount of information is to be processed. However, many data sets that the user may wish to read are already set up on fixed-format records, with every number having a predetermined location on each record. Therefore, some procedures, notably the DATA command, provide the option of reading in commonly used arrays from fixed-format cards with one of three standard formats. These formats will be described in detail in Chapter 4.

Array data are entered as an array name followed by a sequence of numbers terminated with a dollar sign ($), semicolon(;), or double semicolon (;;). The dollar sign signifies the end of the array, the double semicolon signifies the end of the subcommand, and the semicolon signifies the end of the command. For example, in the DATA command the age-specific birth rates are entered with the keyword BIRTHR. Thus one might use the sequence

```
DATA $$ BIRTHR 0 0 0 1.1E-3 0.0674 0.1506 0.1384
        0.0676 0.0287 0.0071 0.0004 $
```

If several identical values in a row are to be entered, they may be indicated with a repeater that consists of a number indicating the number of times the value is repeated, followed by a star (*), followed by the value. For example, if one enters a series of proportions for females at birth, and wishes this to be the same for the entire age range, then

DATA $$ SEXR 0.4886 0.4886 0.4886 0.4886 0.4886 0.4886
0.4886 0.4886 0.4886 0.4886 0.4886 0.4886 0.4886 0.4886
0.4886 0.4886 0.4886 0.4886 0.4886 $

has the same interpretation as

SEXR 19*0.4886 $

Many arrays encountered in SPAHR are labeled arrays, and, in addition to the array name, a further set of qualifying names must be employed to specify the data properly. All sex-specific arrays, for example, have the additional qualifiers MALE and FEMALE. The subarrays thus defined are terminated with a double dollar sign ($$). Some arrays can have two levels of qualifiers. The DDC (Deaths by Cause) array in the DATA command, for example, is specific for both sex and cause of death and so must be entered as

DATA $$ DDC <cause1> MALE ..... $$ FEMALE .....$$
<cause2> FEMALE ...$$MALE ... $

The primary qualifiers <cause1> and <cause2> always precede the secondary qualifiers MALE and FEMALE, but MALE and FEMALE subarrays beneath <cause1> and <cause2> can be entered in any order, and under the keyword DDC, <cause1> and <cause2> may appear in any order.

Following the lowest-level qualifier is the series of numbers. A double dollar sign ($$) should be used to terminate the series of numbers in a labeled array if more subsections of the same array are to follow. If a dollar sign ($) terminates the entire array, the program will continue reading and will search for a following array of subcommands. Terminating the array with a semicolon (;) will cause the immediate cessation of further data input or subcommand searches and will begin execution of the current command.

Unless otherwise indicated, arrays read in SPAHR are age specific. The 19 standard age groups are defined as 0-1,1-4,5-9, ... , 85+. If more than 19 elements are entered, the last one is taken to refer to the age unknown group, and the 20th through the next-to-last will be added into the 19th. If fewer than 19 are specified, then the unspecified slots will be set to zero.

Simple arithmetic manipulations may be performed while reading labeled arrays. At any qualifier level, qualifiers of the same level may be used in arithmetic expressions as if they were variable names. Addition, subtraction, multiplication and division are permitted. Manipulation with exponents, parenthetical groups, and special functions such as logs or square roots are not available. Several examples illustrate this feature.

To create a population in which the male and female numbers were identical at each age, entering

POPULATION MALE ... $$ FEMALE = MALE $

would be appropriate.

To read in data on deaths by cause and to define one cause to be the sum of several others, the statements

```
DATA $$DDC <cause1> MALE ... $$FEMALE ... $$
     <cause2> MALE ... $$ FEMALE ... $$
      <cause3> = <cause1> + <cause2> $
```

would be appropriate.

When reading in array data, great care must be taken regarding the use of the equal sign after the array name and qualifiers, as some ambiguities are possible. For example, the statement

```
DATA $$ DEATHS MALE 6 $$ ...
```

can be compared with the almost identical statement

```
DATA $$ DEATHS MALE =6 $$ ...
```

In the first case, SPAHR will assume that the first age group has suffered 6 deaths, but that no deaths at all have been noted in any other age group. It is equivalent to having entered

```
DATA $$ DEATH MALE 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 $$ ...
```

In the second case, however, the equal sign (=) indicates that male deaths are set to an arithmetic expression, so that six deaths will be recorded in every single age group. This is equivalent to having entered

```
DATA $$ DEATHS MALE 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 $$ ...
```

The arithmetic operations in SPAHR are NOT performed in the standard FORTRAN sequence. Instead they are performed in the order encountered. Thus

```
A * B + C
```

will not be evaluated the in same way as

```
C + A * B
```

3.5 Subcommands

Some commands may have subcommands. These follow double or single dollar sign delimiters after the command and parameters, and they may have their own set of parameters and array data. A subcommand must be terminated with a double semicolon (;;) if more subcommands or arrays are to follow it, or with a single semicolon (;) if it is the last item in the command. Arrays entered within subcommands are, as indicated earlier, terminated with dollar signs unless they are the last item in the subcommand.

A subcommand can initialize parameters for a command in situations where the logical structure of the parameters in question is conceptually separable from the main body of the command. The PROJECT command, for example, has several subcommands. The LIFETAB subcommand is provided because in RATES mode (see the description of the PROJECT command, Chapter 4.12, for details) the output simulates a situation in which the LIFETAB command is called down upon the projected population at 5-year intervals. Instructions regarding the construction of life tables are logically distinct from those that generate the death rates upon which life tables are based.

On the other hand, the ADJUST subcommand is an integral part of the logic of the PROJECT command in that it provides information concerning the way in which the underlying death rates are to be calculated as a function of exposure to hazards. However, several different, logically independent sets of cause-specific death rates may be generated or otherwise influenced by the operations defined in the ADJUST subcommand. Furthermore, the operations involved may themselves depend on arrays entered by the user. Consequently, isolating the operations involved in a single subcommand that can be repeated several times simplifies the logical structure of the total command.

Arrays and subcommands are processed in the same loop in most command processors. Consequently, arrays and subcommands can be intermixed arbitrarily without harm. However, it will probably be to the user's advantage to group the arrays read in directly from a command ahead of the subcommands, especially if the subcommands themselves have arrays, as in the following example:

```
*Command statement with parameters:
PROJECT RETAINDR START=1970 STOP=2170 STOPBR=2000 DOSINT=6 $
*Data array read by command statement:
BIRTHL 0 0 0 0.00097 0.05923 0.1324 0.1185 0.0594 0.0253
  0.0062 0.0004 $
*First subcommand invocation:
ADJUST MODEL=6 LPLAT START=1970 REM=0.1 PRINT ;;
*Second invocation of subcommand:
ADJUST MODEL=6 REL START=1995 PRINT $
*Data array for subcommand:
COEFS BREAST FEMALE 0 0.0069 0.0069 $$ MALE = FEMALE $
*Second Data array for subcommand:
DOSE LEUKEMIA 20*0.5,100*0.3,500*0.15$$ LUNG 500*0.02,300*0.1667$$
BONE 285*1.12 327*0.815$$ OTHER 76*0.8 277*0.6 $$BREAST=OTHER ;
```

3.6 Comments

A comment record is defined by a star (*) in the first column. Any number of comments may be placed in the program, either within or between the commands. Comment records may also appear in the sets of fixed-format data read in through the DATA command, but only before the first card (the NAME card) or following the last data record. A series of comment records following any data block in fixed format must be terminated by a comment record in which both the first and second columns contain stars, unless the NOCOM option of the DATA command has been specified.

In free-format input files (including the command file) only, a record with C$ in the first two columns will also be interpreted as a comment record. Some FORTRAN processors designed for debugging purposes, such as WATFIV, make special use of C$ cards. In particular, the string C$DATA appears at the beginning of many data decks in WATFIV jobs. When used with one of these debugging processors, SPAHR will not read this record. However, if the same data file is used by a regular compiler, then the comment record will suddenly appear. Thus the SPAHR comment convention makes it possible to use the data file with C$ cards without modification.

3.7 The SPAHR Variables

One may create a variable in SPAHR by assigning it a value. For example,

```
X = 6 ;
```

will create a variable name X with the value 6.0. SPAHR variables can be manipulated arithmetically. The variable Y, for example, can be defined in the statement

```
Y = X * 2.3 + 8 ;
```

A variable name may be any string of 8 characters or fewer that begins with a letter. It may contain no delimiters or blanks.

SPAHR also creates its own system variables in the course of executing some commands. These variables are not important to the proper functioning of any of the SPAHR commands. They are copies of some of the data in the current data block, provided to allow the user to easily control the execution of SPAHR. It is assumed that they will be used in IF statements to make decisions about the desired execution sequence of statements. An example of this use follows in this section. System variables are set by SPAHR as the result of certain commands, but they are not otherwise unique. The user may reset their values or use them in computations, just as if they were ordinary variables. Doing this will not affect SPAHR. The table below shows the names and definitions of the system variables defined in version 3.4 and the names of the commands that initialize them:

System Variable Name	Initializing Command	Definition
TP	DATA	Total population size.
TPF	DATA	Total female population.
TPM	DATA	Total male population.
TB	DATA	Total births.
TBF	DATA	Total female births.
TBM	DATA	Total male births.
TD	DATA	Total deaths.
TDF	DATA	Total female deaths.
TDM	DATA	Total male deaths.
ID1	DATA	Primary identification code if present in fixed-format input.
ID2	DATA	Secondary identification code.
DATE	DATA	Date associated with the current population.
NRR	ANALYSIS	Net Reproduction Rate.
GRR	ANALYSIS	Gross Reproduction Rate.
TFR	ANALYSIS	Total Fertility Rate.
GFR	ANALYSIS	General Fertility Rate.
EOF	LIFETAB	Female expectation of life.
EOM	LIFETAB	Male expectation of life.

In most cases described in Chapter 3.3 in which a numeric constant is used, a user-defined or system variable may be substituted. For example, the command statement

```
DATA INPUT=9 PYRAMID=8 FORMAT=3;
```

might also be entered as

```
IN=9 ; PR=8 ; FORM=3 ;
DATA INPUT=IN  PYRAMID = PR  FORMAT= FORM ;
```

Consider the sequence of statements

```
MALE=9 ;
DATA $$ DEATHS MALE 10 5 3 5 12 16 24 48 ... $$
            FEMALE=MALE ;
```

The word MALE in this example has been used in two separate contexts. It has been defined as a SPAHR variable name, but it is also the name of a qualifier for the array DEATHS. In the event of a name conflict such as this, SPAHR will give priority to the qualifier name over the user-defined variable name. Thus this example is equivalent to

```
DATA $$ DEATHS MALE 10 5 3 5 12 16 24 48 ... $$
          FEMALE 10 5 3 5 12 16 24 48 ... ;
```

rather than

```
DATA $$ DEATHS MALE 10 5 3 5 12 16 24 48 ... $$
        FEMALE 9 9 9 9 9 9 9 9 9 9 ... ;
```

3.8 Advanced Language Features

The SPAHR language permits the user to write a SPAHR program almost as if it were a higher level language such as FORTRAN or PL/I. One may, for example, define conditional execution of statements or groups of statements, or repeat groups of statements. For example, we may have a collection of 100 sets of demographic data on a disk file in fixed format. If we have reason to believe that some of the data is of dubious quality, we may want to analyze only those sets of data in which the population is large enough so that the data are likely to be useful. This series of statements will select the desired sets.

```
DO N = 1 TO 100 ;
DATA INPUT=9 FORMAT=1 ;
IF TP GT 5000 THEN DO ;
  LIFETAB ;
  LOTKA ;
  ZEROS ; ANALYSIS ;
  END ;
  ELSE PRINT "POPULATION OF" TP 'IS TOO SMALL TO USE' ;
END ;
END ;
```

The above program reads through the first one hundred sets of data. If the total population (system variable TP) is greater than five thousand, then the data are assumed to be valid, and the statements in the DO group following the IF statement are executed. Otherwise, an informative message is printed out on the log file advising that the population is too small and giving its exact size. (The DATA command causes the name of the population and other identifying data to print out on the log file as well.) Three END statements appear. The first of these closes the conditional DO group. The second one closes the repeating DO group. The third one ends the SPAHR program.

