JOEM • Volume 40, Number 4, April 1998

OCMAP-PLUS: A Program for the Comprehensive Analysis of Occupational Cohort Data

Gary M. Marsh, PhD
Ada O. Youk, PhD
Roslyn A. Stone, PhD
Stephen Sefcik, BS
Charles Alcorn, BS

OCMAP-PLUS: A Program for the Comprehensive Analysis of Occupational Cohort Data

Gary M. Marsh, PhD
Ada O. Youk, PhD
Roslyn A. Stone, PhD
Stephen Sefcik, BS
Charles Alcorn, BS

The Occupational Cohort Mortality Analysis Program (OCMAP) has been redesigned for optimal microcomputer use and extended to include new computing algorithms. The new program, OCMAP-PLUS, offers a comprehensive, flexible, and efficient analysis of incidence or mortality rates and standardized measures in relation to multiple and diverse work history and exposure measures. New features include executable code, minimization of memory requirements, disk file storage of person-day arrays, stratified analyses by geographic area, employment status and up to eight exposure variables, a data imputation algorithm for study members with unknown race, and enhanced algorithms for constructing several time-dependent exposure measures. New modules create grouped data files for Poisson and logistic regression and risk set files for use in relative risk regression analysis. The Mortality and Population Data System (MPDS) provides external comparison rates and proportional mortalities. Analysis from two recent cohort mortality studies illustrate several new features.

A major objective of occupational epidemiology is to identify the health consequences of workplace exposures so that remedial efforts can be implemented when indicated. Other objectives are to provide data useful for setting standards for protection of workers exposed to toxic substances, to make projections of risk to members of less-exposed populations, and to elucidate mechanisms of toxicity and exposure-response relationships. Three of the most common epidemiological study designs used to meet these objectives are the historical cohort, proportional mortality, and case-control designs.[1] Each design requires investigators to enumerate a study population and to collect individual-level data on demographic factors (eg, age, race, sex), employment and exposure history, health or vital status, and factors that can potentially confound exposure-response relationships (eg, cigarette smoking or dietary history). A comprehensive descriptive and analytical evaluation of such data involves the calculation of many time-dependent measures.[2–4]

The Occupational Cohort Mortality Analysis Program (OCMAP),[5–8] developed at the University of Pittsburgh, is a widely distributed modularized computer program that performs many basic data management and statistical analysis procedures for these three epidemiological study designs. While OCMAP is designed primarily for occupational mortality studies, applications generalize easily to studies of other health end-

From the Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pa. (Dr Marsh, Dr Youk, Dr Stone); computer programming consultant, Pittsburgh, Pa. (Mr Sefcik); and CWA Computer Consulting, Waynesburg, Pa. (Mr Alcorn).

Address correspondence to: Gary M. Marsh, PhD, A410 Crabtree Hall, University of Pittsburgh, Pittsburgh, PA 15261.

1076-2752/98/4004-0351$3.00/0

points, such as cancer incidence, and to studies in non-occupational settings.[5] The current OCMAP program is now in use by over 300 institutions in the United States and abroad. Releases of OCMAP have been referenced in over 150 peer-reviewed journals, based on a review of the Scientific Citations Index database.[9]

Other similar programs for the analysis of occupational study data are available.[10–12] None of these currently available programs, however, enables a comprehensive, flexible, and efficient analysis of disease outcome or effect measures in relation to multiple, diverse measures of employment history and/or exposure. In addition, none provides a direct interface with other programs commonly used for more extensive regression modeling of cohort data.

To overcome these and other limitations in the OCMAP (Versions 1.0–2.1) programs, and to meet the specific analytic requirements of an ongoing cohort mortality study of man-made vitreous fiber (MMVF) workers, we redesigned and extended the OCMAP microcomputer software. In this article, we describe the general structure and analytical capabilities of the new OCMAP-PLUS program and provide examples of its use in the MMVF cohort study and in a recent study of formaldehyde-exposed workers.

Major Enhancements to OCMAP Implemented in OCMAP-PLUS

First, we describe the major enhancements implemented in OCMAP-PLUS in relation to existing features in OCMAP and OCMAP/PC (referred to as "prior releases" of the program).

Improved Operational Efficiency

The original OCMAP[5,6] was developed for mainframe environments where memory is abundant but disk storage can be limited. OCMAP/PC[7] was later developed for *personal computer (PC)*-based (MS-DOS; Microsoft Corp, Redmond, WA) applications. However, because OCMAP/PC is basically OCMAP recast in Microsoft FORTRAN, it did not perform optimally in the PC environment when random-access (RAM) memory was limited. To operate optimally within a current PC-based environment, where large disk storage is now affordable and readily available, OCMAP-PLUS minimizes memory requirements but utilizes extensive disk input and output (I/O).

Distribution of OCMAP-PLUS as executable code, rather than source code, eliminates the previously required text-editing, compiling, and linking of the source code. A FORTRAN compiler is no longer required.

More Centralized Processing Structure

Prior releases of OCMAP contained up to five program modules (referred to as "Mod"): Mod 1—Standardized Mortality Ratio Analysis, Mod 2—Proportional Mortality Ratio Analysis, Mod 4—Effective Exposure Modeling,[6] Mod 5—Internally Standardized Relative Risk Analysis, and Mod 6—Directly Standardized Measures.[8] Each of these is a "stand alone" program that performs each of the tasks necessary to input data, perform analysis, and output report tables. These tasks are:

Task 1—Create and process the user control file.

Task 2—Interface with the standard mortality files and construct appropriate cause of death groupings

Task 3—Access the cohort file, select cases and/or jobs within cases, perform status tests and data recodes, analyze and summarize data.

Task 4—For selected cases, accumulate and aggregate data.

Task 5—Prepare and output tables for reports.

Prior releases also contain a cohort data editing module (Mod 3) that is run before the other program modules. Prior releases did not contain any general-purpose data export capability.

In OCMAP-PLUS, data and programs related to setup, execution, and output are organized more centrally into five sequential phases (inputs, pre-processing, processing, post-processing, and reports/data export). This more efficiently structured data processing platform eliminates much of the processing redundancy in the earlier releases. Unlike prior releases, specialized programs in OCMAP-PLUS now handle Tasks 2 through 5 and the additional task of data export (Task 6—create analysis files for use with other software).

The programs in OCMAP-PLUS are classified into three groups: processing, reporting, and utility. These three groups of programs handle all five phases of an OCMAP-PLUS analysis. Figure 1 illustrates the sequential processing structure of OCMAP-PLUS and associated programs and relates these to Tasks 1–5 of prior releases of OCMAP:

Inputs. The study data (cohort data file and standard death rate files) are assembled and the program control files are created in this phase. The creation of program control files is facilitated by a utility program, MAKE-PLUS. This utility can be used to create new control files, edit earlier control files, and check for errors in the control files. MAKE-PLUS replaces Task 1 in prior releases.

