

A SAS® Program for the Computation of Cumulative Exposure Estimates for Analysis of Cause-specific Mortality Rates in a Cohort of Synthetic Rubber Workers

Ilene Brill, University of Alabama at Birmingham, Birmingham, AL

Maurizio Macaluso, University of Alabama at Birmingham, Birmingham, AL

Robert Matthews, University of Alabama at Birmingham, Birmingham, AL

Elizabeth Delzell, University of Alabama at Birmingham, Birmingham, AL

ABSTRACT

The purpose of this SAS macro in a Windows environment is to generate SAS data sets containing person-year and numerator data for use in stratified and regression analyses of mortality rates and ratios. The data layout consists of aggregate person-year and death counts cross-classified according to categories of fixed and time-dependent variables, and of cumulative exposure to butadiene, styrene, and benzene. The program was employed to evaluate the association of chemical exposure and specific causes of death among workers employed in the manufacture of synthetic rubber. Two files containing cumulative exposure indices are created, a decedent file containing one observation for each decedent and a person-year file which contains one observation for each person year of follow-up. Relatively straightforward extensions of this program allow the dynamic allocation of person-years according to lagged exposure variables. The output data are easily employed in stratified and regression analyses of rates. Combination of the output person years data with standard mortality rates leads to the computation of cause-specific expected numbers of deaths for the analysis of mortality ratios. Application of the program is illustrated with runs from a study of causes of death among synthetic rubber workers.

INTRODUCTION

We carried out a retrospective follow-up study of about 15,000 men employed in six styrene-butadiene rubber (SBR) manufacturing plants in the USA and Canada.¹ The purpose of the study was to evaluate the possible carcinogenic effects of exposure to the monomers used to make SBR, 1,3-butadiene (BD) and styrene (STY), and to benzene (BZ). Estimates of exposure to monomers (expressed as parts per million in the air, or as the annual rate of exposures –peaks- above a certain threshold) were developed for 308 work area-job groups and for each calendar year between 1943 and 1991.² Each subject's work history was linked with the exposure estimates to obtain the subject's exposure history. The work history data sets consist of an observation for each unique job and for each year a subject worked in that job. Observations in the work history data set begin with the first day worked in the plant and end with the last day worked or 12/31/91, whichever is earlier. Follow-up time in the study begins after completion of one year of work at the plant or January 1, 1943, whichever was later, and ends on the death date or on 12/31/91, whichever was earlier (for subjects whose vital status was known) or on the last day of employment (for a small proportion of subjects whose vital status was unknown). Cumulative exposure indices and other time-dependent covariates (i.e., variables whose value

changes within the same subject's history as time passes) are recursively computed during each year of follow-up, and the information is stored in a person-year file. Cumulative exposure to butadiene, styrene or benzene is expressed in ppm-years (i.e., the average ppm during one year summed over years of exposure) or in total peaks (i.e., the annual frequency of exposure to peaks summed over the years). A decedent file is also created, which contains information on cumulative exposure and other variables as of the time of death.

While the decedent file contains only one observation for each decedent, the person-year file contains one observation for each year of follow-up contributed by each subject. Combination of the output person years data with local, regional or U.S. standard mortality rates leads to the computation of cause-specific expected numbers of deaths for the analysis of standardized mortality ratios (SMRs). The files can be used for stratified or Poisson regression analyses in SAS to study cause-specific mortality rates and ratios as a function of cumulative exposure to butadiene, styrene and benzene (ppm-years or peaks) adjusting for potential confounders. The macro arguments allow the user to select different synthetic rubber plants and different exposure estimates for butadiene, styrene and benzene. The macro could be adapted for the analysis of diverse exposure-disease relationships in epidemiologic studies conducted in other occupational or in nonoccupational settings, in which exposure is accumulated over time.