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

4.0 THE SPAHR COMMAND REFERENCE

4.1 Overview

This section describes in detail the various command statements that are implemented in version 4.1 of SPAHR.

For demographic procedures, the first section describes in general terms the function of the command. The general description is followed by a detailed breakdown of the single-valued parameters and switches, array items, and sub-commands. Following these in many cases is a description of the data required.

Each language statement is described with examples in the way best suited to its characteristics.

To use the command statements the user should refer to Volume II, the SPAHR Introductory Guide, where specific examples of many of the commands are provided. The more advanced user, however, may wish to try the many options that are available for each command and thus become familiar with the versatility of SPAHR.

4.2 OPTIONS (Single-Valued Parameters)

The OPTIONS command sets default I/O unit numbers and execution modes for the single-valued parameters listed below.

PRINT = File unit number of the FORTRAN file to contain the standard printed output for each future command. Default = 3.

DETAIL = File unit number of the file which will contain the detailed printed output for each succeeding command. Default = 0.

BATCH/TSO = Execution mode switch.

Specifying BATCH forces the batch mode with its full echo of input. Specifying TSO (the default) forces the interactive mode, which suppresses echo of terminal input and causes SPAHR to issue prompts when input is desired.

In batch mode, a set of input records defining the complete job is entered into the computer system, which schedules the job to be executed at some future time at its own convenience. In interactive mode, the SPAHR program is executed while the user communicates with it at an online terminal. Commands are processed as they are entered at the terminal.

TERMFILE = File number for SPAHR system summary log and terminal messages. Default = 6.

NA = Number of age groups. Default = maximum = 19.

NS = Number of sex groups (1 or 2; Default = 2).

PUNCH = File unit number for card-image output. Default = 7.

MSGDO = DO level beyond which selected log messages such as lines indicating the pages printed on output files will not be printed. Default = 0, which means that the print advisories will appear only for commands issued outside of any DO loops.

For example,

OPTIONS BATCH DETAIL = 8;

tells SPAHR that execution will be in noninteractive mode, and that full details will be printed for all subsequent commands. It is assumed that unit 8 has previously been allocated as a print file.

4.3 DATA

The DATA command has as its primary function the input and consistency analysis of raw demographic data. Raw demographic data must be entered so that the other procedures in SPAHR will have enough information to function. The PROJECT command, for example, cannot function without a population to project, and the LIFETAB command needs enough information to calculate age-specific death rates so that it can calculate a life table.

Data may be entered either in free-format mode by the use of the named arrays described in this section or by reading in fixed-format data cards using one of three predefined formats. In addition, data may be entered either in the command file (the one in which the DATA command appears, usually unit 5) or a data file. If the data are not in the command file or on the default data file, the INPUT parameter must be used.

4.3.1 Single-Valued Parameters

Command Control Parameters

PRINT = File unit number for printed output.

PUNCH = File unit number for punched output. When this parameter is coded, a data deck will be written out in the SPAHR standard format, with a dummy comment record at the end consisting of the asterisks in the first two columns.

TERM = Log file information level. With TERM=0, nothing will be printed on the log file. With TERM=1 (the default) the NAME will be printed on the log file, along with any inconsistencies detected in the input data. At this level too, comment lines in fixed-format input files preceding the NAME card will be printed out on the log file. TERM=2 or greater will cause printing of more detailed messages that are primarily of interest to persons trying to find program bugs.

PYRAMID= File unit on which population pyramid is to be printed. If used as a switch, unit specified by PRINT = is assumed. If PYRAMID = 0, unit 6 is assumed.

PR1 Directs printing of data before consistency analysis and redistribution of unknowns.

PR2 Directs printing of data following checks and redistribution.

INPUT = File unit number on which DATA will find basic demographic data. The format of this data (if fixed format) will be specified in the FORMAT parameter. If fixed-format data are read in, free-format data may be entered via the same DATA command. However, the fixed-format data are read in first, so that free-format data will prevail in the event of duplication. Unless the NOCLEAR option has been specified, the default INPUT file number is 4. If the NOCLEAR option has been specified, the INPUT parameter cannot be used. See the description of the NOCLEAR option below for further details.

FORMAT = Fixed-format input definition key. Fixed-format data are restricted to one of the standard formats below.

=1: input in SPAHR standard form.
=2: input in Keyfitz standard form.
=3: input in Preston standard form.

REWIND Causes the file referenced by the INPUT parameter to be rewound before being read. This option is ignored if the input data are on the command file.

NOCOM If mentioned, tells DATA that the coded input file referred to in the INPUT parameter contains comment records. Useful for all formats, but vital for FORMAT = 2.

INCREM = Number of columns per page allowed for cause-specific tables printed in the DATA command.

CLEAR Specifies that cause-of-death definitions and other information normally retained between data blocks will be cleared out (i.e. set to zeros or blanks) before the current data set is processed.

NOCLEAR Specifies that the current data block is not to be cleared out before new data is added as a result of the current call. This option has the effect of causing SPAHR to act as if the data in the current data block had just been read in from a fixed-format file. The data that are subsequently read in free format from the command file are used to override selectively only those items specified. When NOCLEAR is specified, data cannot be read in from an external data file.

NKOS = number of cause-of-death groups to be expected in the data file.

Single-Valued Demographic Data Items

Single-valued demographic data items may be entered in two places: immediately following the command name DATA, as with all other SPAHR commands, or following the array name PARMS.

Except for DATE, the items in this group are all computed by DATA from the appropriate age-specific data arrays. If entered redundantly here, they will be checked against the internally computed values, and any discrepancies will be noted on the log file.

DATE = date to be associated with the data being entered. (Default=1975)

TP = total population

TPM = total male population.

TPF = total female population.

TB = total births.

TBM = total male births.

TBF = total female births.

TD = total deaths.

TDM = total deaths among males.

TDF = total deaths among females.

4.3.2 Array Data

Definition Arrays

PARMS The set of single-valued demographic data items described in the previous section.

NAME 60-character string defining the name to be associated with the population. This may include a SPAHR variable expansion as noted in the TITLE command.

CAUSES 8-character cause-of-death names. No blanks or other delimiting characters may occur within the string. The first character must be a letter or an underscore. Each name of 1-8 characters may be set equal to a string of up to 40 characters enclosed in quotation marks which will be used for labeling purposes on some tables and graphs. Causes may also be defined in this section as sums of other causes. If the cause thus defined is initialized as part of the DDC array, it will be checked for internal consistency. Because the CAUSES array is handled in an anomalous manner, three examples are presented here.

1) Example of simple names:

```
CAUSES TB FLU PULMONRY CARDIO $
```

2) Example of simple names with expanded definitions:

```
CAUSES TB = 'Pulmonary Tuberculosis',
       FLU='Influenza', PULMONARY='Pulmonary diseases',
       CARDIO='Cardiovascular Disease' $
```

3) Example of simple names with extended labels defining one cause as a supergroup of other causes:

```
CAUSES TB='Pulmonary Tuberculosis'
       FLU = 'Influenza'
       PULMONRY = 'Pulmonary Diseases' = TB + FLU
       CARDIO = 'Cardiovascular Disease' $
```

Age-specific Arrays - No Qualifiers

SEXR proportion of births that are female by age of mother.
BIRTHT number of births by age of mother.
BIRTHR age-specific birth rates by age of mother.

Age- and Sex-specific Arrays
Primary Qualifiers MALE and FEMALE

POP population by age and sex.
DEATHS number of deaths by age and sex.
BIRTHS number of births by age of mother and sex of child.
DEATHR age-specific death rates by sex.
PPCT proportional population age distribution.
PPST stable age distribution.
FFX top row of Leslie projection matrix.
SX survival terms of Leslie projection matrix.
VL life table survivors from birth to age x (small ℓ_x column of the life table).
VLL life table ${}_nL_x$ column stationary age distribution.
VMX life table central death rates.
QX age-specific probabilities of death.

Age-, Sex-, and Cause-Specific Arrays
Primary Qualifiers are Cause Names
Secondary Qualifiers MALE and FEMALE

These arrays occupy the same space in internal storage. Consequently, only one may be specified in any DATA command. The cause names used as the secondary qualifiers may be defined in two ways. They may be set using the CAUSES array (see above), or, if this is the first mention of causes of death in the SPAHR job or the CLEAR option has been specified, the cause names specified will be used to define the current cause names. In version 4.1, no more than 36 causes of death may be defined.

DRC age-specific death rates by sex and cause of death.
DDC deaths by age, sex, and cause.

4.3.3 SPAHR Data Files

The principles outlined in Chapter 4 may be illustrated by considering some examples of SPAHR data files.

4.3.3.1 Free-format Data Files: The Minimum Data Set

SPAHR free-format data files consist of information written in a computer dataset in descriptive form. The minimum number of data items that must be supplied before SPAHR jobs will run are:

1) The date (year) associated with the population.
2) The name associated with the population.
3) The number of people in each age group of the population by sex. The only age groups recognized for the present version are 0-1, 1-4, 5-9, 10-14, ... 85+, for a total of 19 age groups.
4) The number of male and female deaths in each age group.
5) The number of male and female births by age of mother in each age group.

The number of deaths and births may be replaced by age-specific rates, defined as the ratio of the number of deaths (or births) in an age group to the number of people at risk in the same age group. In the case of birth rates the persons at risk for both sexes of children are females, and thus the minimum acceptable data set would look like Example Data Set 1.

Example Data Set 1

```
PARMS DATE=1970 $
NAME UNITED STATES (WHITE) 1970 $
POP MALE 1501250 5873083 8633093 9033725 8291270 6940820
 5849792 4925069 4784375 5194497 5257619 4832555 4310921 3647243
 2807974 2107552 1437628 805564 486957 $$
       FEMALE 1433839 5614968 8264333 8647392 8079090 7341007
 5962122 5042368 4936494 5412335 5587023 5169302 4695581 4157467
 3491080 2874531 2114943 1314258 889855 $
 DEATHS MALE 31725 4910 4099 4382 12200 13812 9897 9130 12459 21819
 35992 53092 76502 98781 113614 122829 124979 101556 90339 320 $$
     FEMALE 23151 3714 2646 2410 4672 4826 4360 4899 7447 12557
 20873 28920 39009 50841 67187 90091 113145 116567 142201 143 $
 BIRTHR 3*0 0.0011 0.0674 0.1506 0.1348 0.0676 0.0287 0.0071 0.0004 ;
```

This data set consists, for the most part, of data items that begin with one or more keywords and end with a delimiter. Delimiters are discussed fully in Section 2.4, while keywords are described in Section 4.3.2.

A review of Example Data Set 1 follows. The first keyword encountered, PARMS, tells the DATA command that single-valued data items follow. In this case, the parameter DATE is assigned the value 1970, and all calculations using the data for this population will associate the date 1970 with it. The $ delimiter tells SPAHR that no more single-valued data items are to be read in and to search for another keyword.

The next keyword, NAME, indicates that the following item is the name associated with this population. SPAHR will place everything between the next nonblank character and the first delimiter ($$, $, or ;) into the character string that is printed out at the head of most tables generated in subsequent steps. In this case, a dollar sign is used to terminate the name string and tell SPAHR to search for another keyword.