Pre-Processing. A utility program, the Cohort Data Editor (EDT), is similar to Mod 3 in prior releases. The standard death rate files are preprocessed using a utility program, the Rate File Interface (RFI). This utility performs Task 2 in prior releases and is normally executed once during an analysis. The resultant RFI output files are preserved for access by other software components.

Processing. The processing phase contains three analytic modules (Modules 1, 2, and 4), which replace Task 3 of the prior releases.

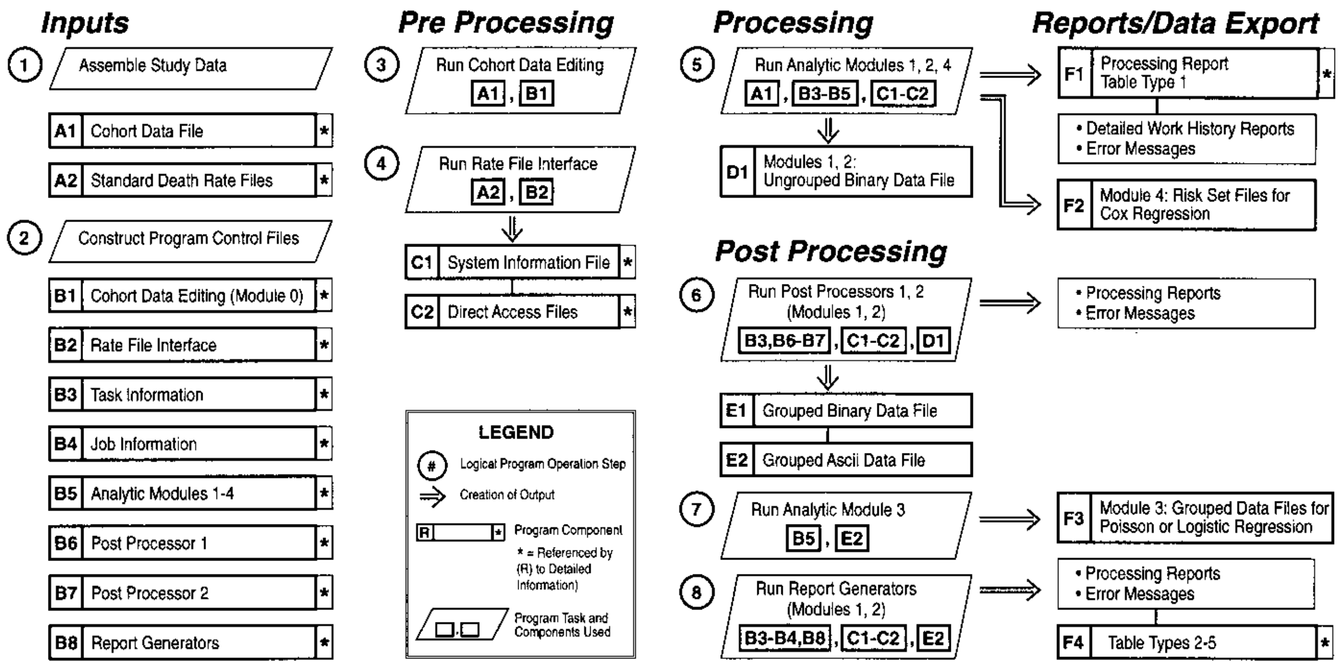
Inputs
Pre Processing
Processing
Reports/Data Export
1 Assemble Study Data
A1 Cohort Data File *
A2 Standard Death Rate Files *
2 Construct Program Control Files
B1 Cohort Data Editing (Module 0) *
B2 Rate File Interface *
B3 Task Information *
B4 Job Information *
B5 Analytic Modules 1-4 *
B6 Post Processor 1 *
B7 Post Processor 2 *
B8 Report Generators *
3 Run Cohort Data Editing A1, B1
4 Run Rate File Interface A2, B2
C1 System Information File *
C2 Direct Access Files *
LEGEND
# Logical Program Operation Step
Creation of Output
R * Program Component
* = Referenced by (R) to Detailed Information)
Program Task and Components Used
5 Run Analytic Modules 1, 2, 4 A1, B3-B5, C1-C2
D1 Modules 1, 2: Ungrouped Binary Data File
Post Processing
6 Run Post Processors 1, 2 (Modules 1, 2) B3,B6-B7, C1-C2, D1
E1 Grouped Binary Data File
E2 Grouped Ascii Data File
7 Run Analytic Module 3 B5, E2
8 Run Report Generators (Modules 1, 2) B3-B4,B8, C1-C2, E2
F1 Processing Report Table Type 1 *
• Detailed Work History Reports
• Error Messages
F2 Module 4: Risk Set Files for Cox Regression
• Processing Reports
• Error Messages
F3 Module 3: Grouped Data Files for Poisson or Logistic Regression
• Processing Reports
• Error Messages
F4 Table Types 2-5 *

Fig. 1. OCMAP-PLUS setup, execution, and output.

Post-Processing. The main post-processor programs (PP1, PP2) utilize the primary output (summarized case data) of the processing programs (Modules 1 and 2). This post-processing replaces Task 4 of the prior releases. Module 3 (OCMXLATE), which creates files for Poisson or logistic regression analysis, is also considered a post-processor.

Reports/Data Export. The Report Generators (RPT2-RPT5) use the primary output of the second post-processing program (PP2) and replace Task 5 of the prior releases. There is a separate reporting program for each Table Type 2 through 5 as defined in the prior releases. Table Type 1 (age distributions) is written as part of a general processing report produced from the processing program Module 1. Task 6 is performed by processing program Module 4 and post processing program Module 3. These programs create data files for relative risk regression analysis and Poisson or logistic regression analysis, respectively. Other utility software provided in OCMAP-PLUS includes:

1. Data Management Program (DMP)—converts binary intermediate files to the ASCII equivalent.
2. Rate File Dumper (RFD)—converts the binary RFI outputs to the ASCII equivalent.
3. Rate File Lister (RFL)—converts the binary RFI outputs to the ASCII equivalent.
4. PRN—selective printing from each report program's output file.

More Flexible Data Input

Like prior releases, the cohort data input file used by almost all processing modules of OCMAP-PLUS contains records of three types: Record Type 1, identification and demographic variables, Record Type 2, case selection variables, and Record Type 3, work history, and exposure variables. Several extensions have been made in OCMAP-PLUS:

Record Type 1. This record contains the basic data required by all analytic modules: record ID, sex, race, geographic area, cause of death, birth date, hire date, starting and stopping date for person-day counts (for Modules 1 and 4) and vital status. In addition to age group and time period, prior releases permitted a maximum of four strata defined by race (white or nonwhite) and sex (male or female). Multiple runs and supplemental programming were required to adjust for geographic area. OCMAP-PLUS permits a maximum of 128 strata defined by geographic area (32 levels), race and sex. The addition of geographic area as a stratification variable allows users to indirectly adjust summary rates for geographic area within a single run of the program.