FUNCTION OF MACRO EXPPYRS

The first DATA step in the macro reads the job exposure matrix data set which includes the following variables used throughout the macro (all date variables are SAS date values):

YDUR= # days worked in a job group during that year
SDATE= start date for job segment
EDATE= end date for job segment
YOD= death date

Macro variables:

&BDEXP= exposure estimate for butadiene
&STEXP= exposure estimate for styrene
&BZEXP= exposure estimate for benzene

```

%MACRO exppyrs(dsn,titl,bdexp,stexp,bzexp);
*****
This DATA step reads the Work History file in the
form of one observation/job/year, sorts by ssn, job
number, calendar year
*****;

data spl;
set in.&dsn._jem(keep=ssn jobnum sdate edate
jobgrp ydur yob yin yout yrout yrin plant icd8
race yod &bdexp &stexp &bzexp);

length calyr 4;
calyr = year(sdate);
run;

proc sort data=spl;
by ssn sdate calyr;
run;

```

DECEDENT FILE

In the data step that creates the decedent file, length of employment in days is calculated for each job in a given calendar year, multiplied by the calendar year-specific exposure indices for butadiene, styrene, and benzene, then summed over all the job by year records until the date of death. The resulting variables are the cumulative exposure indices for each decedent. These variables are named CUMEXPBD, CUMEXPST, and CUMEXPBZ for butadiene, styrene and benzene respectively. Length of employment is also accumulated until death and the variable name is LENEMPL. AGEDEATH is age computed as of date of death. Time since first exposure to butadiene, TIME1BD, and to styrene, TIME1ST, and to benzene, TIME1BZ, are all calculated as of the date of death.

```

*****
Creates decedent file. Outputs only work histories
for decedents.
*****;

data &dsn._dec(keep=ssn race yob plant icd8 yout
yod yin ageddeath lenempl
time1bd time1st time1bz
cumexpbd cumexpst
cumexpbz yrin yrout);

set spl;
by ssn;
where 43 <= mod(year(yod),100) < 92;
length ageddeath 3;

if first.ssn then do;
date1bd=.; date1st=.; date1bz=.;
le=0; cebd=0; cest=0; cebz=0;
end;
retain date1bd date1st date1bz le cebd cest cebz;

*****
Calculates date first exposed to butadiene, styrene
and benzene
*****;

if date1bd = . and &bdexp ne 0 then date1bd = sdate;
if date1st = . and &stexp ne 0 then date1st = sdate;
if date1bz = . and &bzexp ne 0 then date1bz = sdate;

```

```

*****
Cumulates length of employment and cumulative
exposure through date of death
*****;

le = ydur + le;
cebd = cebd + (ydur*&bdexp);
cest = cest + (ydur*&stexp);
cebz = cebz + (ydur*&bzexp);

if last.ssn then do;
agedeath = int((yod - yob)/365.25);
time1bd = (yod - date1bd)/365.25;
time1st = (yod - date1st)/365.25;
time1bz = (yod - date1bz)/365.25;
lenempl = le/365.25;
cumexpbd = cebd/365.25;
cumexpst = cest/365.25;
cumexpbz = cebz/365.25;
output;
end;
run;

data out.&dsn._dec;
set &dsn._dec;
if time1bd = . then time1bd=0;
if time1st = . then time1st=0;
if time1bz = . then time1bz=0;
run;

```

PERSON-YEAR FILE

The next data step creates the cohort person-years of follow-up file. Length of employment and cumulative exposure indices are calculated as in the decedent file through the beginning of follow-up date, YRIN. For the calendar year of YRIN and for every year through the end of follow-up date, YROUT, for that individual an observation is output. For the calendar year of YRIN, length of employment and cumulative exposure are calculated by a similar equation.

LE0 = the length of employment cumulated through the beginning of follow-up

LE1 = LE0 + the number of days worked through the end of the calendar year of YRIN

Note: If YROUT is in the same calendar year as YRIN, then length of employment is only calculated through YROUT

The cumulated length of employment in years output for that calendar year is the midpoint of the first calendar year of follow-up:

$$\text{LENEMPL} = ((\text{LE0} + \text{LE1})/2)/365.25$$

Cumulative exposure for butadiene is calculated as:

CEBD0 = number of days in each job segment X the cumulative exposure index summed over all job segments through the beginning of follow-up

CEBD1 = CE0 + number of days in each job segment X the cumulative exposure index summed over all job segments through the end of the calendar year of YRIN