The numbers of males in each age group of the population in our example are preceded by the keywords POP (indicating population) and MALE (indicating that the numbers refer to men rather than women). These two keywords (which must appear in this order) are followed by the number of males in the 0-1 age group (1,501,250, according to the 1970 census), then by the number of males in the 1-4 age group (5,873,083), and so on to the 19th number (486,957 men over the age of 85). This final number is followed by a $$, indicating that although no more numbers are in this subset of the POP array, more data will follow. The keyword FEMALE, which follows immediately after the $$, informs SPAHR that the numbers to follow refer to females, and since the numbers follow a $$ they are also part of the POP array. The data for females, however, are terminated by a dollar sign, indicating that no more data are in this particular array. The next keyword is DEATHS. The array that follows is entered exactly as is the POP array, except for the 20th number in each section, which represents the age unknown group (explained in the discussion of Example Data Set 2 below).

Following the DEATHS data comes the BIRTHR keyword, which refers to rates rather than to actual births. This particular keyword does not require that MALE and FEMALE qualifiers be used. It is safe to assume that male births will be a little more than half of the total, while female births will number a little less than half. SPAHR will combine the birth rates for both sexes and divide them accordingly. The BIRTHR array is terminated by a semicolon, which signals SPAHR that no more data are in this set.

4.3.3.2 Free Format: A Larger Data Set

The DATA command has a number of data manipulation enhancements that make data entry easier. Inconsistencies in data may be discovered, and data in age groups that do not fit into the SPAHR format may be entered without modification.

Totals for the population, deaths, and births may be entered independently of the POP, DEATHS, and BIRTHS arrays using PARMS options TP, TD, and TB, respectively. Totals specific to each sex may be entered by appending an M or an F to these keywords (e.g. TDM refers to the Total Deaths for Males). The totals thus entered are checked against the totals calculated from the age-specific arrays, and if a discrepancy is found, an informative message is printed out.

The age groups in the data source may not match those defined by SPAHR (0-1, 1-4, and 5-year age groups through 85+). National Center for Health Statistics (NCHS) data, for example, are often given in a series of age groups terminating in the 100+ age group. Occasionally official statistics are associated with no known age. Such situations are met by SPAHR in the following way:

1) Age groups from the 19th through the next-to-last are added into the 85+ age group.

2) If more than 19 age groups are entered, the last is assumed to be the age unknown category. The members of this category are distributed among the known age groups in proportion to their size. Thus the distribution by age is preserved exactly. However, fractional deaths, births, and people may occur in the data set when the DATA routines are finished.

These two features are illustrated in Example Data Set 2, which is identical to Example Data Set 1 except for the addition of some totals in the PARMS field and the inclusion of data for deaths from the cancer types treated in the Reactor Safety Study (USNRC, 1976) and taken from the Vital Statistics of the United States for 1970.

Example Data Set 2

```
PARMS DATE=1970 TP=177748704 $ TDM=739659 TDF=942437 $
NAME UNITED STATES (WHITE) 1970 $
POP MALE 1501250 5873083 8633093 9033725 8291270 6940820
 5849792 4925069 4784375 5194497 5257619 4832555 4310921 3647243
 2807974 2107552 1437628 805564 486957 $$
       FEMALE 1433839 5614968 8264333 8647392 8079090 7341007
 5962122 5042368 4936494 5412335 5587023 5169302 4695581 4157467
 3491080 2874531 2114943 1314258 889855 $
DEATHS MALE 31725 4910 4099 4382 12200 13812 9897 9130 12459 21819
 35992 53092 76502 98781 113614 122829 124979 101556 90339 320 $$
     FEMALE 23151 3714 2646 2410 4672 4826 4360 4899 7447 12557
   20873 28920 39009 50841 67187 90091 113145 116567 142201 143 $
BIRTHR 3*0 0.0011 0.0674 0.1506 0.1348 0.0676 0.0287 0.0071 0.0004$
CAUSES LEUKEMIA LUNG STOMACH ALIMENRY PANCREAS BREAST BONE THYROID
  OTHER CANCER ='CANCER'= LEUKEMIA + LUNG + STOMACH +
   ALIMENRY + PANCREAS + BREAST + BONE + THYROID + OTHER $
DDC LEUKEMIA MALE 25 247 344 229 211 155 120 117 153 182 298 368
       535 695 829 941 928 659 318 89 10 1 0 $$
      FEMALE 36 195 283 163 153 117 82 95 127 166 224 300 369 432
              550 687 761 649 348 109 21 2 0 $$
    LUNG FEMALE 1 2 3 2 3 5 13 47 187 410 885 1358 1645 1648 1477
       1256 1093 711 335 95 18 2 1$$
         MALE 3 2 0 2 7 15 31 100 349 1119 2314 4073 6556 8312 8616
      7359 5136 2417 775 173 20 4 5 $$
     STOMACH FEMALE 0 0 0 0 0 4 5 15 47 83 167 232 375 484 668 823
              985 863 504 145 25 5 0 $$
          MALE 0 0 0 0 3 4 14 24 72 142 266 461 758 1027 1247 1337
                1292 978 450 133 21 0 0 $$
      ALIMENRY FEMALE 7 7 6 6 11 22 40 84 158 372 831 1477 2228 2847
        3714 4379 4607 3728 2080 706 129 11 3 $$
        MALE 7 18 3 2 27 32 51 80 164 403 859 1606 2762 3762 4209
             4431 4167 2908 1433 383 58 6 2 $$
     PANCREAS MALE 0 1 0 0 1 5 8 7 36 106 218 398 571 820 999 1120
               1193 891 504 138 22 1 1 $$
        FEMALE 0 0 0 0 1 3 5 27 78 153 349 657 1014 1314 1506 1430
      1292 770 350 83 12 0 1 $$
     BREAST FEMALE 0 0 0 0 0 12 92 311 680 1408 2525 3181 3657 3360
        3220 2884 2626 1819 997 354 81 9 0 $$
        MALE 6*0 1 1 1 4 12 24 22 25 39 28 36 24 10 2 2 1 0 $$
    BONE FEMALE 2 6 18 54 57 18 21 5 16 18 25 37 65 59 73 79 73 63
      35 18 0 0 $$
    MALE 0 1 14 54 121 54 15 15 14 19 39 54 70 88 106 90 85 68 28 11
       0 0 0 $$
   THYROID FEMALE 0 0 1 0 3 3 2 4 8 7 18 22 44 74 76 125 117 84 44
       16 1 0 0 $$
       MALE 3*0 1 1 3 6 5 6 12 14 29 42 40 51 44 40 32 10 2 3*0$$
   OTHER FEMALE 31 178 193 191 246 302 426 393 940 1695 3075 4100
             4686 6046 6526 6654 6230 4491 2332 786 7 14 9 $$
       MALE 30 233 347 238 371 601 557 572 771 1358 2312 3621 5374
               7241 8088 8657 8492 3802 1369 188 55 3 1 ;
```

As in Example Data Set 1, the first keyword is PARMS. DATE is once more set to 1970. In addition, we now specify that the total population (TP) should be checked against 177,748,704; that the total male deaths should number 739,659; and that the total female deaths should be 942,437. SPAHR will check the totals in the POP and DEATHS arrays against these numbers and notify us if they do not correspond. The rest of this example is identical up through the BIRTHR array, which is now terminated by a $. This signals SPAHR that another array is to be anticipated, and the keyword CAUSES appears on the following line. This is an array of cause names that tells SPAHR which causes of death are to be entered, and to define relationships between some of the causes. In particular, it defines a cause named 'CANCER' as the sum of all of the rest of the defined causes. This array is likewise terminated with a dollar sign, to indicate that more data follow.

The keyword DDC is next. This mnemonic represents Deaths by Cause. The next keyword is a particular cause of death, in this case LEUKEMIA. The additional keyword MALE that follows is a secondary qualifier denoting the sex of those among whom the leukemia deaths are occurring. Following immediately is the number 25, the number of male infants less than 1 year of age who died of leukemia during 1970. The 19th number is 318. However, that is not the last number in the set, because our NCHS tables have defined more age groups. The 318, therefore, is not the number of deaths in the 85+ age category, but is instead the number in the 85-89 category. The 89 that follows is the number in the 90-94 age group that died; 10 men between the ages of 95 and 99 died of leukemia, and 1 person over 100 years of age died of leukemia. The 0 that follows is necessary to tell SPAHR that no one in the <u>age unknown</u> category died of leukemia. Otherwise, the 1 would be assumed to be a person of unknown age, and the 10 people would be assumed to be in the 95+ age category. A $$ delimiter follows, and SPAHR calculates the 85+ age group as 418 = 318 + 89 + 10 + 1. After the $$ is the keyword FEMALE. The next segment also ends with a $$, indicating that more data follows in the current array.

If the next keyword entered was either MALE or FEMALE, then the data for that sex would be entered again. The user would do this if he was running SPAHR interactively and had made an error in the previous few lines. However, because the data for leukemia entered to this point are correct, we have entered instead a keyword giving the name of the next cause of death (LUNG). When SPAHR sees that the keyword LUNG is not a secondary qualifier, it steps up a level and assumes that it is a primary qualifier.

The keyword LUNG is followed by FEMALE instead of MALE, unlike all previous arrays in which the data for males were entered first. Within any level, the order in which array segments are entered may be altered.

In this example the cause names were defined in the course of entering the DDC array. Cause names defined in this way are fixed in the internal data blocks for the rest of the SPAHR run, so that when subsequent data files are read in, the names of the causes must be the same as those defined in the first file. The order in which the cause groups are entered in the later files does not matter.

To define the causes in a different order than they are entered in Example Data Set 2, which is necessary in some of the mortality adjustment routines in the PROJECT command, or to associate names of more than 8 characters with the causes of death when certain tables are printed in other steps, the CAUSE array must be used. The CAUSE array was described in detail in Chapter 4.3.2.

4.3.3.3 Fixed-format Data Files

The DATA command permits the use of fixed-format as well as free-format data through the INPUT and FORMAT options.

Three standard formats are presently available for SPAHR data files. The first is the SPAHR standard, a flexible format designed to store various types of raw demographic data. However, several other standardized formats for demographic data are already in wide use, and two of them are incorporated in SPAHR to minimize the effort involved in initial use. Therefore, the second SPAHR format is derived from the standard employed by Keyfitz in his set of portable programs for population analysis (Keyfitz and Flieger, 1971). The third SPAHR format is that employed by Preston, Keyfitz, and Schoen in the original data base for their *Causes of Death* study (1972). The programs used in their study were never organized for public distribution. However, the data base for their study implemented a format that has been widely used.

Comments may be inserted in data decks immediately after the last card in the set and before the NAME card for the next set. Comment cards are denoted by a star (*) in column 1. The last comment card in the group is denoted by stars in columns 1 and 2. If the data deck does not contain comment cards, the NOCOM parameter must be coded in the DATA command. In the Keyfitz standard format, the Keyfitz comment convention may also be employed, but is not recommended.

When the NAME field is entered using one of the standard formats, excess blanks in the field are eliminated. In addition, if an ampersand (&) is entered in the NAME field with a previously defined SPAHR variable name, this & name segment will be replaced by the value of the variable as described for the TITLE command (Chapter 4.13). This happens only in the case of fixed-format input. In free-format NAME strings, the & name segment will be read through intact, with variable substitution delayed until the NAME is printed or punched out.

In the fixed-format as well as the free-format data, the totals of the age-specific data (e.g. population, deaths, and births) are used only for cross-checking the validity of the data. If the spaces for the totals are left blank, they will be calculated from the age-specific data.