Record Type 2. This record contains up to 24 user-specified non-time dependent variables. Record Type 2 allows subgroups of the cohort to be selected using the death rate-linking variables or other variables (eg, year of hire, year of birth, prior employment). In OCMAP-PLUS, the control file has been simplified so that users no longer need to supply the FORTRAN-type input format to process selection records and selection variables.

Record Type 3. This record contains detailed work history and exposure history data, which can be used in the analytic modules to examine cohort mortality in relation to duration of employment, job type and/or indicators of job exposure(s). Record

Type 3 contains one record for each job held during the employment history and, for each job, includes the starting and finishing dates, job type, and job exposure profile. Although the exposure profile can contain qualitative or quantitative exposure values for one to 99 agents, a maximum of eight agents can be considered in a given analysis. OCMAP-PLUS includes the following enhancements with respect to Record Type 3:

- For each job in the work history, its "type" can be described by up to eight components (eg, a typical four-level hierarchical scheme might include building, department, work area, and job title). Job definition and selection can be done in much greater detail than was previously possible.
- Job exposure fields within work history records are no longer a fixed length of 15 characters. The available options in OCMAP-PLUS are:
- *Option 1 (default)*: The exposure fields are all of the same length, which is user-specified. Different precisions can be attached to each exposure field by the placement of a decimal point.
- *Option 2:* The exposure fields vary in length, with the FORTRAN-type format specified in a user-defined file.

Marsh[13] describes a strategy, based on OCMAP Record Type 3, for merging and analyzing work history data in industry-wide occupational studies.

Extended Analytic Capabilities

OCMAP-PLUS contains four analytic modules (referred to as "Module"). Module 1 and Module 2 are enhanced versions of the Mod 1 and Mod 2 of prior releases. Modules 3 and 4 are new programs in OCMAP-PLUS. Mod 4–6 of prior releases are not included in the current version of OCMAP-PLUS, but many of their computational algorithms have been included in the current modules. This section describes the basic features of the four analytic modules and the new analytic capabilities of OCMAP-PLUS. Marsh[1,13,14] provides a more complete discussion of the statistical algorithms in OCMAP-PLUS.

Module 1—Cohort Mortality Rate Analysis. This module computes person-days at risk, internal cohort mortality rates, and indirectly standardized mortality ratios (SMRs) for subgroups of the cohort defined by one or more of the:

1. Standard death rate-linking variables (age, time period, race, sex, geographic area and cause of death),
2. Selection variables in Record Type 2,
3. Time-dependent work history variables (duration of employment, time since first employment), and
4. Time- and agent-specific exposure variables (duration of exposure, cumulative exposure and average intensity of exposure).

Storing the person-day counts on disk (for use by post-processing programs) eliminates the redundant person-day calculations in prior releases. Standard population death rates used in the calculation of SMRs can be specified by the user or obtained from the Mortality and Population Data System (MPDS) developed at the University of Pittsburgh.[15] Module 1 also computes 95% and 99% confidence intervals for SMRs using Poisson probabilities.[16]

Module 2—Proportional Mortality Analysis. Module 2 performs analogous computations to Module 1 for proportional mortality rather than mortality rates. Module 2 computes internal proportional mortality and indirectly standardized proportional mortality ratios (SPMRs) for the same subgroups defined for Module 1. Proportional mortality and SPMRs can be based on the distribution of total deaths or on any user-specified causes of death. Standard population proportional mortality can be specified by the user or obtained from the MPDS system. Module 2 computes 95% and 99% confidence intervals for SPMRs using the test-based approach of Miettinen.[17]

Module 3—Create Grouped Data Files for Poisson or Logistic Regression. The stratified analytic methods available in Modules 1 and 2 may not be adequate in studies that relate mortality risk to multiple exposure and/or potential confounding variables. Poisson regression[3,18] can be used to summarize data, to investigate the dependence of rates or SMRs on combinations of these variables, and to estimate adjusted risk (rate) ratios. Similar questions can be addressed for proportional mortality and SPMRs using logistic regression.[3]

In Poisson regression, the observed number of deaths in a particular cross-classification of the variables is assumed to follow a Poisson distribution with a mean that depends on the person-years at risk (for internal cohort rates) or the expected number of deaths (for SMRs), and the effects of the classification factors. A multiplicative model for the cohort rates is given by

$$\begin{aligned}&\mathit{log}\ E(\mathit{observed\ deaths})\\&\quad= \mathit{log}(\mathit{person\text{-}years}) + \beta'\boldsymbol{x}\end{aligned} \qquad [1]$$

and for the SMRs is

$$\begin{aligned}&\mathit{log}\ E(\mathit{observed\ deaths})\\&\quad= \mathit{log}(\mathit{expected\ deaths}) + \beta'\boldsymbol{x}\end{aligned} \qquad [2]$$

where E denotes statistical expectation, log denotes natural logarithm, β is a p-dimensional vector of regression coefficients to be estimated and **x** is the corresponding vector of covariates.

For proportional mortality analysis, the age-time specific proportion of deaths from a cause of interest can be modeled using a logistic regression model:

$$\mathit{logit}(p) = \beta'\boldsymbol{x} \qquad [3]$$

TABLE 1

OCMAP-PLUS Module 3 Output: Format of Grouped Data Files for Poisson Regression

The file created by Module 3 (OCMXLATE) is a space-delimited flat ASCII file with each record in the following format:

Field Number	Description	Location	Width	Type[*]	Notes
1	File ID	1:4	4	C or I	User supplied—four characters that uniquely identify the input file
2	Plant	6:7	2	I	Plant Code
3	Sex	9:9	1	I	Sex Code (1—Male, 2—Female)
4	Race	11:11	1	I	Race Code (1—White, 2—Nonwhite)
5	Latency	13:13	1	I	Latency Interval
6	Exposure Index 1	15:15	1	I	First Selected Exposure Interval
7	Exposure Index 2	17:17	1	I	Second Selected Exposure Interval
8	Exposure Index 3	19:19	1	I	Third Selected Exposure Interval
9	Exposure Index 4	21:21	1	I	Fourth Selected Exposure Interval
10	Exposure Index 5	23:23	1	I	Fifth Selected Exposure Interval
11	Exposure Index 6	25:25	1	I	Sixth Selected Exposure Interval
12	Exposure Index 7	27:27	1	I	Seventh Selected Exposure Interval
13	Exposure Index 8	29:29	1	I	Eighth Selected Exposure Interval
14	Age Time Recode	31:32	2	I	Age Time Recode
15	Time Index	34:35	2	I	Time Interval
16	Age Index	37:38	2	I	Age Interval
17	Number of Unknown Causes of Death	40:47	8	I	Number of Unknown Causes of Death
18	Number At Risk	49:56	8	I	Ignore—internal programming code
19	Age	58:65	8	I	Ignore—internal programming code
20	Time	67:74	8	I	Ignore—internal programming code
21	Cell	76:83	8	I	Ignore—internal programming code
22	Observed Number of Deaths—1	85:92	8	I	Observed Number of Deaths for First Selected Cause of Death
	.				.
	.				.
	.				.
XX	Observed Number of Deaths—*n*		8	I	Observed Number of Deaths for the nth Selected Cause of Death
XX	Expected Number of Deaths—1[†]		15	F	Expected Number of Deaths for First Selected Cause of Death
	.				.
	.				.
	.				.
XX	Expected Number of Deaths—*n*[†]		15	F	Expected Number of Deaths for the nth Selected Cause of Death
XX	Person Years[‡]		15	F	Accumulated Person-Year Counts (Module 1 only)
XX	Total Deaths		15	F	Observed Counts of Death for All Causes of Death (Module 2 only)

[*] I, Integer; C, Character; F, Floating Point.