Note: If YROUT is in the same calendar year as YRIN then cumulative exposure is only calculated through YROUT

Cumulative exposure in ppm- years output for that calendar year is the approximate mid-point cumulative exposure for that year:

$$\text{CUMEXPBD} = ((\text{CEBD0} + \text{CEBD1})/2)/365.25$$

Similarly, variables for styrene are CEST0, CEST1 and CUMEXPST, and for benzene are CEBZ0, CEBZ1, and CUMEXPBZ.

```
data &dsn._pyrs(keep=ssn race yob plant icd8 yout
                yod yin yrin yROUT lenempl
                time1bd time1st time1bz
                cumexpbd cumexpst
                cumexpbz py calyr
                age compress='yes');

set spl;
by ssn calyr;
length m d y 3 age 3;
if first.ssn then do;
    date1bd=.; date1st=.; date1bz=.;
    le=0; cebd0=0; cest0=0; cebz0=0; le0=0; le1=0;
    cebd1=0; cest1=0; cebz1=0;
end;

retain date1bd date1st date1bz le le0 le1 cebd cest
cebz cebd0 cest0 cebz0 cebd1 cest1 cebz1;

*****
Calculates date first exposed to butadiene, styrene
and benzene
*****

if date1bd = . and &bdexp ne 0 then date1bd = sdate;
if date1st = . and &stexp ne 0 then date1st = sdate;
if date1bz = . and &bzexp ne 0 then date1bz = sdate;

*****
Cumulates length of employment and cumulative
exposure through beginning of follow-up (YRIN)
*****
    if edate <= yrin then do;
        le = le + ydur;
        cebd = cebd + (ydur*&bdexp);
        cest = cest + (ydur*&stexp);
        cebz = cebz + (ydur*&bzexp);
    end;

    else if sdate <= yrin and edate > yrin then do;
        le = le + (yrin - sdate);
        cebd = cebd + (yrin - sdate)*&bdexp;
        cest = cest + (yrin - sdate)*&stexp;
        cebz = cebz + (yrin - sdate)*&bzexp;
        le0 = le;
        cebd0 = cebd;
        cest0 = cest;
        cebz0 = cebz;
        le1 = le0 + ((edate + 1) - yrin);
        cebd1 = cebd0 + ((edate + 1) - yrin)*&bdexp;
        cest1 = cest0 + ((edate + 1) - yrin)*&stexp;
        cebz1 = cebz0 + ((edate + 1) - yrin)*&bzexp;
    end;
```

For years other than the year of YRIN, LE0 and CEBD0, CEST0 and CEBZ0 are calculated as the accumulation at the beginning of the calendar year. For the calendar year of interest, LE1, CEBD1, CEST1, and CEBZ1 are accumulated as above except that calculations are through the end of that calendar year. If the calendar year is the year of YROUT, then LE1, CEBD1, CEST1, and CEBZ1 are only accumulated through YROUT.

The midpoint length of employment, LENEMPL, and the approximate midpoint cumulative exposures for butadiene (CUMEXPBD), styrene (CUMEXPST), and benzene (CUMEXPBZ) are calculated as illustrated above for the year of YRIN.

```
*****
Cumulates length of employment and cumulative
exposure for job segments after the beginning date
of follow-up and which do not include YRIN
*****
    else if sdate > yrin then do;

        if first.calyr then do;
            le0 = le1;
            cebd0 = cebd1;
            cest0 = cest1;
            cebz0 = cebz1;
        end;
        le1 = le1 + ((edate + 1) - sdate);
        cebd1 = cebd1 + ((edate + 1) - sdate)*&bdexp;
        cest1 = cest1 + ((edate + 1) - sdate)*&stexp;
        cebz1 = cebz1 + ((edate + 1) - sdate)*&bzexp;
    end;

*****
Calculates the approximate midpoint length of
employment and cumulative exposure for each
calendar year
*****

    if last.calyr then do;

        lenempl = ((le0 + le1)/2)/365.25;
        cumexpbd = ((cebd0 + cebd1)/2)/365.25;
        cumexpst = ((cest0 + cest1)/2)/365.25;
        cumexpbz = ((cebz0 + cebz1)/2)/365.25;
```

The program calculates person years (PY) for the year of YRIN as the end of the calendar year - YRIN. If YROUT is in the same year as YRIN then PY = YROUT - YRIN. For all calendar years between YRIN and YROUT, PY = 1. In the year of YROUT, PY = YROUT - the beginning of the calendar year.