4.3.3.4 The SPAHR Standard

Card	Columns	Data Item
1	1-72	NAME.
2	1-4	DATE.
2	5-16	Total Population (TP).
2	17-28	Total Births (TB).
2	29-40	Total Deaths (TD).
2	53-55	ID1 (For U.S. Data this is the FIPS State code).
2	56-58	ID2 (FIPS County code).
3	1-4	Number of years covered in birth data.
3	5-8	Number of years covered in death data.
3	9-12	Number of causes of death (NKOS). Maximum is 28, default is 0.
3	13-22	Total female population (TPF).
3	23-32	Total female births (TBF).
3	33-42	Total female deaths (TDF).
3	43-52	Total male population (TPM).
3	53-62	Total male births (TBM).
3	63-72	Total male deaths (TDM).
4-5	(10F8.0)	Female population by age.
6-7	"	Male population by age.
8	(9F8.0)	Female births by age of mother, starting with age group 10-14, ending with 45-49. The ninth field is the age of mother not known category.
9	"	Male births by age of mother, as above.
10-11	(10F8.0)	Female deaths by age.
12-13	"	Male deaths by age.
14	(10F8.0)	Total female deaths for each of the causes listed (may cover more than one card).
15	"	Total male deaths for each cause.
16-17	"	Female deaths by age for first cause.
18-19	"	Male deaths by age for first cause.

(Succeeding blocks of four cards for each cause.)

Cards 14 and beyond are omitted if NKOS is 0.

Example Data Set 3 shows the data from Example Data Set 2 as it would appear in the SPAHR standard format.

Example Data Set 3

UNITED STATES (WHITE) 1970

1970 177748704 2986491 1682096

1 1 10 91027840 1459199 739659 86720864 1527290 942437

1433839	5614968	8264333	8647392	8079090	7341007	5962122	5042368	4936494	5412335
5587023	5169302	4695581	4157467	3491080	2874531	2114943	1314258	889855	0
1501250	5873083	8633093	9033725	8291270	6940820	5849792	4925069	4784375	5194497
5257619	4832555	4310921	3647243	2807974	2107552	1437628	805564	486957	0
4648	266058	540174	392685	166546	69224	18776	1092	0	
4865	278473	565381	411009	174318	72454	19652	1143	0	
23151	3714	2646	2410	4672	4826	4360	4899	7447	12557
20873	28920	39009	50841	67187	90091	113145	116567	142201	143
31725	4910	4099	4382	12200	13812	9897	9130	12459	21819
35992	53092	76502	98781	113614	122829	124979	101556	90339	320
5869	11197	5430	27453	9045	27216	742	649	49550	137151
7454	47388	8229	27373	7040	232	946	338	54281	153281
36	195	283	163	153	117	82	95	127	166
224	300	369	432	550	687	761	649	480	0
25	247	344	229	211	155	120	117	153	182
298	368	535	695	829	941	928	659	418	0
1	2	3	2	3	5	13	47	187	410
885	1358	1645	1648	1477	1256	1093	711	450	1
3	2	0	2	7	15	31	100	349	1119
2314	4073	6556	8312	8616	7359	5136	2417	972	5
0	0	0	0	0	4	5	15	47	83
167	232	375	484	668	823	985	863	679	0
0	0	0	0	3	4	14	24	72	142
266	461	758	1027	1247	1337	1292	978	604	0
7	7	6	6	11	22	40	84	158	372
831	1477	2228	2847	3714	4379	4607	3728	2926	3
7	18	3	2	27	32	51	80	164	403
859	1606	2762	3762	4209	4431	4167	2908	1880	2
0	0	0	0	1	3	5	27	78	153
349	657	1014	1314	1506	1430	1292	770	445	1
0	1	0	0	1	5	8	7	36	106
218	398	571	820	999	1120	1193	891	665	1
0	0	0	0	0	12	92	311	680	1408
2525	3181	3657	3360	3220	2884	2626	1819	1441	0
0	0	0	0	0	0	1	1	1	4
12	24	22	25	39	28	36	24	15	0
2	6	18	54	57	18	21	5	16	18
25	37	65	59	73	79	73	63	53	0
0	1	14	54	121	54	15	15	14	19
39	54	70	88	106	90	85	68	39	0
0	0	1	0	3	3	2	4	8	7
18	22	44	74	76	125	117	84	61	0
0	0	0	1	1	3	6	5	6	12
14	29	42	40	51	44	40	32	12	0
31	178	193	191	246	302	426	392	940	1695
3075	4100	4686	6046	6526	6654	6230	4491	3139	9
30	233	347	238	371	601	557	572	771	1358
2312	3621	5374	7241	8088	8657	8492	3802	1615	1
77	388	504	416	474	486	686	980	2241	4312
8099	11364	14083	16264	17810	18317	17784	13178	9674	14
65	502	708	526	742	869	803	921	1566	3345
6332	10634	16690	22010	24184	24007	21369	11779	6220	9

4.3.3.5 The Keyfitz Standard

Card	Columns	Data Item (Free-Format Name)
1	1-72	Name associated with population (NAME).
2	1-4	Year associated with population (DATE).
2	10-18	Total population (TP).
2	19-27	Total Deaths (TD).
2	28-36	Total births (TB).
2	37-45	Male births (TBM).
2	46-54	Female births (TBF).
2	63	Number of years covered in birth and death data. (Defaults to 1 year.)
3-4	(10F8.0)	Male population by age (19 age groups).
5-6	(10F8.0)	Female population by age.
7-8	(10F8.0)	Births by age of mother (BIRTHT).
9-10	(10F8.0)	Male deaths by age.
11-12	(10F8.0)	Female deaths by age.

Example Data Set 4 shows the minimum data set of Example Data Set 1 as it would appear in the Keyfitz standard format:

Example Data Set 4

```
UNITED STATES (WHITE) 1970
1970     177748704  1682096  2986491  1527290  1459199
 1433839 5614968 8264333 8647392 8079090 7341007 5962122 5042368 4936494 5412335
 5587023 5169302 4695581 4157467 3491080 2874531 2114943 1314258  889855       0
 1501250 5873083 8633093 9033725 8291270 6940820 5849792 4925069 4784375 5194497
 5257619 4832555 4310921 3647243 2807974 2107552 1437628  805564  486957       0
       0       0       0    9513  544531 1105555  803694  340864  141678   38420
    2235       0       0       0       0       0       0       0       0       0
   23155    3715    2647    2410    4673    4827    4361    4900    7448   12559
   20877   28926   39017   50851   67200   90108  113167  116590  142228       0
   31736    4912    4100    4383   12204   13817    9900    9133   12463   21826
   36004   53110   76528   98815  113653  122871  125021  101590   90370       0
**
```

4.3.3.6 The Preston Standard

The following data set appears twice, the first time for males and the second for females.

Card	Columns	Data Item
1	1-32	Country name.
1	33-39	Sex name (MALE or FEMALE).
1	40-44	Year associated with population.
1	45	Sex code (M or F).
1	48-58	Total population for this sex.
1	68-77	Total deaths for this sex.
2-3	(10F8.0)	Population by age.
4-5	(2x,F6.0, 9F8.0)	Total deaths by age.
6-7	"	Deaths from pulmonary tuberculosis by age.
8-9	"	Other infectious and parasitic diseases.
10-11	"	Neoplasms, both benign and malignant.
12-13	"	Cardiovascular diseases.
14-15	"	Influenza, pneumonia, & bronchitis.
16-17	"	Diarrhea, enteritis, and gastritis.
18-19	"	Certain degenerative diseases.
20-21	"	Maternal mortality.
22-23	"	Certain diseases of infancy.
24-25	"	Motor vehicle accidents.
26-27	"	Other accidents and violence.
28-29	"	All other causes & unknown.
30	(2x,12F6.0)	Total deaths for each cause.

4.3.4 Reading Data Files with the DATA Command

This section will describe the method for placing the data from the above files in the SPAHR current data block. For the purpose of this discussion, it is assumed that Example Data Set 1 has been allocated to file unit 11, Set 2 to file unit 12, Set 3 to 13, and Set 4 to 4 (the default unit).

Example Data Sets 1 and 2 may both be read the same way, by coding the word DATA and setting the parameter INPUT to 11 or 12:

```
DATA INPUT=12 ;
```

To see an exact copy of the data file printed out on the log file, use the ECHO option in the DATA command:

```
DATA ECHO INPUT=11 ;
```

To see the input data organized in a table, specify the PR1 option; to see the data after the age unknown category has been redistributed, specify the PR2 option. Any combination of these options may be specified. For example:

```
DATA ECHO PR1 PR2 INPUT=11 ;
```

The fixed-format data files (Example Data Sets 3 and 4) are read in exactly the same way, except that an additional parameter (FORMAT) must be specified to tell SPAHR which of the special formats (Chapter 4.3.1) is to be used. To read the SPAHR standard format, FORMAT should be set to 1:

```
DATA INPUT=13 FORMAT=1 PR1 ;
```

To read a Keyfitz standard file, FORMAT should be set to 2.

The ECHO parameter does not work for data that have been read in from a fixed-format file; however, such data may be modified by using free-format data input in the command file as part of the DATA command. For example, to read in the data from the Keyfitz standard input deck while changing the name of the population, enter:

```
DATA FORMAT=2 PR2 $$ NAME United States of America (White Population) 1970 ;
```

Because the INPUT parameter is not specified, the default (INPUT=4) is assumed.

Finally, the input data can be entered simply as part of the DATA command, without reference to external files:

```
DATA PR 1 $$ PARMS DATE=1970 $ NAME UNITED STATES (WHITE) 1970 $ POP MALE
 1501250
5873083 8633093 9033725 8291270 6940820
 5849792 4925069 4784375 5194497 5257619 4832555 4310921 3647243
 2807974 2107552 1437628 805564 486957 $$
      FEMALE 1433839 5614968 8264333 8647392 8079090 7341007
 5962122 5042368 4936494 5412335 5587023 5169302 4695581 4157467
 3491080 2874531  2114943 1314258 889855 $ DEATHS MALE 31725 4910 4099
4382 12200 13812 9897 9130 12459 21819
 35992 53092 76502 98781 113614 122829 124979 101556 90339 320 $$
     FEMALE 23151 3714 2646 2410 4672 4826 4360 4899 7447 12557
 20873 28920 39009 50841 67187 90091 113145 116567 142201 143 $
 BIRTHR 3*0 0.0011 0.0674 0.1506 0.1348 0.0676 0.0287 0.0071 0.0004 ;
```

4.4 OUT

The OUT command is a utility for producing a general purpose, machine-readable summary file containing an abstract of the demographic data concerning the current population. The output is placed on the default punch file unit (which can only be altered in the OPTIONS command) and has no parameters. The file it produces consists of six 80-column card images at each call, with a fixed format suitable for input to other computer programs.

IMPORTANT NOTE!!!!

The output will be generated whether or not SPAHR has calculated the data items involved, so the user must take care that the OUT statement follows the DATA, LIFETAB, and ANALYSIS commands. If OUT is called out of order, SPAHR will print gibberish in the uninitialized slots without any warning or error messages.

The machine-readable data block generated by the OUT command has the following structure:

Card	Columns	Data Item
1	1-72	NAME
2	5-12	Combined infant mortality rate.
2	13-20	General fertility rate (GFR).
2	21-28	Net reproduction rate (NRR).
2	29-36	Total fertility rate (TFR).
2	37-44	Mean age at birth of mothers.

Cards 3 and 4 refer exclusively to females.

Card	Columns	Data Item
3	1- 8	Geometric mean of the age-specific death rates.
3	9-16	Infant mortality rate.
3	17-24	Age-standardized death rate (Preston, Keyfitz, and Schoen's Young population).
3	25-32	Age-standardized death rate (Preston, Keyfitz, and Schoen's Old population).
3	33-38	Expectation of working life.
3	39-44	Expectation of life at birth.
3	45-50	Expectation of life at age 10.
3	51-56	Expectation of life at age 45.
3	57-62	Expectation of life at age 55.
3	63-68	Expectation of life at age 65.
4	1-14	Total population.
4	15-28	Total deaths.
4	29-42	Total births.
4	43-56	Total deaths before 1 year of age.