[†] For Module 1, expected number of deaths based on standard population death rates. For Module 2, expected number of deaths based on standard population proportional mortality.

[‡] The output excludes records with zero person-years.

Comparisons to an external population can be made by including an offset term in the model:

$$logit(p) = logit(p^{*}) + \beta'\mathbf{x}. \qquad [4]$$

where p* denotes the corresponding age-time specific proportions in the standard population.

Maximum likelihood estimates of Poisson and logistic regression parameters can be obtained using available software such as the Generalized Linear Interactive Modeling (GLIM4) program,[19] EGRET,[20] and Epicure.[21]

In the past, we created the grouped data files for Poisson or logistic regression from the cohort data input file using separate programs. In OCMAP-PLUS Module 3 (OCMXLATE), these grouped data files are constructed from the output files of Post Processor 2 (PP2), which uses as input the output from Module 1 for Poisson regression files or Module 2 for logistic regression files. Both of these grouped data files are in a flat ASCII format suitable for input into other statistical packages. Table 1 provides an annotated listing of the variables (and format) created by Module 3.

Module 4—Create Risk Set Files for Relative Risk Regression Analysis. Continuous-time relative risk regression models provide an alternative analysis to Poisson regression of the internal cohort rates. The Cox proportional hazards model[22,23] and its generalizations are discussed in an occupational cohort setting in Breslow and Day.[3] The general form of the model is given by

$$\lambda(t) = \lambda_0(t)\, exp\{\boldsymbol{x}(t)\beta\} \qquad [5]$$

where $\lambda_0(t)$ is the hazard of an event at time t for an individual with baseline levels of all covariates, $\mathbf{x}(t)$ is a vector of covariates (exposures and/or potential confounding variables evaluated as of each event time t), and β is the corresponding parameter vector estimated by partial likelihood. The Cox model essentially compares the exposure(s) of the case (usually a cause-specific death) to the average exposure(s) of the members of the corresponding risk set (the case and the other cohort members alive and at risk at the time the case died). When there are multiple time-dependent covariates, the most feasible way to actually fit model [5] is to explicitly construct the risk sets and to estimate the parameters using a conditional logistic regression program such as EGRET,[20] Epicure,[21] or LogXact Turbo.[24] The algorithms used to enumerate these risk sets and compute the values of the time-dependent covariates at the appropriate times also provide matched sets for nested case-control studies.

OCMAP-PLUS, Module 4, contains four separate programs that are executed sequentially:

RISK-SET—From the input cohort file (Record Types 1–3), the cases are identified and the event times ascertained. Age is the primary time dimension. For each case, a risk set is defined as the case plus all cohort members alive and at risk at the age the case died. Individuals are not considered to be "at risk" prior to the age at which they meet the cohort entry criteria ("delayed entry"). Risk sets are matched within a specific range ("caliper matched") on year of birth. For each member of each risk set, each covariate $\mathbf{x}(t)$ is evaluated as of the age the corresponding case died. Table 2 provides an annotated listing of the variables (and format) created by RISK-SET.

RS-MATCH—The full risk sets can be further matched on values of variables from Record Types 1–3, with exact matching on categorical covariates, such as race, and caliper matching on continuous variables.

RS-RANDOM—The non-cases in each risk set can be sub-sampled at random, based on either a fixed sample size per risk set, as in a nested case-control study with 1:M matching, or a fixed sampling fraction, as in incidence density sampling.[4]

RS-STRIP—The user specifies the variables to be output in the flat ASCII file that contains the explicitly constructed risk sets.

Other New Features in OCMAP-PLUS

Algorithm to Allocate Person-Years for Persons of Unknown Race (Modules 1–3)

In prior releases, race-specific person-days were accumulated for only those study members of known race. Race information can be obtained from death certificates for deceased workers, but may be missing for workers who were alive at the end of follow-up and employed during time periods when such data were not routinely collected. When race information was unknown for only a small percentage of the cohort, and most of those with known race were white (or nonwhite), our past approach has been to assume that all study members were of the predominant race and calculate the expected numbers of deaths based on the corresponding race-specific standard mortality rates. Neither the prior releases of OCMAP nor other similar programs included the person-years of study members of unknown race in external mortality comparisons. OCMAP-PLUS now provides such an analytic approach to include study members of unknown race.

In OCMAP-PLUS, the person-years accumulated by study members of unknown race are assigned to the white or nonwhite categories in proportion to the person-year distributions of study members of known race, using a Proportional Allocation Method (PAM). For example, if a cohort with 1300 person-years included 800 white person-years, 200 nonwhite person-years and 300 person-years of unknown race, 800/1000 (80%) of the 300 unknown race person-years would be assigned to the white category, and 200/1000 or (20%) would be assigned to the nonwhite category, resulting in estimated total race-specific distributions of 1040 white and 260 nonwhite person-years. To reduce bias, this allocation is performed within strata defined by geographic area, sex, age group and time period. The resulting total numbers of allocated white and nonwhite person-years represent weighted averages of the stratum-specific assignments. Sparse data problems have precluded further stratification on other study variables, such as duration of employment or time since first employment.

Although the PAM does not impute race for individual study members, the person-years of study members of unknown race are included in the grouped external mortality comparisons, under the assumption that the person-years distributions of whites and nonwhites in a given stratum of workers with unknown race are proportional to the corresponding distributions based on persons of known race. A model-based iterative allocation of person-years developed by Youk[25] can be used in conjunction with standard OCMAP-PLUS output to impute race at an individual level.

