

## SAS<sup>®</sup> Programs for Adding Functionality to the Occupational Mortality Analysis Program (OCMAP) Software

Robert S. Matthews, University of Alabama at Birmingham, Birmingham, Alabama  
David A. Brown, PhD, Wyeth Research, Pittsburgh, Pennsylvania

### Abstract

*The OCMAP-PLUS program is a software package comprised of several component modules. This software allows the user to perform many of the conventional statistical analyses associated with epidemiologic research studies, particularly those among occupational groups. We have been using this software for more than a decade and have developed several extensive pre- and post- processing modules in SAS for use in working with OCMAP software. The process of manually abstracting and transforming the raw output from OCMAP to create corresponding Microsoft Word tables was enormously time consuming, resource exhausting and prone to errors. This paper describes the automated processes we have developed, using SAS/AF software, for using OCMAP to generate several types of standardized mortality ratio (SMR) tables, extracting the data from these tables into a SAS dataset, and directly exporting the results into a MS Word document. We estimate that this automated procedure reduced the amount of time needed to produce approximately 60 MS Word tables from several weeks to less than 1 hour. Although our use was specific to occupational epidemiology research, it can be readily applied to a variety of other applications.*

### Introduction

We have utilized the OCMAP software on a variety of research projects for both mortality and cancer incidence studies. One of the primary types of tables that we produce from this software is a table in Microsoft Word that shows, for each cause of death, the observed and expected numbers of death, the SMR, and the 95% lower and upper confidence intervals. The final tables that we generate are typically in one of two forms. The first form is typified by figure 3; a single OCMAP run is used to create a table that has five columns. One column is defined for each race/gender combination and a total column for the combined race/gender groups. The output for these five columns comes from five distinct tables in the OCMAP output (table type 4). The other type of table template that we typically create is one that is composed of the same race/gender group from different cohort subgroups. This involves doing separate OCMAP runs on different sub cohorts (typically through the use of a selection variable) and then extracting the desired table from each individual OCMAP run (see figure 2).

Creating these final tables manually is a very laborious, error-prone process. Several of the cause of death categories in the final tables are defined based on the values from multiple categories in the

OCMAP output. Combining the values and having to recheck and recompute the confidence intervals was very time-consuming and typically involved multiple passes to identify any typographical or computational errors.

Hence, the reason for developing a fully automated system of programs that eliminates having to rekey any data from the OCMAP runs into a Microsoft Word document. The intermediate step in going from OCMAP to MS Word was to create a SAS dataset. This SAS dataset can then be used for other purposes, such as producing graphs or other types of listings.

The programs that we have developed allow the user to run a series of OCMAP analyses, execute a SAS/AF screen to select the desired tables, and, by pressing a single button, automatically create a SAS dataset of the results and a corresponding Microsoft Word document. Figure 1 shows a flowchart of all the components necessary to go from a cohort file to a final MS Word document.

### Details

SAS/AF software was used to create a user-friendly front-end for the table conversion program. The screen (see figure 4) has several tabs for easy

navigation and allows the user to specify different parameters to produce the final document.

Individual tables within a single OCMAP run are stratified on plant/race/sex/latency/age/time and exposure variables. The AF screen allows the user to select which OCMAP tables should be placed into each column in the MS Word document. For example, Figure 2 shows a partial printout of an OCMAP run for the total values in each stratification level. In Figure 3, these data have been restructured and placed into the last column of the table. Restructuring the raw data is important because data can be combined from multiple OCMAP runs, rounding is applied to each value based on a set of user-defined rules, and computation of aggregate categories is applied. This allows hundreds of pages of output to be reduced into a compact, readable, word-processing document.

### **Translation Table dataset**

A key part of the conversion process involves the use of a SAS dataset that contains a separate observation for each cause of death category in an OCMAP run. Variables in this dataset include the text description of each cause of death category, the corresponding text for the MS Word table, indentation values for OCMAP and Word, and a selection variable to indicate whether or not each individual cause of death categories should be included in the Word table. Several cause of death fields are unique to a particular gender (e.g. prostate cancer for men). There is a variable in the translation table to indicate these gender specific causes of death.

### **Parsing Algorithm**

After a table is read in, three passes are made through data. The first is to extract all the fields that are identified in the translation dataset. The second pass handles any computed fields, and the final pass applies rounding to each value.

The variable in the translation dataset named 'ocmapcod' is used to identify each cause of death that will be parsed when an exact match is found between the variable text and the input record from the raw OCMAP table type 4 output. When a matching record is found, the observed number of deaths, expected number of deaths, the SMR and the upper and lower 95% confidence intervals are extracted from their respective columns. These values are stored in a two-dimensional array and each cause

of death specified in the translation table is scanned and stored in the array. After all values have been read, the array is read from the beginning and each line is written to the user-specified SAS dataset. Two datasets are generated each time the data is exported.