Cards 5 and 6 are identical to cards 3 and 4, except that they refer exclusively to males.

4.5 LIFETAB

The LIFETAB command computes an ordinary single-decrement life table. Refer to Volume I for a description of the life table technique.

Single-Valued Parameters

PRINT = File number for printed output.

PUNCH = File number for punched output, which takes the form of four cards containing the $_nL_x$ column of the life table in FORMAT (10F8.0) for consistency with Keyfitz and Flieger's life table program.

TERM Controls the log message. If specified, causes a line to be printed in the log file giving the expectations of life at birth for each sex. By default the message is printed. To suppress it, set TERM=0.

GRAPHM File unit on which SPAHR will print a crude semi-log graph of the age-specific death rates on the line printer.

GRAPHL Prints a crude line-printer graph of the probability of survival function $_nL_x$.

RATES Specifies that the life table will be based on the death rates in the current data block rather than on the raw population and deaths.

NGP Number of graphs per page. May be either one or two. (Default is two.)

DRMAX Maximum death rate to be plotted out on the graphs generated by the GRAPH and PLOT commands.

DRMIN Minimum death rate to be plotted.

PLOTM File to be used to produce special output file for later use in a graphics postprocessor (not yet available).

OVLM Specifies the overlay level for the plots generated by the PLOTM option. See the forthcoming graphics section.

=0: Each set of death rates will be plotted alone.
=1: Male and female plot overlaid.
=2: Plots overlaid by cause of death.
=3: Current plot overlaid on previous plot.

Required Data

POP and DEATHS must be read in DATA.

4.6 MULDEC

A detailed explanation of the multiple decrement life table and the associated single decrement table is presented in Volume I. The construction and use of multiple decrement life tables are discussed by Chiang (1968) and by Preston, Keyfitz, and Schoen (1972). Further examples of the analysis of such tables are given by Preston (1976).

MULDEC is designed to calculate measures of actuarial interest relating to various causes of death. It is also the principal method for initializing the array of death rates by age, sex, and cause used by other routines such as PROJECT and FACTOR. It will accept as input either deaths by cause or death rates by cause. However, if deaths are used as input, these data will be replaced by the corresponding death rates. The CAUSE option described below should therefore not be used in the first call following initialization. The first call to MULDEC notifies SPAHR that rates rather than deaths are present. Thus if a single cause is analyzed first, SPAHR will assume that all deaths have been converted into rates, and the deaths from those causes not treated in the first call will afterwards appear to SPAHR as death rates greatly in excess of one.

Single-Valued Parameters

PRINT = File unit number for printed output.

DETAIL = File unit number for detailed output.

PUNCH = File unit number for punched output.

TERM Controls the log file message. By default, a line is printed in the log file for each cause of death giving the probability of eventual death from, and the years of life lost to, each cause. Setting TERM=0 will suppress the printing.

INCREM = Maximum number of columns in summary tables to be printed. Default = 14.

GRAPHM Directs the printing of a semilog plot of the age-specific death rates for each cause.

NGP = Number of graphs per page. Default = 2. User may choose 1 or 2.

CAUSE = Index number of the cause of death to be processed. May also be used to indicate the first of a range of causes to be processed. If used to indicate a single cause of death, forces DETAIL output level and suppresses the normal print output.

CAUS2 = Final index number of range of causes of death to be processed. If CAUS2 is specified and CAUSE is not, CAUSE defaults to 1.

Required Data

Either:

1) POP, DEATHS, DRC or DDC read in DATA; call to LIFETAB, or

2) DR and DRC from a DATA entry.

4.7 ROOT

The ROOT procedure calculates the intrinsic rate of natural increase (Lotka's r) by three different methods. The first method is the direct solution of the numerical approximation to the integral equation. The second method uses the Laplace Transform approach. The third method estimates r as the dominant root of the Leslie projection matrix. The methods were described by Keyfitz and Flieger, whose ROOT program (Keyfitz & Flieger, 1971) is the core of this command.

Single-Valued Parameters

PRINT = File unit number for printed output.

TERM Specifies a single line of output to go to the log file.

Array Data

NMF = Net maternity function. This may be typed in, but is needed only if other data from which it can be computed have not been read in through DATA.

Required Data

Either: 1) NMF entered as an array item, or

2) POP and BIRTHS or BIRTHT entered in a call to DATA and either
   a) DEATHS entered in DATA and a call to LIFETAB, or
   b) VLL (see Chapter 4.3.2) entered in DATA.

4.8 ZEROS

This command calculates all roots, both real and complex, of the net maternity function. It is an adaptation of the Keyfitz-Flieger routine of the same name.

Single-Valued Parameters

PRINT = File number for printed output.

TERM Specifies a condensed output on the log file, consisting of the first three distinct roots.

Array Data

NMF = Net maternity function. This may be typed in, but is needed only if other data from which it can be computed have not been read in through DATA.

Required Data

Either: 1) NMF entered as an array item, or

2) POP and BIRTHS or BIRTHT entered in DATA and either
 a) DEATHS entered in DATA and a call to LIFETAB, or
 b) VLL (see Chapter 4.3.2) entered in DATA.

4.9 LOTKA

The LOTKA command fits the net maternity function to the normal distribution curve by the method of moments and performs an analysis of its properties under that assumption. It is taken from Keyfitz and Flieger's set of portable routines (Keyfitz & Flieger, 1971).

Single-Valued Parameters

PRINT = File number for printed output.

TERM Specifies a single line of output to go to the log file.

Array Data

NMF = Net maternity function. This may be typed in, but is needed only if other data from which it can be computed have not been read in through DATA.

Required Data

Either: 1) NMF entered as an array item, or

2) POP and BIRTHS or BIRTHT entered in DATA and either
 a) DEATHS entered in DATA and a call to LIFTAB,or
 b) VLL (see Chapter 4.3.2) entered in DATA.

4.10 STABLE

The STABLE procedure generates one or more stable-population age distributions from the current life table and a set of user-supplied intrinsic growth rates. If the FIT option is specified, the age distribution of the current population is then fitted to the set of derived distributions by the method of ogives described by Coale in the United Nations Manual IV (1967). If STABLE is called with no options, the value of R consistent with the set of birth and death rates will be used. In this case, or if R is the only option specified, only a single stable population will be computed.

Single-Valued Parameters

R = Estimate of the intrinsic rate of natural increase. If not included, it will be estimated from existing information if possible.

PRINT = File unit number for printed output.

FIT = File unit. Specifies that the existing age distribution is to be compared with all possible stable populations. For each age group, the value of R yielding the same cumulative age distribution is printed.

R2 = Final value of a series of R values to be used in calculating a series of stable age distributions.

RINC = Increment between successive R values. Default is 0.005.

N = Number of values of R to use. Default is the maximum permissable value of 12.

NOTE: Given any two of the parameters R, R2, RINC, and N, the remaining parameters will be calculated or assumed.

4.11 ANALYSIS

This command combines portions of the results of several other commands into a single analysis. It generates the stable age distribution, net maternity function, and mean ages of fertility under various assumptions and analyzes them.

Single-Valued Parameters

PRINT = File unit number for printed output.

DETAIL = File unit number for detailed output.

PUNCH = File unit number for punched output.

4.12 PROJECT

This command produces a projection in five-year intervals of the current population. Options available include the ability to override the input population, birth rates, or death rates, and dynamic modification of mortality and fertility rates during the course of the projection. Both period (cross-sectional) and cohort projections may be generated. Unique to this projection model is the ability to project deaths by age group and cause, and the option to perform Monte Carlo projections in addition to deterministic ones.

4.12.1 Single-Valued Parameters

Operating Mode Selection

The PROJECT command can produce several types of projections. The parameters in this section are used to specify the ones to be generated in the current invocation.

PERIOD — If mentioned, directs that a period (cross sectional) projection be performed. This is the default mode.

COHORT — Forces the production of a cohort projection. If mentioned as a switch, all five-year birth cohorts in the projection will be analyzed. If set to the initial birth year of a cohort, a projection will be performed only for those cohorts between the one mentioned and the one whose initial birth year is given in the parameter C2 (see below).

C2 = Initial birth year of last birth cohort within the range to be projected. Defaults to the value of COHORT, so that if it is not specified and COHORT is, only the birth cohort specified by the COHORT parameter will be analyzed.

CONSTANT — If mentioned, suspends the projection of the population age structure over time. The initial population will instead be propagated through unchanged. This option is included to allow determination of the effects of a constant versus a dynamic age structure; its results should therefore be compared with results of the PERIOD option.

RANDOM = Number of Monte Carlo simulation cycles to be performed. When projecting small populations, or when simulating relatively rare events, the variance of the estimate may be as important as the expected value. When RANDOM is specified, the indicated number of projections will be run from the same initial state as Monte Carlo simulations. Both births and survivors are randomly perturbed at each projection interval, and deaths by age and cause within each interval are likewise allocated with random perturbations. Means and standard deviations for all items in the summary result tables are printed out.

RANDOM mode is incompatible with COHORT and RATES modes. The variances of a cohort projection, however, are easily derived from Chiang's (1968) discussion of the variances of life tables.

SEED = Initializing parameter for the random number generator. It is needed only if RANDOM mode has been selected. It should be at least a six-digit number, preferably a large prime number, for best results. If it is not specified, or if it is specified as a negative number, the default seed will be selected. It is wise to specify your own seed for production work, especially if several runs of identical conditions are planned.

RATES If mentioned, causes the life tables and other vital rates to be projected normally. Instead of using the rates thus generated to analyze the future population, however, PROJECT produces a detailed analysis of the rates. The form and level of detail of this analysis are controlled by the LIFETAB and MULDEC subcommands (not to be confused with the commands of the same names). The use of this parameter results in the destruction of the data originally in the current data block.

ACCUM If mentioned, causes PROJECT (in PERIOD mode only) to write records into an intermediate file of the standard summary output. If ACCUM is set to a positive number, this summary file is rewound at the end of the current projection and then is read back in to generate a projection consisting of the pooled results of all previously ACCUMulated projections in the current run. It is important that all projections thus treated have the same number of causes of death and be projected over the same number of years.

POSTPROC If mentioned in conjunction with the ACCUM parameter forces PROJECT to generate the pooled projection from the existing summary file without performing an additional projection first.

File Unit/Switch Control Options

PRINT = File unit number for summary printed output.

DETAIL = File unit number for extensive data on each PROJECT interval.

PYRAMID= File unit number for projected population pyramids to be printed.

PRJSRV = File unit number for output from Leslie matrix calculations.

PRJKOS = File unit for projected deaths by cause. This generates a set of deaths by age, sex, and cause for each projection interval. INCREM species that causes of death will print out for each line.

DOSINT = File unit number on which PROJECT will print an integrated dose response analysis. Causes the program to accumulate the product of total person-years of exposure and annual dose rate for each dose and cause-of-death combination for each year of the projection. From this a gross risk estimator, defined as the ratio of cumulative excess deaths to cumulative integrated exposure, is computed at the end of the projection. Such estimators have been used in various radiation risk assessments, as for example the GESMO study (USNRC, 1976).

Interpolation Control Parameters

Baseline rates of mortality and fertility can be made to vary over time by using linear interpolation between the initial rates BIRTHR and DEATHR and the final rates BIRTHL and DEATHL (described in Chapter 4.12.2 below). Prior to the year STARTxx, the initial rates apply. Between the years STARTxx and STOPxx, the mean values of the birth rates are calculated for each projection interval and used as baseline rates. After year STOPxx, the final rates apply.

STARTDR= Initial year for death rate interpolation. Until this year, the initial death rates prevail.