TABLE 2

OCMAP-PLUS: Module 4 Output—Format of Risk Set File for Relative Risk Regression

The file created by RISK-SET is a space-delimited ASCII flat file with each record in the following format:

Field Number	Description	Location	Width	Type*	Notes
1	Case Number	1:6	6	I	Stratum Number
2	Case or Control	8:8	1	I	Case or Control (1—case, 2—control)
3	Record Number	10:15	6	I	Record Number from Cohort File
4	Sex	17:17	1	I	1—Male, 2—Female
5	Race	19:19	1	I	1—White, 2—Nonwhite
6	ICDA	21:24	4	C or I	ICDA Code, 4102 means 410.2
7	Birth Month	26:27	2	I	Date of Birth from Record Type 1
8	Birth Day	29:30	2	I	Date of Birth from Record Type 1
9	Birth Year	32:35	4	I	Date of Birth from Record Type 1
10	Hire Month	37:38	2	I	Date of Hire from Record Type 1
11	Hire Day	40:41	2	I	Date of Birth from Record Type 1
12	Hire Year	43:46	4	I	Date of Birth from Record Type 1
13	Start Month	48:49	2	I	Start Date from Record Type 1
14	Start Day	51:52	2	I	Start Date from Record Type 1
15	Start Year	54:57	4	I	Start Date from Record Type 1
16	Stop Month	59:60	2	I	Stop Date from Record Type 1
17	Stop Day	62:63	2	I	Stop Date from Record Type 1
18	Stop Year	65:68	4	I	Stop Date from Record Type 1
19	Vital Status	70:70	1	I	Standard OCMAP Coding of Vital Status
20	Plant	72:73	2	I	Plant Number from the Rate File Link
21	Event Date Month	75:76	2	I	Date the Control Reaches the Event Time
22	Event Date Day	78:79	2	I	Date the Control Reaches the Event Time
23	Event Date Year	81:84	4	I	Date the Control Reaches the Event Time
24	Entry Date Month	86:87	2	I	Date of Entry into the Study
25	Entry Date Day	89:90	2	I	Date of Entry into the Study
26	Entry Date Year	92:95	4	I	Date of Entry into the Study
27	Separation Date Month	97:98	2	I	Date of Separation
28	Separation Date Day	100:101	2	I	Date of Separation
29	Separation Date Year	103:106	4	I	Date of Separation
30	Vital Status	108:108	1	I	Dead or Alive (1—Dead, 0—Alive)
31	Working Status	110:110	1	I	Working Status at the Date of the Event (1—Working, 0—Not Working)
32	Days Since Last Worked	112:117	6	I	Number of Days Since They Last Worked
33	Plant at Event Time	119:122	4	I	Plant where Employed
34	Age	124:129	6	I	Age at Event Date
35	Latency	131:136	6	I	Latency in Days
36	Number of Selection Variables	138:140	3	I	Number of Selection Variables
37	Number of Exposures	142:144	3	I	Number of Exposures
38	Number of Average Exposures	146:148	3	I	Number of Average Exposures
XX	Selection Variables		*n**15	F	*n* is the Number of Selection Variables
XX	Exposures		*n**15	F	*n* is the Number of Selection Variables
XX	Average Exposures		*n**15	F	*n* is the Number of Selection Variables
XX	Time Since First Exposure		*n**6	I	*n* is the Number of Selection Variables
XX	Record ID		35	C	ID from Record Type 1
XX	End of Record Identifier		1	C	The "I" Character

* I, Integer; C, Character; F, Floating Point.

Time-Dependent Stratification by Active and Inactive Work Status (Modules 1–4)

In occupational cohort mortality studies, the time-dependent variable "employment status" (ES) can potentially confound the association between workplace exposure and health outcome (mortality). ES is necessarily related to exposure (only employed persons can receive workplace exposures), and ES may be related to the risk of death (either because a change in ES may signify ill health, or because being unemployed increases the risk of death).[26,27]

A new feature of Modules 1–4 enables the user to stratify analyses according to ES. In Module 1, for example, person-days of observation and observed deaths are categorized at any time point according to whether the study member was actively working in the study plant or inactive at that time. The stratification by ES is done concurrently with the other time-dependent variables. An optional switch allows the user to extend the "actively employed" status a period of k days beyond the termination date.

Incorporation of Multiple Time-Dependent Indices of Exposure (Modules 1–4):

Increased Number of Exposures Considered Simultaneously. Person-days of observation can be accrued simultaneously with respect to a maximum of eight exposures variables (and in conjunction with age, time, race, sex and geographic area, up to a maximum of 128 subgroups). Prior releases permitted only two simultaneous exposures.

New Exposure Indices. Like prior releases, OCMAP-PLUS permits the calculation of two summary measures of exposure, duration of exposure and time-weighted cumulative exposure. These time-dependent measures can be computed over the total work history or over subsets defined by job type. The duration of exposure measure for a given agent (Agent_Dur) computed over N_j jobs during exposure period j can be expressed as

$$Agent\_Dur_j = \sum_{i=1}^{N_j} Time_i \cdot Exp_i, \text{ where}$$

$$Exp_i = \begin{cases} 1 & \text{if exposed} \\ 0 & \text{if unexposed} \end{cases} \qquad [6]$$

where $Time_i$ and Exp_i represent the duration of employment and exposure level, respectively, of the ith job in the work history during exposure period j. Duration of employment can be viewed as a special case of this duration of exposure measure where $Exp_i = 1$ for all jobs in the work history.

For exposure agents measured quantitatively, the time-weighted cumulative exposure (Agent_Cum) is defined analogously:

$$Agent\_Cum_j = \sum_{i=1}^{N_j} Time_i \cdot Exp_i \qquad [7]$$

where Exp_i is the quantitative exposure value for the ith job.

In OCMAP-PLUS, a third summary exposure measure, average intensity of exposure, can be computed separately or in conjunction with duration of exposure and/or cumulative exposure. For exposure agents measured quantitatively, the average intensity of exposure (Agent_AIE), computed over N_j jobs during exposure period j, is calculated by dividing Agent_Cum by Agent_Dur or

$$Agent\_AIE_j = \frac{Agent\_Cum_j}{Agent\_Dur_j} \qquad [8]$$

Agent_AIE is the average intensity of exposure during periods when the worker is exposed to that agent. The default "exposure period" j described in[6–8] is the total work history (ie, the time interval from date of hire to date of termination from employment, accounting for gaps such as sick leaves, layoffs, strikes, etc).

The interpretation of internal rates or SMRs relative to the exposure measures in[6–8] relies on the implicit assumption that mortality during a given observation period is related to the duration, cumulative or average intensity of exposure received from hire date up to the point of observation. The actual "effectiveness" of an agent to cause disease may change over time, and may be negligible during parts of the work history. To enable some adjustment for this changing effectiveness, Module 1 also includes the lagging algorithm from the effective exposure modeling program in prior releases of Mod 4,[6] which can be applied to each of the summary exposure measures.[6–8] For each summary measure (SM), the lagging algorithm incorporates a binary weighting factor, which is a function of the time between exposure and subsequent risk of disease, to construct an effective summary measure (ESM) as

$$ESM_k = \sum_{j=1}^{k} w_{k-j+1} \cdot SM_j,$$

$$\text{where } w_{k-j+1} = \begin{cases} 0 \text{ when } (k-j+1) \le l \\ 1 \text{ when } (k-j+1) > l \end{cases} \qquad [9]$$

where ESM_k is the exposure potentially effective in the causation of disease (mortality) in time period k; SM_j is the summary exposure measure in exposure period j (Agent_Dur, Agent_Cum or Agent_AIE); w_{k-j+1} is the exposure weighting (lagging) factor; and l is the lagging period.