The first dataset contains the data in the format needed for the MS Word table and contains three observations for each cause of death. The first observation contains the observed and expected numbers, the second contains the SMR and the third contains the confidence intervals.

The second dataset transforms the data generated from the first dataset and places all the data for a given cause of death into a single observation.

### **Final Table**

The last step in generating a final Word document is to run a Word macro that converts the tab-delimited SAS output into a Word table. This macro also adjusts the column widths and handles some other formatting issues.

### **Conclusion**

The software programs we have developed for transforming OCMAP output into SAS datasets and Microsoft Word documents has proved invaluable for some of our recent analyses. We subsequently developed a procedure to graph our results and will continue to add other functionality as time and opportunity allows.

### **Additional Information**

OCMAP-Plus is written and supported by the Department of Biostatistics at the University of Pittsburgh.

For more documentation about how these programs work or to request a copy of the AF frame and SCL code, contact Robert Matthews at one of the following addresses:

Robert Matthews  
University of Alabama at Birmingham  
Department of Epidemiology and International Health  
1665 University Blvd. RPHB 517C  
Birmingham, AL 35294-0022  
rsm@uab.edu or <http://www.epi.soph.uab.edu/rsm/>

Dr. David Brown  
brownd6@wyeth.com

**Flowchart showing the steps involved in converting  
OCMAP output to a SAS dataset and a MS Word table**

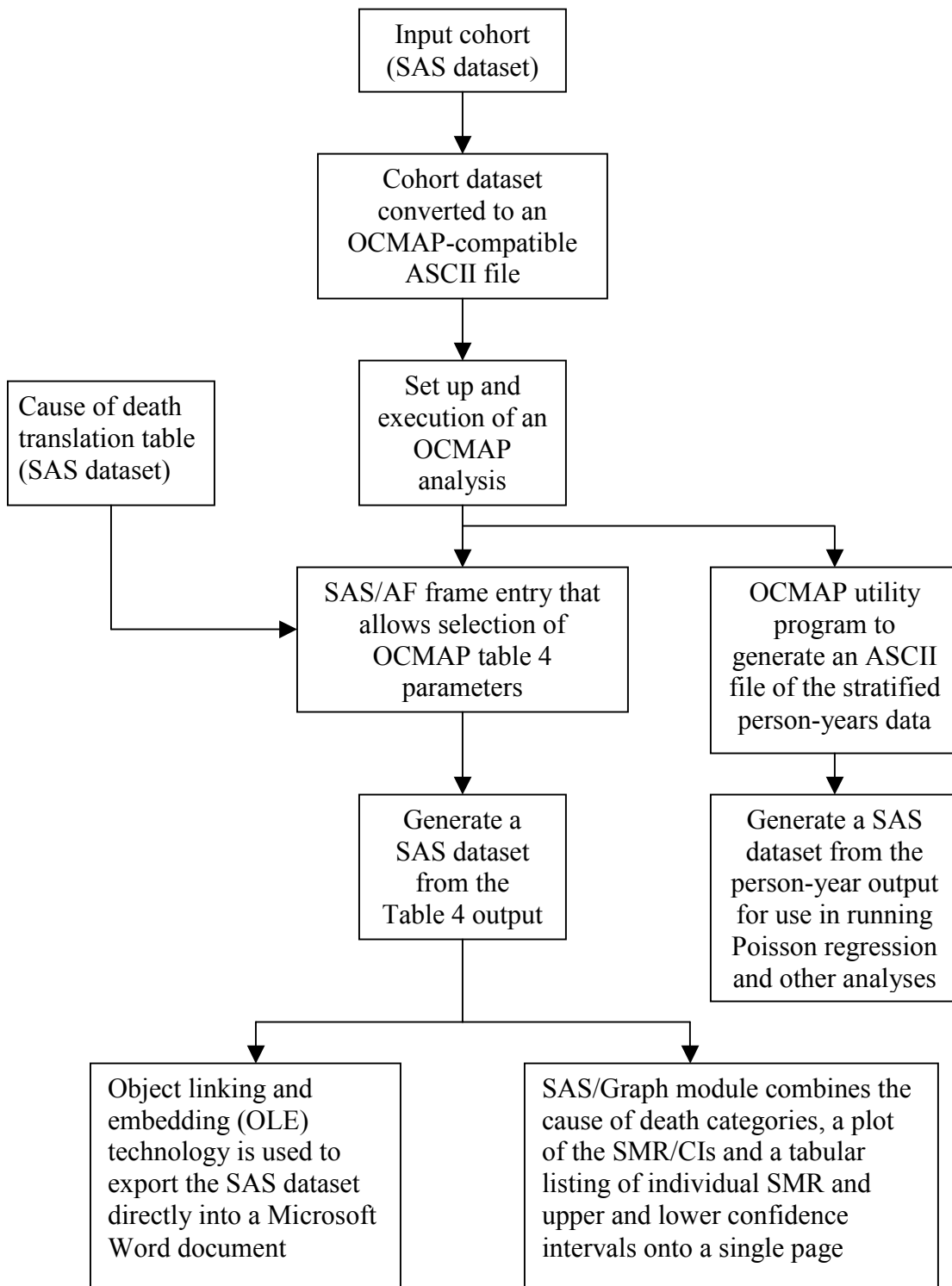


Figure 1

# Original OCMAP output

Project: Generic study  
 Cohort: Final master cohort  
 Rates: US rates  
 Exposures: Exposure type 2, years

Plant: 01  
 Race: TOTAL  
 Sex: TOTAL  
 Latency: TOTAL  
 Age: TOTAL  
 Time: TOTAL

Exposure 1: TOTAL  
 Exposure 2: TOTAL  
 Exposure 3: TOTAL  
 Exposure 4: TOTAL  
 Exposure 5: TOTAL  
 Exposure 6: TOTAL  
 Exposure 7: TOTAL  
 Exposure 8: TOTAL

No. at Risk: 85648 Pers Yrs: 1419021.0

Standardized Mortality Ratio Analysis: Table Type 4 - CAUSE SPECIFIC SMRS AND CONFIDENCE INTERVALS