STOPDR = Final year for death rate interpolation. After this year the death rates entered via the DEATHL array will prevail.

STARTBR= Initial year for birth rate interpolation. Until this year, the initial birth rates prevail.

STOPBR = Final year for birth rate interpolation. After this year, the birth rates entered via BIRTHL will prevail.

Other Switches and Assigned-Value Options

INCREM = Number of columns to be printed in cause-of-death tables. (Default = maximum = 14.)

NCOLS = Number of columns to be printed in the summary printed output tables, exclusive of the cause-of-death tables. (Default = maximum = 11.)

START = Starting year of the projection. (Defaults to the value of DATE in the DATA command.)

STOP = Terminal year of projection. Default = START+25.

POPSIZE= Override for initial population. If specified, the number of people in each age and sex group will be scaled accordingly.

RETAINP If this parameter is not specified, SPAHR does not alter the current population or any of its characteristics during the projection, unless the RATES option was selected. Thus when the projection is finished, the original population still exists. When RETAINP is specified, however, the population as projected will replace the starting population in the current data blocks, and the deaths and births in the current population will be recalculated to be consistent with the original birth and death rates.

RETAINDR Causes the mortality rates, both total and by cause, that prevail at the end of the projection to replace the ones in the current data block. Only those causes of death that produced deaths in the last projection interval will be retained. Note that both RETAINP and RETAINDR must be specified to retain the complete projected population and projected rates as the new current population.

YEAR = Number of years for which the projection should be performed. For example, YEARS = 30 would project the initial population 30 years ahead.

4.12.2 Array Data

POP Override population which will be projected in place of the one in the current data block.

BIRTHR Initial birth rates to be used in the current projection. Defaults to the current birth rates.

BIRTHL Birth rates by age of mother that will prevail at the end of the interpolation period specified by the parameters STARTBR and STOPBR. If not entered, the parameters STARTBR and STOPBR will be ignored.

DEATHR Initial death rates to be used in the current projection. Defaults to the current death rates.

DEATHL Death rates by age and sex that will prevail after year STOPDR and will be used between that year and year STARTDR to interpolate the death rate in conjunction with the starting death rates, which are entered either in the current data block or via array DEATHR. If not entered, the parameters STARTDR and STOPDR will be ignored.

4.12.3 Subcommands

4.12.3.1 ADJUST

The PROJECT command has the option of varying mortality rates dynamically over the course of the projection. One method for doing this is provided by the interpolation between the initial death rates and the ones specified in the DEATHL option. However, it is often of interest to assume that the change in mortality rates results from exposure to some adverse factor in the environment. The ADJUST subcommand implements several procedures for defining a dose history of a toxic substance for the population and for calculating the resulting changes in death rates over time.

Each of the models within the ADJUST subcommand defines one or more new causes of death in addition to those present in the original data. In REL or BACKGROU mode, it is further assumed that a cause of death with a name identical to one of the previously defined cause names has been entered in the DATA command. The causes of death newly defined in the ADJUST subcommand are assigned unique names by taking the previously defined names, converting any trailing blanks to underscores, and converting the eight characters into a digit corresponding to the order of the current ADJUST call. Thus if we invoke Model 6 in REL mode, it is assumed that a cause of death named LUNG that gives the mortality from cancer of the lung and bronchus (along with other cancers as indicated for this model) has been defined in the DATA command. If Model 6 is invoked in REL mode in the second call to ADJUST in the current projection, then a new cause of death with the name LUNG_2 will be defined for the duration of the projection. The death rates for this new cause of death will not be retained in the current data block after the projection is finished unless the RETAINDR parameter is coded.

The causes of death are not limited to those defined by default. (These are given for each model in the description of the MODEL parameter below.) The user may add causes, up to a limit of 28 for each ADJUST call, simply by adding members to the COEFS array by defining new cause names and associating them with appropriate coefficient values.

If more than one cause of death is defined in a particular call to ADJUST, then a summary cause named _EXCESSn is created as the sum of all other defined causes for call n. In addition, if more than one ADJUST invocation is made, a grand _EXCESS_ category is computed for all invocations.

Up to five ADJUST subcommands may be entered for one projection. Their effects will be added to the original rates. Any effects that depend on relative risk assumptions will be calculated on the basis of mortality rates as they appeared prior to the effects of any other ADJUST invocation. The total number of causes of death that may be projected, including those defined in DATA, is 36. Thus there is an effective upper limit of 17 on the number of causes that may be defined in relative risk mode.

NOTE: When constant exposure is assumed (i.e., when the dose is entered via the OPTIONS section of the subcommand, and the end of the exposure is greater than or equal to the end of the projection), then the rates prevailing at the end of 95 years will be held constant from that time forward and not recomputed.

4.12.3.1.1 Single-Valued Parameters

PRINT = File unit for printed output from the ADJUST routines. This output is typically a half page summarizing the actions of the subroutines specific to a particular model. Default is 0 for no printing. (Note: a summary of the net effects of all adjustments is given by using the ADJDUR option in PROJECT.)

DETAIL = File unit for detailed printed output from the ADJUST subroutines specific to the model. This output is typically one to three pages of printed output for each model for each projection interval.

START = Year of onset of exposure, Defaults to the initial year of the projection, unless the BACKGROU option has been selected, when the default will be 100 years earlier.

STOP = Year of cessation of exposure. Defaults to the last year of the projection.

MINAGEF= Age at onset of exposure for females.

MINAGEM= Age at onset of exposure for males.

MAXAGEF= Maximum age at which exposure takes place among females.

MAXAGEM= Maximum age at which exposure takes place among males.

MODEL = Mortality model index. Defines the type of hazard exposure and the dose response model that will be used to alter the mortality rates. Each of the models described here has two operating modes. In absolute mode (the default) the excess risks derived depend only upon the exposure level and duration, age, and sex. In relative mode, on the other hand, the increment in risk depends in addition on the level of mortality that would prevail in the absence of exposure to the risk factor of interest. The operating mode is determined by the REL parameter described below.

The first model is based on mortality due to cigarette smoking. Models 2 through 5 are derived by determining the equivalent dose in cigarettes per day required to duplicate the results in the studies cited, with the assumption that smoking began at birth. Model 6 is based on radiation risk assessments and uses a model developed by the BEIR committee (NAS, 1972). Models 7, 8, and 9 are derived from the BEIR committee report of 1980 (NAS, 1980). All of these models are described fully in Volume I.

MODEL = 1 Invokes a model indexed on cigarettes and based on studies by Hammond (1966), whose data were fitted to a model developed by

Lundy (Lundy & Grahn, 1977). Defines by default the single cause _EXTRA_.

= 2 Invokes a model indexed to suspended sulfates and based on work by Morris (Morris & Novak, 1976; Finch & Morris, 1977). Defines by default the single cause _EXTRA_.

= 3 Invokes a model indexed to SO_2 and total suspended particulates (TSP) derived from studies by Lave and Seskin (1977). Defines by default the single cause _EXTRA_.

= 4 Invokes a model indexed to benzo[a]pyrene and derived from the studies reviewed by Carnow and Meier (1973). Defines by default the single cause _EXTRA_.

= 5 Invokes a mortality change model indexed on total suspended particulates and derived from studies by Winkelstein (Winkelstein et al., 1968). Defines by default the single cause _EXTRA_.

= 6 Invokes a radiation-indexed model corresponding to the linear nonthreshold model, uncorrected for dose rate, presented in the 1972 BEIR report (NAS, 1972) as modified in the Reactor Safety Study (WASH-1400), often referred to as the Rasmussen Report (USNRC, 1975, 1976). By default, the following causes are defined:

LEUKEMIA - All leukemias, both lymphatic and myeloid.
LUNG - Cancers of the lung and bronchus.
STOMACH - Stomach cancer.
ALIMENRY - All cancers of the digestive tract except stomach and pancreatic.
PANCREAS - Cancer of the pancreas.
BREAST - All breast cancers.
BONE - Osteogenic sarcomas.
THYROID - Cancer of the thyroid.
OTHER - All remaining malignancies.

= 7 Invokes a radiation-indexed model corresponding to the linear nonthreshold model presented in the 1980 BEIR report (NAS, 1980) and modified by procedures outlined in Volume I. By default the following causes are defined:

THYROID - Cancer of the thyroid.
BREAST - All breast cancers.
LUNG - Cancers of the lung and bronchus.
ESOPH - Cancer of the esophagus.
STOMACH - Cancer of the stomach.

INTEST - Cancer of large and small intestines.
LIVER - Cancer of the liver.
PANCREAS - Cancer of the pancreas.
URINARY - Cancer of the urinary organs.
LYMPHOMA - Lymphosarcoma, reticulum cell sarcoma, and multiple myeloma.
LEUKEMIA - All leukemias.
BONE - Cancer of the bone.
OTHER - All other cancers.

= 8 Invokes a radiation-indexed model corresponding to the linear-quadratic, nonthreshold model presented in the 1980 BEIR report (NAS, 1980) and modified by procedures outlined in Volume I. The cause names and use of the option REL are identical to the descriptions for Model 7.

= 9 Invokes a radiation-indexed model corresponding to the pure quadratic, nonthreshold model presented in the 1980 BEIR report (NAS, 1980) and modified by procedures outlined in Volume I. The cause names and use of the option REL are identical to the descriptions for Model 7.

REL Relative mode switch. If not mentioned, the absolute risk version of each of the models specified in the MODEL parameter will be used. If REL is specified with Models 6-9, then an appropriate set of cause-of-death data must have been entered via the DATA command, with names identical to those listed above for the particular model. The only exception to this rule occurs for the special cause name _EXTRA_, which is assumed relative to total mortality prior to the effects of any ADJUSTments.

BACKGROU Corrects the initial or baseline rates of mortality in all the models to remove the assumed effects of background exposure to the pollutant in question. Both total and identically named cause-specific death rates, if entered, will be decreased accordingly.

LPLAT Long PLATeau switch is used only for Model 6. The effect of radiation exposure for all causes except leukemia persists for a maximum of 30 years (the default) or the lifetime of the individual exposed. LPLAT forces the latter assumption. This option has no effect on leukemia, which has a short plateau, even under conservative assumptions in the BEIR (1972) report.

The identities and dose levels of the pollutants are given by the keywords:

SO_2 = $\mu g/m^3$ of sulfur dioxide.

SO_4 = $\mu g/m^3$ of suspended sulfates.

TSP = $\mu g/m^3$ of total suspended particulates.

POM = $\mu g/m^3$ of benzo[a]pyrene.

REM = Whole-body equivalent dose in REM/year.

CIG = Cigarettes smoked per day.

When dealing with TSP pollution in the presence of particulate control technology, it must be noted that the various control techniques do not trap particulates of all size ranges with equal ease. Therefore, the following parameters are used to estimate the effective TSP dose:

RFRAC = Proportion of particulates in the uncontrolled emissions that are included in the overall efficiency of particulate removal.

ENRESP = Efficiency of removal of nonrespirable particulates.

ERESP = Efficiency of removal of respirable particulates.

Given RFRAC and any one of the remaining two parameters in this group, the biologically effective TSP dose multiplier is computed and applied to the input TSP value.

4.12.3.1.2 Array Data

All parameters of all models within the ADJUST subcommand are supplied with default values. However, the default values may all be modified as desired by the user.

COEFS Each model has an associated set of coefficients. The values of these coefficients, their derivations, and the forms of the models to which they apply are described in Volume I. Each model is sex-specific, so the secondary qualifiers are always MALE and FEMALE. The primary qualifiers are always the names of the cause(s) of death generated in the model. In Models 1 through 5, only one cause (named _EXTRA_) is generated by default. In Model 6, eight cancers and a residual cancer group are defined by default. In Models 7, 8, and 9, twelve cancers and a residual cancer group are defined by default. New causes may be defined in any model simply by adding additional cause names to the COEFS array. For the radiation models, defining new causes also modifies the names of the dose terms expected.