As an alternative adjustment for changing exposure effectiveness, Module 1 can accrue person-days of observation relative to a "moving average" intensity of exposure (MAIE). That is, rather than relating mortality risk to the "cumulative" AIE[6] computed across the entire work history, the MAIE relates mortality risk to the AIE received in a user-specified time interval j. In symbols,

$$Agent\_MAIE_{j'} = \frac{Agent\_Cum_{j'}}{Agent\_Dur_{j'}} \qquad [10]$$

where j′ is the user-specified exposure period. The MAIE can also be used in conjunction with the lagging algorithm.[9] Unlike Agent_Dur and Agent_Cum, which are monotonically increasing, Agent_AIE and Agent_MAIE can increase or decrease over time.

In Module 1, OCMAP-PLUS accrues person-days at risk for up to eight agents in terms of the summary exposure measures in,[6–10] concurrently with the other time dependent variables age, time period, duration of employment and the time since first employment.

TABLE 3
University of Pittsburgh Mortality and Population Data System (1995): ICDA Default Cause of Death List—63 Causes

Label	Cause of Death	6th & 7th Revision (1950–1967)	8th Revision (1968–1978)	9th Revision (1979+)
01	All Causes of Death	001–999	000–999	001–999
02	Tuberculosis	001–019	010–019	010–018
03	All Malignant Neoplasms	140–205	140–209	140–208
04	Buccal Cavity and Pharynx	140–148	140–149	140–149
05	Digestive Organs and Peritoneum	150–159	150–159	150–159
06	Esophagus	150*	150	150*
07	Stomach	151	151	151*
08	Large Intestine	153	153	153*
09	Rectum	154	154	154*
10	Biliary Passages and Liver Primary	155	155, 156	155, 156*
11	Pancreas	157*	157	157*
12	All Other Digestive	152, 156, 158, 159*	152, 158, 159	152, 158, 159*
13	Respiratory System	160–164	160–163	160–165
14	Larynx	161*	161*	161*
15	Bronchus, Trachea, Lung	162, 163	162	162*
16	All Other Respiratory	160, 164	160, 163	160, 163, 164, 165*
17	Breast	170	174	174, 175
18	All Uterine (female only)	171, 172–174	180, 181, 182.0*, 182.9	179, 180, 181, 182*
19	Cervix (female only)	171	180*	180*
20	Other Female Genital Organs (female only)	175, 176	183–184*	183–184*
21	Prostate (male only)	177	185	185*
22	Testis and Other Male Genital Organs (male only)	178, 179*	172.5, 173.5, 186*, 187	186, 187*
23	Kidney	180	189.0, 189.1, 189.2	189.0, 189.1, 189.2
24	Bladder and Other Urinary Organs	181	188, 189.9	188, 189.3, 189.4, 189.8, 189.9
25	Malignant Melanoma of Skin	190	172.0–172.4*, 172.6–172.9	172*
26	Eye	192*	190*	190*
27	Central Nervous System	193*	191, 192*	191, 192*
28	Thyroid Gland and Other Endocrine Glands and Related Structures	194, 195*	193, 194*	193, 194*
29	Bone	196	170*	170*
30	All Lymphatic and Hematopoietic Tissue	200–205	200–209	200–208*
31	Lymphosarcoma and Reticulosarcoma	200	200	200*
32	Hodgkin's Disease	201	201	201*
33	Leukemia and Aleukemia	204	204–207	204–208
34	All Other Lymphopoietic Tissue	202, 203, 205	202, 203, 208, 209*	202, 203*
35	All Other Malignant Neoplasms	165, 191, 197–199	171, 173.0–173.4*, 173.6–173.9, 195–199	171, 173*, 195–199
36	Benign Neoplasm	210–239	210–239	210–239
37	Diabetes Mellitus	260	250	250
38	Cerebrovascular Disease	330–334	430–438	430–438
39	All Heart Disease	400–402, 410–443	390–398, 400.1, 400.9, 402, 404, 410–414, 420–429	390–398, 402, 404, 410–429
40	Rheumatic	400–402, 410–416	390–398	390–398
41	Ischemic	420, 422.1	410–414	410–414
42	Chronic Disease of Endocardium and Other Myocardial Insufficiency	421, 422.0, 422.2	424, 428	424, 428*
43	Hypertension with Heart Disease	440–443	400.1, 400.9, 402, 404	402, 404*
44	All Other Heart Disease	430–434	420–423, 425–427, 429	415–417, 420–423, 425–427, 429*
45	Hypertension w/o Heart Disease	444–447	400.0, 400.2, 400.3, 401, 403	401, 403, 405
46	Nonmalignant Respiratory Disease	241, 470–527*	460–519*	460–519*
47	Influenza and Pneumonia	480–483, 490–493	470–474, 480–486	480–487
48	Bronchitis, Emphysema, Asthma	501, 502, 527.1, 241*	490–493	490–493*
49	Bronchitis	501, 502	490, 491	490, 491
50	Emphysema	527.1	492	492
51	Asthma	241	493	493
52	Other Nonmalignant Respiratory	470–475, 500, 510*–527.0, 527.2	460–466*, 500–519	460–466*, 470–478, 494–496, 500–519

* Comparability ratios (CR) unavailable at present time. Rates are unadjusted (CR = 1.0) for these causes.

Mortality and Population Data System (MPDS)

Since 1980, the Department of Biostatistics at the University of Pittsburgh has maintained a data repository and retrieval system for detailed mortality data provided by the National Center for Health Statistics and the US Census Bureau. This Mortality and Population Data System (MPDS) contains the underlying cause of death code (using International Classification of Diseases (ICD) four-digit codes) for all per-

TABLE 3—Continued

University of Pittsburgh Mortality and Population Data System (1995): ICDA Default Cause of Death List—63 Causes

Label	Cause of Death	6th & 7th Revision (1950–1967)	8th Revision (1968–1978)	9th Revision (1979+)
53	Ulcer of Stomach and Duodenum	540, 541	531–533	531–533
54	Cirrhosis of Liver	581	571	571
55	Nephritis and Nephrosis	590–594*	580–584	580–589
56	All External Causes of Death	800–999*	800–999	E800–999*
57	Accidents	800–962	800–949	E800–949
58	Motor Vehicle Accidents	810–835	810–823	E810–825
59	All Other Accidents	800–802, 840–962	800–809, 824–949	E800–807, E826–949
60	Suicides	963, 970–979	950–959	E950–959
61	Homicides and Other External	964, 965, 980–999*	960–978*, 980–999	E960–978*, E980–999
62	All Other Causes	020–138, 206–207,* 240, 242–254, 270–326, 340–398, 450–468, 530–539, 542–580, 582–587, 600–795	000–009*, 020–136, 240–246, 251–389, 440–458, 520–530, 534–570, 572–577, 590–796	001–009*, 020–139, 240–246, 251–389, 440–459, 520–530, 534–570, 572–579, 590–799
63	Acquired Immunodeficiency Syndrome (AIDS) (1987–1992 only)	not applicable	not applicable	042–044, 795.8*

* Comparability ratios (CR) unavailable at present time. Rates are unadjusted (CR = 1.0) for these causes.

