WEB APPENDIX 1

Technical Note

The expected value of Y_i^a for a person who experienced the hazard rate, h^a, over the span of T years, $\sum_{u=0}^{T_i} h^a(u|\boldsymbol{W}_i)\, S^a(u-1|\boldsymbol{W}_i)L(u)\ \ldots$, Web Expression 1, can be readily estimated in the setting of a constant hazard and constant span T by the expression $E(Y_i^a) = 1 - \exp(-h^a T)$. To see this, let $F(t)$ denote cumulative distribution function, which is the quantity we want to estimate (i.e., the cumulative probability of the event up to time t). Let $S(t) = 1 - F(t)$, and assume constant hazard, such that $h(t) = h$. Denote the derivative of $F(t)$ with respect to time as $f(t) = dF/d\mathrm{t} = -dS(t)/dt = h(t)S(t)$, as dt goes to 0. This is the quantity that we integrate over in Web Expression 1 via the discrete time summation to time T, yielding an estimate of $F(t)$. A constant hazard, h, implies an exponential distribution for the time until an event occurs, $F(t) = 1 - \exp(-ht)$. Hence, in this setting we can bypass the summation over discrete time intervals to obtain cumulative probability, $F(t)$.

WEB APPENDIX 2

Calculation of Expected Events

While SMRs are often calculated using data that have been summarized in a grouped data structure defined by cross-classification of factors such as attained age, calendar period, race, and sex, the same result can be obtained using the individual person-time data. The expected number of deaths is obtained by multiplying the appropriate age-, calendar-period-, race-, and sex-specific reference rate by the amount of person-time contributed by each person in each period of observation. These products are summed over the all person-periods that were contributed by all individuals to obtain the expected number of deaths.

For the calculations described here, we suppose a person-period file is constructed with one record for each person-period from study entry until end of follow-up. Setting survival prior to the first period at 1, for each person-period, we calculate the product of the reference hazard rate function for that period and estimated reference survival function for the prior period. Following each person-period, we update the reference survival function. The reference hazard rate, $h^0(t|\boldsymbol{W})$, represents race- and sex-specific national death rates for 5-year age and calendar period intervals. The attained age and calendar period is determined by the time on study and the person's age at entry and calendar date of entry into the study. A SAS program is provided (see below) to calculate the expected number of deaths in a single pass through the data by summation of the contribution made by each person-period over the total possible person-time.

A 95% confidence interval may be derived for the CMR using an approximation of the Poisson distribution of the number of observed events (1, 2) analogous to its use when calculating a confidence interval for the classical SMR. Exact methods also may be used to derive a confidence interval for the ratio of observed to expected events and are the preferred method if the observed number of deaths is small (e.g., fewer than 6) (3).

SAS code for all-cause mortality calculations

Assume the study data consist of a person-period file, named *DS*. Each person is identified by a unique study ID, *i*. There is one record per unit of person-time, with time indexed by the variable *u* which takes a value of 0 at study entry and increases monotonically until time T_i, the administrative end of follow-up. Each record of this data set includes the following:

i, a unique study ID for each person;
u, a variable that indexes potential follow-up time, from 0 at study entry to *T* at end of follow-up;
age, a variable equal to age-at-entry plus *u*,
period, a variable equal to date-of-entry plus *u*;
sex, an indicator of person's sex;
race, an indicator of person's race (if reference rates depend upon race);
c, a binary variable that equals 1 for time prior to date of last observation ($u \leq T'$), else 0
rate, reference death rate, $h^0(u|\boldsymbol{W_i})$, expressed in deaths per person-year at risk (e.g., reference rate for the race and sex group of subject *i*, and the 5-year age and calendar period intervals associated with follow-up time *u*);
lu, unit of person-time (e.g., 1 if each record represents one person-year).

The data are sorted by *i* and *u*.

The following code calculates *E*; as in the text this denotes the expected number of deaths if the cohort had experienced the reference hazard rate. The code also calculates the expected number of deaths using the classic SMR method; in the text this quantity is denoted $\tilde{E}$, while here we denote this value by *Q*.

```
data MR ;
set DS end=eof ;
```

```
 by i u ;
retain S_i expected_i expected_smr_i  E Q;
if _n_=1 then do;   E=0; Q=0; end;
if first.i then do;   S_i=1; expected_i=0; expected_smr_i=0; end;
expected_i=expected_i+ (rate * S_i * lu); expected_smr_i= expected_smr_i + (rate * c * lu);
S_i=1-expected_i ;
if last.i then do ;  E=E+expected_i;  Q=Q+expected_smr_i;  end;
if eof then output MR; run;

proc print data=MR ;  var E Q; run;
```

SAS code for cause-specific mortality

Suppose that the data set, DS, also includes the variables *rateA* and *rateB*, reference death rates for causes A and B, $h_A^0(u|\boldsymbol{W_i})$ and $h_B^0(u|\boldsymbol{W_i})$, expressed in deaths per person-year at risk (e.g., reference rates for causes A and B for the race and sex group of subject *i*, and the 5-year age and calendar period intervals associated with follow-up time *u*). The data are sorted by *i* and *u*. The following code calculates *E* and *Q* for a specific cause of death, A, allowing for mortality due to competing causes, B.

```
data MR ;
set DS END=EOF ;
by i u ;
 retain S_i expectedA_i expectedB_i expected_smr_i   E Q ;
if _n_=1 then do;  E=0; Q=0; end;
if first.i then do;  S_i=1; expectedA_i=0; expectedB_i=0; expected_smr_i=0; end;
expectedA_i=expectedA_i+ (rateA * S_i * lu); expectedB_i=expectedB_i+ (rateB * S_i * lu);
expected_smr_i= expected_smr_i  + (rateA * c * lu);
S_i=1-expectedA_i-expectedB_i ;
if last.i then do ;   E=E+expectedA_i;   Q=Q+expected_smr_i; end;
if eof then output MR; run;

proc print data=MR ; var E Q; run;
```

Web References

1. Ury HK, Wiggins AD. Another shortcut method for calculating the confidence interval of a Poisson variable (or of a standardized mortality ratio). *Am J Epidemiol* 1985;122(1):197–198.
2. Vandenbroucke JP. A shortcut method for calculating the 95 percent confidence interval of the standardized mortality ratio. *Am J Epidemiol* 1982;115:303–304.
3. Rothman K, Boice JD. Epidemiologic analyses with a programmable calculator. Washington, DC: US Department of Health, Education, and Welfare, 1979.