PLATEAU This is an auxilliary array used only with the radiation risk models derived from the two BEIR reports (Models 6, 7, 8, and 9). For each cause it represents the duration of risk after the period of latency following exposure has passed. There is one member of this array for each member of the COEFS array, except that PLATEAU is not sex-specific.

LATENT — This is an auxilliary array used only with the radiation risk models derived from the TWO BEIR reports (Models 6, 7, 8, and 9). For each defined cause it represents the number of years of latency that must elapse between exposure and the earliest observed increased risk of death. It is not sex-specific, so the only qualifiers are the cause names.

MORBID — This array allows SPAHR to calculate incidence of disease from mortality estimates. It is cause and sex specific. Default values are provided only for Models 6, 7, 8, and 9. Each cause and sex of the MORBID array has three members. The first value is the factor by which calculated deaths must be multiplied to generate the number of cases. If the case fatality rate is 50%, then this first coefficient would have the value 2.0. The second value is the mean duration (in years) between diagnosis and death for those who die. The third value is the mean number of years spent in an officially defined morbid state by those who recover.

DOSE — This array contains the doses for all models in the event that a time-varying or organ-specific dose is utilized. Models 1-5 in the present version of SPAHR cannot handle time-varying doses. However, provision has been made for future enhancements to these models, and doses entered with the appropriate qualifiers will not generate an error message. Dose is not assumed to be sex specific, so there is only a single set of qualifiers. The qualifiers recognized in this version are:

MODEL	Primary Qualifiers
1	CIGARETTES
2	SO_4 (Suspended Sulfates)
3	SO_2 (Sulfur Dioxide)
3	TSP (Total Suspended Particulates)
4	POM (Polycyclic Organic Matter)
5	TSP (Total Suspended Particulates)
6,7,8,9	WHBI (WHole-Body Irradiation) and organ doses named identically to the defined cause names

WARNING: The ADJUST STOP parameter does not apply to entries in the DOSE array. It affects only the global constant dose specified in the single-valued parameter section of the ADJUST subcommand. If a dose history extending past the value of the STOP parameter is entered in the DOSE array, it will not be automatically terminated in year STOP.

4.12.3.2 LIFETAB and MULDEC

Whenever the RATES or RETAINDR parameters are invoked, PROJECT calculates life tables using the command routines for the LIFETAB and MULDEC commands.

The output from these calculations may be controlled by the use of the LIFETAB and MULDEC subcommands. The forms and options of the subcommands are identical to those of the commands. The user is referred to the descriptions of the LIFETAB and MULDEC commands (Chapters 4.5 and 4.6) for the available options. The defaults in subcommand mode differ from those for the commands, however, in that TERM and PRINT default to zero, thus omitting the log file and printed outputs, which must be requested explicitly.

WARNING !! With the MULDEC command in particular it is possible to generate a large amount of output by using the DETAIL option. Since these routines may be invoked many times by PROJECT, be careful not to request more output than can feasibly be studied in the time available.

4.13 TITLE

The TITLE command sets a character string that will be printed by SPAHR at the top of every page except for the joblog or terminal file. TITLE has the general form

TITLE <string of up to 40 characters>;

The text can be terminated with $$, $, or ; and therefore may not include these symbols. Alternatively, the text may be enclosed by either single (') or double quotation (") marks, in which case only the semicolon is prohibited. The title string can include the current value of a SPAHR variable. This is known as a variable expansion. If the variable name is preceded by an ampersand (&), the value of the variable will be printed instead of the name of the variable. For example:

TITLE TEST RUN;

TITLE "Much Fancier $64,000//Test Run//";

N = 12 ; TITLE Test number &N ;

In this example, the title that prints out is

Test number 12

4.14 SET

The SET statements permit the user to create and manipulate scalar variables. SET statements are similar in function to assignment statements in other higher-level languages such as FORTRAN, PL/I, and BASIC, and they have the form

[SET] <variable name> = <constant or expression> ;

The keyword SET is optional, as this type of statement may also be detected by the presence of the = as a delimiter of an otherwise command-level string. Addition, subtraction, multiplication, and division are permitted. Manipulation with exponents, special functions such as log and square root, and parenthetical groups are not available. Either the keyword SET or the equal sign may be omitted, because SPAHR uses either one as the signal to invoke the SET routines. Thus

```
SET XX = 7 ;
SET XX 7 ;
XX = 7 ;
```

will all be interpreted identically. All three statements will define a SPAHR variable named XX with the value 7.

4.15 PRINT

PRINT permits the user to print out the current value of any user-defined or scalar system variable. It has the form

PRINT <list> ;

where the <list> may include any combination of system variables, user-defined variables, file indicators, column indicators, and character constants.

Data

When a variable name is encountered in the <list>, the value currently associated with that variable is printed out. The value will be left-justified starting in the current column position of the print line, expressed in scientific notation with six digits of precision, and non-significant zeros will be omitted.

Character

A character constant is any string of characters enclosed in either single or double quotation marks. It will be transferred exactly as entered (omitting the quotation marks) to the print line starting at the current column.

System

The strings read in by using the TITLE command or the NAME parameter in the DATA command may be referenced in the PRINT statement as _TITLE_ and _NAME_ respectively. The special name _PAGE_ causes a page to be skipped and a title message to be printed out.

Column

By default, the current column is the column immediately following the last item entered. However, the current column number may be reset by entering a @ symbol immediately followed by an integer giving the column to be used as the current column number.

File

By default, the print line will appear on the terminal or log file (unit 3). However, another file unit may be specified by including a # symbol followed immediately by a number representing a different file unit. If more than one file indicator is entered in a PRINT statement, the last one entered will apply. The entry

```
X = 2.3 ; Y = 0.00034 ; Z = 1005684.8 ;
  PRINT #6 "X=" X @40 Y @20 "Z = " Z ;
```

will print the following line on unit 6:

```
X=2.3                3.4E-4                Z = 1.00568E6
```

4.16 DO

The DO statement is used to define an associated group of statements. The end of the group must be demarcated by an END statement. Statements in a DO group may be executed repetitively and may be nested in up to nine levels. The DO statement has the form

```
DO <variable name> = <initial value> TO <final value>
                BY <increment value> ;
```

The control variable thus created is a user-defined variable and may therefore be used or altered with a SET statement. The BY value is optional and defaults to 1.0. If no control variable is specified, then the defined group is executed only once.

SPAHR processes DO statements by reading them into a buffer and reinterpreting them on each pass without reprinting them. Therefore, messages printed out will not be easily associated with the DO statements to which they refer. To alleviate this problem, the user may wish to print out a message containing the value of the control variable at the beginning of each pass, as

```
DO X = 1 TO 10 BY 2 ;  PRINT "====== X LOOP:  X = ",X," =======" ;
```

```
.
.
.  (other SPAHR statements)
.
.
.  END ;
```

In addition, the informative messages that print out on the log file in response to each command may generate more output than the user needs. Most commands are therefore equipped with a TERM parameter that defines the importance level required for a message to be printed out. Setting TERM=0 in most commands suppresses most terminal output. For example, the sequence

```
DO NPOP=1 TO 1000 ;
  DATA INPUT=18 FORMAT=1 TERM=0 ;
  LIFETAB TERM=0 ; ANALYSIS TERM=0 ;
  PRINT 'NO.' NPOP _NAME_ EOM EOF 'NRR=' NRR ;
  END ;
```

will cause only a single line to be printed out for each population, rather than the five to seven lines that would print out if TERM were not set to zero.

4.17 IF

The IF statement allows conditional execution of SPAHR commands. It has the form

```
IF <logical expression> THEN <command or DO group> ;
                   ELSE <command or DO-group> ;
```

The logical expression may be either a single simple expression or a compound logical expression connected with AND and/or OR operands. The simple expressions have two forms: the name of an array specifiable in the DATA command that returns the value <u>true</u> if the array has been properly initialized and <u>false</u> if it has not; or a relational comparison between two system or user variables, or between one variable and a constant (variables are discussed in Chapter 3.7). Unlike FORTRAN and PL/I, arithmetic expressions <u>may not</u> be imbedded within the logical expression. The relational operators are defined as follows:

Operator	Meaning
EQ	Equal to
NE	Not equal to
LT	Less than
LE	Less than or equal to
GT	Greater than
GE	Greater than or equal to

The logical variables corresponding to the DATA arrays return the value true if they have been explicitly mentioned in the DATA command, even if in that command they were set to unreasonable values, except for primary data arrays POPULATION, BIRTHS, and DEATHS. For these arrays, true is returned if and only if at least one positive entry age group for each sex is greater than 0.5.

The statement or DO group immediately following the keyword THEN will be executed if and only if the value of the logical expression is true. The statement or DO group following the keyword ELSE will be executed if and only if the value of the logical expression is false. The THEN clause is required at all times; the ELSE clause is optional and may be omitted.

The keyword THEN is used to terminate the logical expression. Therefore, a variable with the name of THEN may never be used on the left side of a relational expression.

4.18 END

The END statement has no operands. It serves two functions. The first is to terminate DO groups. The second is to terminate execution of SPAHR entirely. A separate END statement must be provided for each DO statement.

If SPAHR encounters an end of file in the command file or any free-format data file, an END statement is automatically generated. Consequently, the final END statement in a SPAHR program is optional.

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

5.0 RUNNING SPAHR

5.1 The Interactive Version of SPAHR

Because SPAHR is so versatile, it may be difficult to construct programs for sophisticated analyses. Therefore, an interactive question and answer capability is available. This procedure is self documenting, so that the user needs little or no familiarity with either the language structure or commands. Currently three interactive packages are available in SPAHR. These are PRIMER, SITE, and WORKER. Each of these packages has been designed to address a separate series of questions. A detailed discussion of the use of these interactive packages is available in Volume III of this series.

5.2 SPAHR under CMS

The use of the interactive packages reduces the flexibility of SPAHR, but SPAHR can be run directly on CMS. This section outlines two procedures for doing so. However, only users familiar with the SPAHR language structure and command reference should attempt this. CMS is an interactive system; the commands are proccessed as they are entered, and SPAHR provides prompts as needed. In Example Session 1, the user must log on to CMS as described in Volume III, SPAHR Interactive Package Guide. In addition, the data set DATA1 has been selected for analysis in this example. A complete listing of all data sets available for analysis is provided by entering:

TYPE WORKER DATA A

In the sample sessions below, lines that are printed in lower-case letters and begin with a period are lines entered by the user. Lines that do not begin with a period and are printed in upper-case letters are generated by SPAHR. The period at the beginning of the input lines is printed out as a prompt by CMS to tell the user that it is ready to receive input. Each line typed in at the terminal must end with a carriage return, or the computer will not know that typing is finished.