TABLE 4

Observed (Obs) Deaths and Standardized Mortality Ratios (SMRs) for Selected Causes of Death for Short-Term and Long-Term Workers, N-Cohort Males, Local County Comparison, 1946–1989*

	N-Cohort			
	Short-Term Worker		Long-Term Worker	
Cause of Death (ICD 8th Revision Codes)	Obs	SMR	Obs	SMR
All causes	277	110	408	97
All malignant neoplasms	73	125	108	108
Respiratory system (RSC)	33	143	35	98
NMRD×IP	13	131	23	116
Cirrhosis of liver	7	108	6	81
Nephritis and nephrosis	4	201	6	190
External causes	34	96	25	85

* Adopted from Marsh et al (1996), reference 28.

sons who died in the United States between 1950 and 1994 (limited to deaths from malignant neoplasms for the 1950–1961 period). Individual death records include codes for sex, race, age of death, year of death, and geographic location (county and state of residence at time of death).[15]

In MPDS, individual death records are categorized and linked with the corresponding population data to form death rates specific for five-year age groups, five-year time periods, race (white and non-white), sex, geographic location and cause of death. Cause of death can be defined by any individual ICD code or combination of ICD codes. Table 3 provides a listing of the standard 63 cause of death categories that are suitable for most cohort analyses. The listing shows the ICD codes for the sixth through ninth revisions, and indicates which categories can be made comparable to any one revision using comparability ratios (CR) provided by the NCHS. This approach, in which death rates are "adjusted" to a specific base ICD revision, is appropriate for studies that code all deaths to the base revision. MPDS death rates are also available in unadjusted form. Such rates are appropriate for studies that code all deaths to the revision of the ICD in effect at the time of death.

The MPDS standard death rate files can be written to OCMAP-PLUS format specifications and input directly as standard population data in comparative mortality analyses. The MPDS data base is updated annually as new NCHS data are released. According to NCHS policy, detailed mortality and population data for counties of population size less than 100,000 cannot be distributed. To our knowledge, the MPDS is the most comprehensive and accessible data repository and retrieval system of its kind.

Examples from Recent Historical Cohort Studies

As part of our ongoing man-made vitreous fiber (MMVF) worker cohort mortality study at the University of Pittsburgh, Marsh et al[28] reported the 1946 to 1989 mortality experience of 3,035 workers employed at least one year between 1945 and 1978 in one or more of five rock wool or slag wool (RSW) manufacturing facilities in the United States. This study includes a nested case-control study to examine the relationship between respiratory system cancer (RSC) and exposure to RSW fibers with adjustment for the potential confounding factors (year of hire, plant, and smoking status) and co-exposures (up to seven other agents).

In the MMVF study, we used OCMAP-PLUS Module 1 to compute total and cause-specific internal cohort death rates and SMRs. The output from Module 1 for selected causes of death is illustrated in Table

TABLE 5

Estimated Risk Ratios for Respiratory System Cancer Mortality by Cumulative Exposure to Respirable Fiber (RFib_Cum) With and Without Adjustment for Smoking History, N-Cohort Males*

		Case-Control Study Data (1970–1989)[‡]	
RFib_Cum (fibers/cc-mos.)	Full Risk Sets (1946–1989)[†] (68 cases, 8017 controls)	Unadjusted for Smoking (54 cases, 107 controls)	Adjusted for Smoking[§] (54 cases, 101 controls)
<3	1.00	1.00	1.00
3–14	.72	.70	.64
15–39	.94	.59	.55
40+	.83	.71	.58
(Global *P* value)	(0.79)	(0.76)	(0.64)

* Adopted from Marsh et al (1996), reference 28.

[†] Matched on year of birth.

[‡] Adjusted for year of birth.

[§] Smoking categorized as ever/never; 6 controls with unknown smoking excluded.

TABLE 6

Summary of Unadjusted and Adjusted Risk Ratios for Cancer of the Lung by Duration (years) of Exposure to Formaldehyde ≥ .2 ppm, Wallingford White Male Cohort, 1945–1984*

				Adjusted Risk Ratio Estimates					
Duration of Exposure	Observed Deaths	SMR[†]	Unadjusted Ratio[‡] of SMRs	SMR Based[§]	Global *P* Value	Trend *P* Value	Cohort Rate Based[‖]	Global *P* Value	Trend *P* Value
0	41	93	1.00	1.00	.05	.02	1.00	.03	.01
Under .5	34	166	1.78	1.74			1.64		
.5–	22	145	1.56	1.57			1.54		
5+	16	170	1.83	1.80			2.21		

* Adopted from Marsh et al (1996), reference 29.

[†] SMR internally adjusted for age group and calendar time period.

[‡] Relative to baseline category.

[§] All models adjusted for year of hire and time since first employment via Poisson regression.

[‖] All models adjusted for age, time, year of hire, and time since first employment via Poisson regression.

4. In this example, observed numbers of deaths and the SMRs are shown separately for short-term (duration of employment <5 years) and long-term (duration of employment ≥5 years) male workers. The short-term contribution of the long-term workers, which is necessarily mortality-free, has been excluded from this analysis.

Table 5 shows the estimated risk ratios (RRs) for RSC by a time-dependent cumulative exposure (Agent_Cum) to respirable RSW fiber. The RRs are shown for the full risk sets and for the case-control data (limited to 1970–1989), with and without adjustment for smoking history. The OCMAP-PLUS Module 4 program RISK-SET was used to enumerate the risk sets, matched on year of birth, and compute the appropriate time-dependent covariates at each event time. RS-RANDOM was used to subsample from the full risk sets and select two controls for each case in the nested case-control study. The parameters in these relative risk regression models were estimated using the conditional logistic regression program in EGRET.[20]

A recent cohort mortality study of formaldehyde-exposed workers[29] illustrates the use of the Module 3 grouped data file export feature for Poisson regression analysis. This study examined the 1945–1984 mortality experience of 6,039 male workers with emphasis on the relationship between formaldehyde exposure (with or without particulate exposure) and lung cancer. Table 6 shows observed deaths and SMRs for lung cancer by four categories of a time-dependent duration of exposure (Agent_Dur) to formaldehyde greater than 2 parts per million (ppm). Also shown are estimated risk ratios for lung cancer based on both SMRs and internal cohort rates, with adjustments made for several potential confounders in the Poisson regression models. These models were fit using GLIM4.[19]

Hardware and Software Specifications

OCMAP-PLUS is written in FORTRAN-77, except for Modules 3 and 4, which are written in ANSI-C. OCMAP-PLUS can be used on an IBM-compatible DOS-based PC (IBM, White Plains, NY) with a 386 (or higher) processor, math coprocessor and Smartdrive capability. OCMAP-PLUS requires 640K RAM and a hard drive sufficiently large to allow an approximate 5:1 ratio of free space to system size (data file, program executables and rate files). OCMAP-PLUS can also be compiled for use on VAX-based mainframes and UNIX-based workstations. These versions of OCMAP-PLUS are compiled on individual machines, due to differing system requirements and machine architecture. A version of OCMAP-PLUS for Microsoft® (Redmond, WA) Windows-based operating systems is under development.