Age and time since first exposure to butadiene, styrene and benzene, for the year of YRIN are calculated as of the midpoint between the end of the calendar year of YRIN and YRIN. For all other calendar years except the year of YROUT, age and time since first exposure are calculated as of July 1 of the calendar year of interest. For the year of YROUT, these variables are calculated as of the midpoint between YROUT and the beginning of the calendar year of YROUT. The variable names are AGE, and TIME1BD, TIME1ST and TIME1BZ for butadiene, styrene and benzene respectively.

```
*****
Calculates for the first calendar year of follow-up: time
since first exposure to butadiene, styrene, and benzene,
and age
```

If the calendar year of YROUT is greater than the calendar year of YRIN, then the midpoint is midway between the end of the calendar year and YRIN. If the year of YRIN equals the year of YROUT, then the midpoint is midway between YROUT and YRIN.

The person years value for the year of YRIN equals the end of the calendar year – YRIN unless YROUT is in the same calendar year as YRIN then person years = YROUT-YRIN.

```
*****;
```

```
if calyr = year(yrin) then do;
```

```
if calyr ne year(yrout) then do;
```

```
m=12; d=31; y=calyr;
```

```
endyr=mdy(m,d,y);
```

```
end;
```

```
else if calyr eq year(yrout) then endyr=yrout-1;
```

```
py = ((endyr + 1) - yrin)/365.25;
```

```
midpoint = ((endyr + 1) - yrin)/2 + yrin;
```

```
age = int((midpoint - yob)/365.25);
```

```
time1bd = (midpoint - date1bd)/365.25;
```

```
time1st = (midpoint - date1st)/365.25;
```

```
time1bz = (midpoint - date1bz)/365.25;
```

```
output;
```

```
end;
```

```
*****;
```

Calculates for the years of follow-up after the first (excluding the calendar year of YROUT): time since first exposure to butadiene, styrene, and benzene, and age, as of July 1 of the calendar year, except for the year of YROUT when the variables are computed as of the midpoint between January 1 and YROUT.

The person years value always = 1 except for the year of YROUT when person years equals YROUT - Jan 1 of the same calendar year.

```
*****;
```

```
else if calyr > year(yrin) and calyr ne year(yrout)
then do;
```

```
m=7; d=1; y=calyr;
```

```
midpoint = mdy(m,d,y);
```

```
py = 1;
```

```
age = int((midpoint - yob)/365.25);
```

```
time1bd = (midpoint - date1bd)/365.25;
```

```
time1st = (midpoint - date1st)/365.25;
```

```
time1bz = (midpoint - date1bz)/365.25;
```

```
output;
```

```
end;
```

```
else if calyr > year(yrin) and calyr eq year(yrout)
then do;
```

```
m=1; d=1; y=calyr;
```

```
begyr = mdy(m,d,y);
```

```
endyr = yrout;
```

```
py = (endyr - begyr)/365.25;
```

```
midpoint = (endyr - begyr)/2 + begyr;
```

```
age = int((midpoint - yob)/365.25);
```

```
time1bd = (midpoint - date1bd)/365.25;
```

```
time1st = (midpoint - date1st)/365.25;
```

```
time1bz = (midpoint - date1bz)/365.25;
```

```
output;
```

```
end;
```

```
end; * if last.calyr;
```

After the last date worked, an observation is output for each calendar year through the end of follow-up. Length of employment and cumulative exposure values remain the same as of the last day worked for each consecutive year through the end of follow-up. Other variables output to the data set are SSN, race, and calendar year.

```
*****
```