CMS Example Session 1

```
.spahr * data1

SPAHR VERSION 4.1

COMMAND-
.data ;
<DATA> UNITED STATES(WHITE) 1970
COMMAND-
.lifetab ;
<LIFE TABLE> EXPECTATION OF LIFE FEMALE 75.616 MALE 67.942
<LIFETAB > PRINTED PAGES 1 TO 1 ON FILE 3
```

```
COMMAND-
.muldec ;
<MULDEC> CAUSE OF      COHORT DEATHS          YEARS LOST
          DEATH       FEMALE      MALE      FEMALE     MALE
         LEUKEMIA       671.       743.      0.129      0.131
         LUNG          1240.      4895.      0.219      0.677
         STOMACH        712.       898.      0.087      0.103
         ALIMENRY      3499.      2967.      0.459      0.352
         PANCREAS      1061.       779.      0.160      0.085
         BREAST        3040.        25.      0.552      0.003
         BONE            80.        89.      0.017      0.018
         THYROID         82.        35.      0.011      0.005
         OTHER         5756.      5643.      0.970      0.808
         CANCER       16141.     16074.      2.756      2.334
<MULDEC> PRINTED PAGES   2 TO 5 ON FILE 3
COMMAND-
.project stop=2070 term/
MORE ?
.adjust model=6 rem=0.1 ;
<PROJECT> (  0) POP. IN 2070 IS    292544256.
<PROJECT> PRINTED PAGES  6 TO 15 ON FILE 3
COMMAND-
.end ;
SPAHR NORMAL TERMINATION.
```

The user initiated the SPAHR run by issuing the command spahr * data1 at the terminal. This command told the system to execute the EXEC file named SPAHR EXEC. The * was defined in the exec file to mean that the command file would be entered in interactive mode directly from the terminal. The DATA1 told the system to attach a file named DATA1-DATA to unit 4. NOTE: Because SPAHR is written in FORTRAN, all CMS disk files that it reads must be fixed-length files (RECFM F) with a logical record length of 80 bytes (LRECL 80). Variable-length CMS files or those with one of the compressed formats may not be used.

SPAHR then printed out a line identifying the version number and prompted the user for a command by typing COMMAND-. At this point the user entered a command, which in this case was a DATA command. SPAHR interpreted the command, read in a population from unit 4, and printed out the name of the population. Another command prompt was then issued, which was answered with a LIFETAB command. LIFETAB executed as ordered and printed out its log file line by way of confirmation. Because LIFETAB wrote something on the print file, the SPAHR command processor also printed out a log file line indicating which page(s) in the print file are to be associated with this command. A similar sequence was followed for the MULDEC command. When the PROJECT command was entered, however, the first line terminated with a slash (/). SPAHR recognized from this that the command was not yet finished and so prompted MORE?. After finishing the projection, the END command was entered, and SPAHR terminated, returning the user to the CMS system. The print file (unit 3) remained behind in the form of a file named SPAHR 3 LISTING A. It can be printed out by using the CMS PRINT command or viewed by using the text editor. NOTE: If SPAHR is executed in this fashion twice in a row, the print file SPAHR 3 LISTING A generated the first time will be overwritten by the print file generated the second time. The sequence of runs must be planned accordingly.

To run SPAHR on CMS in a noninteractive mode, the user can create a file named COMFILE COMMAND with instructions similar to those in CMS Example Session 1, as follows:

Example CMS File COMFILE COMMAND A

```
OPTIONS BATCH ;
DATA ;
LIFETAB ;
MULDEC TERM=0
PROJECT STOP=2070 TERM DOSINT=6 / ADJUST MODEL=6 REM=0.1 ;
END ;
```

The commands have changed slightly. The MULDEC terminal output has been suppressed by specifying TERM=0 in the MULDEC command. The amount of log file output from the PROJECT command has been increased by specifying DOSINT=6.

Now we can execute a SPAHR job with a single command, the first line shown here:

CMS Example Session 2

```
.spahr-comfile data1
<SPAHR> VERSION 4.1
COMMMAND-
data ;
<DATA> UNITED STATES (WHITE) 1970
lifetab ;
<LIFE TABLE> EXPECTATION OF LIFE FEMALE 75.616 MALE 67.942
<LIFETAB > PRINTED PAGES   1 TO   1 ON FILE   3
MULDEC TERM=0
<MULDEC > PRINTED PAGES   2 TO   5 ON FILE   3
PROJECT STOP=2070 TERM DOSINT=6 / ADJUST MODEL=6 REM=0.1 ;
INTEGRATED DOSE ESTIMATORS
```

```
CAUSE        DEATHS        MAN-DOSE      ESTIMATOR
LEUKEMI1       5.274E+04     2.421E+09     2.179E-05
LUNG___1       4.723E+04     2.421E+09     1.951E-05
STOMACH1       1.417E+04     2.421E+09     5.853E-06
ALIMENR1       4723.         2.421E+09     1.951E-06
PANCREA1       4723.         2.421E+09     1.951E-06
BREAST_1       3.852E+04     2.421E+09     1.591E-05
BONE___1       1.037E+04     2.421E+09     4.286E-06
THROID_1       2.075E+04     2.421E+09     8.571E-06
OTHER__1       3.164E+04     2.421E+09     1.307E-05
_EXCESS1       2.249E+05     0.0           0.0
<PROJECT> (  0) POP. IN 2070 IS    292544256.
<PROJECT > PRINTED PAGES    6 TO    15 ON FILE  3
END ;

<SPAHR> NORMAL TERMINATION.
```

Notice that SPAHR issued a single COMMAND- prompt line at the beginning and did not echo back the OPTIONS statement. This is because until SPAHR processed the OPTIONS statement, it assumed that it was executing interactively.

The SPAHR exec file is self-documenting. If the user enters the line SPAHR ?, the current options available in the exec file will be listed. These typically include methods for sending the print file directly to a line printer instead of leaving it on disk, for placing the log file on disk instead of printing it out at the terminal, and for various other useful items. The default data file name will also be identified. Entering the exec file given in the SPAHR Programmer's Guide will produce the following response:

SPAHR fn1 fn2 (PRINT xxx LOG yyy)

where fn1 is the file name of fn1 COMMAND *, which is the SPAHR command file. If a * is entered here, command lines will be taken directly from the terminal.

fn2 is the file name for fn2 DATA *, which is a SPAHR data file that will be allocated to unit 4. This is set by default to TEST.

xxx (default DISK) is the disposition of the SPAHR print file allocated to unit 3. It may be one of:

DISK - Directs that the print file be placed on disk file SPAHR3 LISTING A.

PRINT - Directs that the print file be allocated to the user's CMS virtual printer.

yyy (default TERM) is the disposition of the SPAHR log file allocated to unit 6. It may be one of:

TERM - Directs that the log file be printed out at the user's terminal.

DISK - Directs that the log file be placed on disk file SPAHR6 LISTING A.

PRINT - Directs that the log file be allocated to the user's CMS virtual printer.

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

BIBLIOGRAPHY

Carnow, B. W. and P. Meier 1973: "Air Pollution and Pulmonary Cancer," *Arch. Env. Health.* 27:207-218.

Cook, J. R., B. M. Bunger, and M. K. Barrisk 1978: *A Computer Code for Cohort Analysis of Increased Risk of Death* (CAIRD), ORP Technical Report 520/4-78-012, Environmental Protection Agency, Washington, D.C.

Chiang, C. L. 1968: *Introduction to Stochastic Processes in Biostatistics,* Wiley and Sons, New York.

Finch, S. J. and S. C. Morris 1977: *Consistency of Reported Health Effects of Air Pollution,* BNL-21808, Brookhaven National Laboratory, Upton, NY.

Hammel, E. A., D. W. Hutchinson, K. W. Wachter, R. T. Lundy, and R. Z. Deuel 1976: *The SOCSIM Demographic-Sociological Microsimulation Program,* Institute of International Studies Monograph No. 27, Institute of International Studies, Berkeley.

Hammond, E. C. 1966: "Smoking in Relation to the Death Rates of One Million Men and Women," in Haenzel, W., Ed., *Epidemiological Study of Cancer and Other Chronic Diseases,* National Cancer Institute Monograph 19, U.S. Public Health Service, Washington, D.C.

Keyfitz, N. and W. Flieger 1971: *Population: Facts and Methods of Demography,* W. H. Freeman, San Francisco.

Lave, L. and E. P. Seskin 1977: *Air Pollution and Human Health,* Johns Hopkins University Press, Baltimore.

Lindberg, W. 1968: *The General Air Pollution in Norway,* Norwegian Smoke Damage Council, Oslo.

Lundy, R. T., and D. Grahn 1977: "Predictions of the Effects of Energy Production on Human Health," *Proceedings of the Social Statistics Section, Part II: 672-675,* American Statistical Association, Chicago.

Morris, S. C. and K. M. Novak 1976: *Handbook for the Quantitation of Health Effects due to Coal* (Draft), Brookhaven National Laboratory, Upton, NY.

National Academy of Sciences, Advisory Committee on the Effects of Ionizing Radiation 1972: *The Effects on Populations of Exposure to Low Levels of Ionizating Radiation,* National Academy of Sciences, Washington, D.C.

National Academy of Sciences, Advisory Committee on the Effects of Ionizing Radiation 1980: *The Effects on Populations of Exposure to Low Levels of Ionizing Radiation,* National Academy of Sciences, Washington, D.C.

National Cancer Institute, SEER Program 1976: *Cancer Patient Survival, Report Number 5,* DHEW Publication No. (NIH) 77-992, Washington, D.C.

Patrick, C. H. 1977: "Trends in Public Health in the Population near Nuclear Facilities," *Nuclear Safety* 18(5):647-662.

Polissar, L. H. 1980: "The Effects of Migration on Comparison of Disease Rates in Geographical Studies in the United States," *Am. J. Epidemiology* 111(2):175-182.

Preston, S. H. 1976: *Mortality Patterns in National Populations,* Academic Press, New York.

Preston, S. H., N. Keyfitz, and R. Schoen 1972: *Causes of Death: Life Tables for National Populations,* Seminar Press, New York.

Thomas, T. 1973: *An Investigation into Excess Mortality and Morbidity due to Air Pollution,* Ph.D. Thesis (Environmental Science), Purdue University.

United Nations, Department of Economic and Social Affairs, Population Branch 1967: *Methods of Estimating Basic Demographic Measures from Incomplete Data,* Manual IV, Population Study No. 42, United Nations, New York.

U. S. Nuclear Regulatory Commission 1975: *Reactor Safety Study: Appendix VI Calculation of Reactor Accident Consequences,* WASH-1400 (NUREG 75/014) USNRC, Washington, D.C.

U. S. Nuclear Regulatory Commission 1976: *Final Generic Environmental Statement on the Use of Recycled Plutonium in Mixed Oxide Fuel in Light Water Cooled Reactors,* Vol. III, Chapter IV, NUREG-0002, ES, Washington, D.C.

Winkelstein, W. Jr., S. Kantor, E. W. Davis, C. S. Maneri, and W. E. Mosher 1968: " The Relationship of Air Pollution and Economic Status to Total Mortality and Selected Respiratory System Mortality in Men II: Oxides of Sulfur," *Arch. Env. Health* 16:401-405.

Do Not Microfilm Distr

Distribution for NUREG/CR-2364 (ANL-81-59)

Internal:

M. K. Butler
D. Grahn
M. W. Rosenthal (4)
F. S. Williamson
ANL Patent Dept.
ANL Contract File
ANL Libraries (2)
TIS Files (6)

External:

USNRC, for distribution per RH (205)
DOE-TIC (2)
Manager, Chicago Operations Office, DOE
President, Argonne Universities Association
Division of Biological and Medical Research Review Committee:
- T. W. Clarkson, U. Rochester
- L. Grossman, Johns Hopkins U.
- J. W. Osborne, U. Iowa
- H. C. Pitot III, U. Wisconsin-Madison
- M. Pollard, U. Notre Dame
- F. W. Putnam, Indiana U.
- P. O. P. Ts'o, Johns Hopkins U.

U. S. Nuclear Regulatory Commission:
- E. Branagan
- J. D. Foulke
- M. E. Ginevan
- R. Gotchy
- C. Willis

U. S. Department of Energy:
- N. F. Barr
- C. W. Edington
- R. Goldsmith
- M. L. Minthorn, Jr.
- J. W. Thiessen

J. J. Collins, American Cyanamid Co., 1 Cyanamid Plaza, Wayne, NJ 07470 (21)
C. Land, National Cancer Institute, Bethesda, MD 20205
R. T. Lundy, Alameda, CA 94501