Availability of OCMAP-PLUS and the MPDS Data

OCMAP-PLUS is very flexible and easy to use. It is useful for biostatisticians and epidemiologists engaged in occupational or environmental health research, as well as other allied health professionals with some knowledge of statistics or computer programming (eg, physicians, industrial hygienists, engineers, technicians). OCMAP-PLUS is suitable for classroom use to give students "hands on" experience in courses on industrial and environmental epidemiology or biostatistics.

The OCMAP-PLUS package includes a user manual, program executable code, test data, sample control and rate files, and a setup program for hard drive installation. Single or multiple user licenses and optional maintenance programs can be purchased from the Department of Biostatistics. Discounts are available to non-profit organizations and to current OCMAP users. MPDS data can be obtained either as text (hard copy) or in machine-readable OCMAP-PLUS format for a nominal cost. Persons interested in obtaining OCMAP-PLUS and/or MPDS data should contact the first author (G.M.).

Acknowledgments

The development of the OCMAP-PLUS and MPDS programs was supported, in part, by a research contract between the University of Pittsburgh, Department of Biostatistics, and the North American Insulation Manufacturers Association (Alexandria, VA). Other funding is provided by licensing and service fees from outside users of the two programs. The authors wish to acknowledge the OCMAP/MPDS user community and the many OCMAP-PLUS beta testers for their continued interest and support of these programs. The authors also wish to thank Ms Linda Kobus for her administrative assistance on both the OCMAP-PLUS and MPDS projects and Ms Christine Gause who served as an "in-house" beta-tester of OCMAP-PLUS.

References

1. Marsh GM. Epidemiology of occupational diseases. In: Rom WN, ed. *Environmental and Occupational Medicine*, 2nd ed. Boston: Little, Brown; 1992:35–50.
2. Checkoway H, Pearce NV, Crawford-Brown DJ. *Research Methods in Occupational Epidemiology*. New York: Oxford University Press; 1989.
3. Breslow NE, Day NE. The design and analysis of cohort studies. In: *Statistical Methods in Cancer Research*, vol II. [IARC Scientific Publications No. 82.] Lyon, France: International Agency for Research on Cancer; 1987.
4. Breslow NE, Day NE. The design and analysis of case-control studies. In: *Statistical Methods in Cancer Research*, vol I. [IARC Scientific Publications No. 32.] Lyon, France: International Agency for Research on Cancer; 1980.
5. Marsh GM, Preininger ME. OCMAP. A user-oriented occupational cohort mortality analysis program. *Am Statistician.* 1980;34:245–246.
6. Caplan RJ, Marsh GM, Enterline PE. A generalized effective exposure modeling program for assessing dose-response in epidemiologic investigations. *Comput Biomed Res.* 1984;16:587–596.
7. Marsh GM, Ehland J, Paik M, Preininger M, Caplan R. A user oriented cohort mortality analysis program for the IBM PC. *Am Statistician.* 1986;40:308–309.
8. Marsh GM, Co-Chien H, Rao BR, Ehland J. OCMAP Module 6—a new computing algorithm for proportional mortality analysis. *Am Statistician.* 1989;43:127–128.
9. Scientific Citations Index, Institute of Science International, 1996. [Searchable database.]
10. Monson RR. Analysis of relative survival and proportional mortality. *Comput Biomed Res.* 1974;7:325–332.
11. Waxweiler RJ, Beaumont JJ, Henry JA, et al. A modified life-table analysis program system for cohort studies. *J Occup Med.* 1983;25:115–124.
12. Beaumont JJ, Steenland K, Minton A, Meyer S. A computer program for incidence density sampling of controls in case-control studies nested within occupational cohort studies. *Am J Epidemiol.* 1989;129:212–219.
13. Marsh GM. A strategy for merging and analyzing work history data in industry-wide occupational epidemiology studies. *Am Ind Hyg Assoc J.* 1987;48:414–419.
14. Marsh GM. Basic occupational epidemiologic measures. In: Bang KM, ed. *Occupational Medicine: State of the Art Reviews*. vol 11. Philadelphia: Hanley & Belfus; 1996:421–431.
15. Marsh GM, Ehland J, Sefcik S, Alcorn C. *Mortality and Population Data System (MPDS)*. [Department of Biostatistics Technical Report.] Pittsburgh, PA: University of Pittsburgh; 1996.
16. Bailar JC, Ederer F. Significance factors for the ratio of a Poisson variable to its expectation. *Biometrics.* 1964;20:639–642.
17. Miettinen OS. Estimability and estimation in case-referent studies. *Am J Epidemiol.* 1976;103:226–235.
18. Frome EL, Checkoway H. Use of Poisson regression models in estimating rates and ratios. *Am J Epidemiol.* 1985;121:309–323.
19. Francis B, Green M, Payne C. *The GLIM System, Release 4 Manual.* Oxford: Clarendon Press; 1993.
20. *EGRET, Version 0.26.6.* Seattle, WA: Statistics and Epidemiology Research Corporation; 1991.
21. *EPICURE, User's Guide.* Seattle, WA: HiroSoft International Corporation; 1988–93.
22. Cox DR. Regression models and life tables [with discussion]. *J R Stat Soc.* 1972;34B:187–220.
23. Cox DR. Partial likelihood. *Biometrika.* 1975;62:269–276.
24. *LogXact Turbo—Software for Exact Logistic Regression.* Cambridge, MA: CYTEL Software; 1993.
25. Youk AO. *Iterative Allocation of Partially Classified Data in Occupational Epidemiologic Studies.* [Department of Biostatistics Doctoral Dissertation.] Pittsburgh, PA: University of Pittsburgh; 1996.
26. Pearce N. Time-related confounders and intermediate variables. *Epidemiology.* 1992;3:279–281.
27. Steenland K, Stayner L. The importance of employment status in occupational cohort mortality studies. *Epidemiology.* 1991;2:418–423.
28. Marsh GM, Stone RA, Youk AO, et al. Mortality among United States rock wool and slag wool workers: 1989 update. *J Occup Health Safety Aust NZ.* 1996;12:297–312.
29. Marsh GM, Stone RA, Esmen NA, Henderson VH, Lee KY. Mortality patterns among chemical workers in a factory where formaldehyde was used. *Occup Environ Med.* 1996;53:613–617.