For calendar years after the last date worked until the end of follow-up, cumulative length of employment, and exposure to butadiene, styrene and benzene, retain their same values as of the last date worked.

```
*****;
```

```
if last.ssn and (calyr+1) <= year(yrout) then do;
```

```
do i=(calyr+1) to (year(yrout)-1);
```

```
m=7; d=1; y=i;
```

```
midpoint = mdy(m,d,y);
```

```
py = 1;
```

```
age = int((midpoint - yob)/365.25);
```

```
time1bd = (midpoint - date1bd)/365.25;
```

```
time1st = (midpoint - date1st)/365.25;
```

```
time1bz = (midpoint - date1bz)/365.25;
```

```
calyr = i;
```

```
lenempl = le1/365.25;
```

```
cumexpbd = cebd1/365.25;
```

```
cumexpst = cest1/365.25;
```

```
cumexpbz = cebz1/365.25;
```

```
output;
```

```
end;
```

```
m=1; d=1; y=year(yrout);
```

```
begyr = mdy(m,d,y);
```

```
endyr = yrout;
```

```
if endyr > begyr then do;
```

```
py = (endyr - begyr)/365.25;
```

```
midpoint = (endyr - begyr)/2 + begyr;
```

```
age = int((midpoint - yob)/365.25);
```

```
time1bd = (midpoint - date1bd)/365.25;
```

```
time1st = (midpoint - date1st)/365.25;
```

```
time1bz = (midpoint - date1bz)/365.25;
```

```
calyr = i ;
```

```
lenempl = le1/365.25;
```

```
cumexpbd = cebd1/365.25;
```

```
cumexpst = cest1/365.25;
```

```
cumexpbz = cebz1/365.25;
```

```
output;
```

```
end;
```

```
end; * if last.ssn;
```

```
run;
```

```
data out.&dsn._pyrs(compress='yes');
```

```
set &dsn._pyrs;
```

```
if time1bd = . then time1bd = 0;
```

```
if time1st = . then time1st = 0;
```

```
if time1bz = . then time1bz = 0;
```

```
run;
```

```
%mend;
```

Last, the macro is invoked for different rubber plants and different exposure values for butadiene, styrene and benzene.

```
%exppyrs(cop,Copolymer,bdtwalo,stytwalo,bztwmid);
```

```
%exppyrs(fir,Firestone,bdtwalo,stytwalo,bztwmid);
```

```
%exppyrs(gdy,Goodyear,bdtwalo,stytwalo,bztwmid);
```

```
%exppyrs(as,Ameripol Synpol,bdtwalo,stytwalo,bztwmid);
```

CONCLUSION

Dynamic allocation of follow-up time to categories of time-dependent variables is an important programming aspect of the analysis of follow-up studies. Epidemiologists and biostatisticians have used a variety of packages to accomplish this, while many use SAS both for the preliminary manipulations of the data and for the final statistical analysis (e.g., PROC GENMOD for fitting Poisson regression models). During the past few years, we have developed an array of programs that attempt to fill the gap,^{3,4} and obviate the need for leaving the SAS work environment for this intermediate data processing step. The program described in this paper addresses the important issue of evaluating the effect of environmental exposures whose dose accumulates at variable rates over an extended follow-up time.

REFERENCES

1. Delzell E, Sathiakumar N, Hovinga M, Macaluso M, Julian J, Larson R, Cole P, Muir DCF(1996), A follow-up study of synthetic rubber workers. *Toxicology*, 113, 182-189.
2. Macaluso M, Larson R, Delzell E, Sathiakumar N, Hovinga M, Muir D, Julian J, Cole P(1996), Leukemia and cumulative exposure to butadiene, styrene and benzene among workers in the synthetic rubber industry. *Toxicology*, 113,190-202.
3. Macaluso M(1992), Exact stratification of person-years. *Epidemiology*, 3,441-448.
4. Honda Y, Macaluso M, Brill I. A SAS program for the stratified analysis of follow-up data (1998), *J Occup Health*, 40, 154-157.

SAS and SAS/STAT software are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

AUTHOR CONTACT INFORMATION

Ilene Brill
Department of Epidemiology
University of Alabama at Birmingham
Mortimer Jordan Hall, Room 108
1825 University Blvd.
Birmingham, AL 35294-2010
Work Phone: (205) 975-5359
Fax: (205) 975-2435
Email: ibrill@mrg.soph.uab.edu