CAUSE OF DEATH (Based on Revision at Time of Death)	OBS		EXP		SMR		LOWER		UPPER		95% LIMITS		99% LIMITS	
	---	---	---	---	---	---	---	---	---	---	---	---	---	---
All Causes of Death	11413	16886.84	67.6	**	66.4	68.8	66.0	69.2						
All Malignant Neoplasms	3766	4859.21	77.5	**	75.0	80.0	74.3	80.8						
Cancer of Buccal Cavity & Pharynx	59	116.88	50.5	**	38.4	65.1	35.2	70.0						

Figure 2

# Final Microsoft Word table

Causes of Death

FILE:	OUT0200	OUT0200	OUT0200	OUT0200	OUT0200	OUT0200	OUT0200	OUT0200	OUT0200
RACE:	WHITE	NONWHITE	WHITE	NONWHITE	WHITE	NONWHITE	WHITE	NONWHITE	TOTAL
SEX:	MALE	MALE	FEMALE	FEMALE	FEMALE	FEMALE	FEMALE	FEMALE	TOTAL
N:	64480	4094	15170	1904	85648	1419021	85648	1419021	85648
PY:	1127369	57581.2	212264.9	21806.4	212264.9	21806.4	212264.9	21806.4	1419021.0
Obs/Exp	10096/14781	288/668	967/1341	62/98	11413/16887	11413/16887	11413/16887	11413/16887	11413/16887
SMR	68	43	72	64	68	68	68	68	68
95% CI	67-70	38-48	68-77	49-81	66-69	66-69	66-69	66-69	66-69
Obs/Exp	3225/4189	101/155	411/488	29/27	3766/4859	3766/4859	3766/4859	3766/4859	3766/4859
SMR	77	65	84	107	78	78	78	78	78
95% CI	74-80	53-79	76-93	72-153	75-80	75-80	75-80	75-80	75-80
Obs/Exp	54/103	2/7.3	3/6.1	0/0.5	59/117	59/117	59/117	59/117	59/117
SMR	52	27	49	0	51	51	51	51	51
95% CI	39-68	3-99	10-144	0-760	38-65	38-65	38-65	38-65	38-65

Figure 3

OCMAP output and translation table      Printing options and help topics      Column definitions      Help topics

	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5
OUTname:	OUT0200	OUT0200	OUT0200	OUT0200	OUT0200
Plant:	01	01	01	01	01
Race:	WHITE	NONWHITE	WHITE	NONWHITE	TOTAL
Gender:	MALE	MALE	FEMALE	FEMALE	TOTAL
Latency:	TOTAL	TOTAL	TOTAL	TOTAL	TOTAL
Age:	TOTAL	TOTAL	TOTAL	TOTAL	TOTAL
Time:	TOTAL	TOTAL	TOTAL	TOTAL	TOTAL
Exp 1:	TOTAL	TOTAL	TOTAL	TOTAL	TOTAL
Exp 2:					
Exp 3:					
Exp 4:					

**Range of OUT files**  
 start      end  
     

Multiple datasets

Welcome to the OCMAP Table 4 to MS Word conversion program!  
 Fill in the columns on this screen from the header information at the beginning of each Table 4. Case is unimportant. The "Export Data" option can be used to export data directly into a Microsoft Word document and/or a SAS dataset.

Figure 4