Statistical Analysis of the Standardized Mortality Ratio and Life Expectancy

Dejian Lai, Robert J. Hardy, and Shan P. Tsai

A new theoretical relation that does not require the constant age-specific mortality ratio assumption is established between the standardized mortality ratio (SMR) and the life expectancy. A set of regression equations is developed from the theoretical relation to derive estimates of the future expectation of life from estimates of the SMR. Curves are presented showing the changes in life expectancy that are associated with a given SMR for individuals aged 25, 45, and 65 years. These results will provide practical applications in estimating remaining life expectancy in epidemiologic studies in which the SMR is the summary statistic. An application is shown for studies in occupational health to develop and illustrate the method. Am J Epidemiol 1996;143:832-41.

epidemiologic methods; life expectancy; mortality; regression analysis; statistics; vital statistics

The crude mortality rate of a given population depends on the age structure and the age-specific death rates of the population. Comparing crude mortality rates from populations with differing age structure does not reveal the true difference of the mortality experience of the populations. One commonly used summary statistic in this kind of comparison is the standardized mortality ratio (SMR). The SMR measures the excess or deficit of the mortality in the selected study population compared with a chosen standard population (1) and has been widely used in studying the patterns of disease affecting workers by occupational health professionals (2).

The SMR is calculated from the age-specific death rates of a standard population and from the age structure of the study population and total number of deaths in the study population. If there are sufficient data to estimate the age-specific mortality rates in the study population, then another commonly used summary statistic, the expectation of life (or life expectancy), can be calculated for the study population. Life expectancy at each specific age is an estimate of the average future life span for a person who survives to a given age if the age-specific mortality rates do not change in the future. The life expectancy is calculated by using life table methods (3), and its use is limited in occupation studies by the effort needed to adequately follow working cohorts to provide stable estimates of the age-specific rates.

A theoretical relation between life expectancy and the SMR has been developed under the assumption of a constant age-specific mortality ratio for all age groups providing a technique to estimate the life expectancy from SMR (1). In this paper, a new theoretical relation is derived that relaxes the assumption of a constant mortality ratio and provides a more general statistical procedure for estimating the life expectancy from an estimate of the SMR. Based on the new relation, a set of regression equations is established for men and women aged 25, 45, and 65 years from the US population. These equations provide an easy conversion of SMR of a study population to the corresponding life expectancy. Previous studies (1) were limited to age 25. Curves are presented showing the changes in life expectancy that are associated with a given SMR. The results of this paper extend the findings presented previously in both methodology and application. A comparison of the new and old model is provided.

THEORETICAL RELATION

In general, the SMR is the ratio of observed deaths and expected deaths, adjusting or taking into account the structure of a population with respect to factors such as age or age and sex. In this paper, the SMR is calculated when only age structure is considered, and separate regression equations are provided for each sex. The SMR for a study population is defined as

$$\text{SMR} = \frac{\sum \mu^*(t)P^*(t)}{\sum \mu(t)P^*(t)}$$

(1)
where \( \mu^*(t) \) and \( P^*(t) \) are the age-specific mortality rate and person-years at age \( t \) of the observed (or study) population, respectively, and \( \mu(t) \) is the age-specific mortality rate at age \( t \) of the standard population.

It can be shown that if \( \mu^*(t)/\mu(t) \) is constant for all age groups (homogeneity of age-specific mortality ratio), then \( \text{SMR} = \mu^*(t)/\mu(t) \). Based on this assumption, a relation between life expectancy and SMR was derived previously (1). In this paper, we derive the relation (in instances) when \( \mu^*(t)/\mu(t) \) may not be a constant, i.e., if \( \mu^*(t)/\mu(t) \) is not exactly SMR and it fluctuates around SMR. Hence, we assume that

\[
\mu^*(t)/\mu(t) = \text{SMR} + \delta(t),
\]

where \( \Sigma w(t)\delta(t) = 0 \) and \( \delta(t) \) is the residual fluctuation of age-specific mortality ratio around SMR, and \( w(t) = \mu(t)P^*(t)/\Sigma \mu(t)P^*(t) \) are weights proportional to the expected number of deaths.

To develop the relation between the expectation of life and the SMR, we consider the formal definition of the expectation of life as a function of the force of mortality \( \mu(t) \). Following the result in Chiang (3), the life expectancy at age \( t_x \) for the study population is

\[
e_x^* = \int_{t_x}^{\infty} e^{-\int_{t_x}^{t} \mu^*(r) dr} dt,
\]

and the life expectancy at age \( t_x \) for the standard population is

\[
e_x = \int_{t_x}^{\infty} e^{-\int_{t_x}^{t} \mu(r) dr} dt.
\]

The relation between the expectation of life and the SMR of the study population can be derived under equation 2 as follows:

\[
e_x^* = \int_{t_x}^{\infty} e^{-\int_{t_x}^{t} \mu^*(r) dr} dt = \int_{t_x}^{\infty} e^{-\int_{t_x}^{t} \mu^*(r)/\mu(t) \mu(r) dr} dt
\]

\[
= \int_{t_x}^{\infty} e^{-\int_{t_x}^{t} (\text{SMR} + \delta(t)) \mu(r) dr} dt = \int_{t_x}^{\infty} e^{-\text{SMR} \int_{t_x}^{t} \mu(r) dr - \int_{t_x}^{t} \delta(t) \mu(r) dr} dt
\]

\[
= \int_{t_x}^{\infty} e^{-\text{SMR} \int_{t_x}^{t} \mu(r) dr} dt.
\]

Hence,

\[
e_x^* = \int_{t_x}^{\infty} e^{-\int_{t_x}^{t} \mu^*(r) dr} dt - \int_{t_x}^{\infty} \delta(t) \mu(t) dr dt.
\]

From equation 5 and the mean value theorem, there exists a \( t' \in (t_x, \infty) \), such that

\[
e_x^* = e^{-\text{SMR} \int_{t_x}^{t'} \mu(r) dr} dt + \int_{t_x}^{t'} \delta(t) \mu(t) dr dt.
\]

Equation 6 implies that

\[
\ln \left( \frac{e_x^*}{e_x} \right) = -(\text{SMR} - 1) \int_{t_x}^{t'} \mu^*(r) dr - \int_{t_x}^{t'} \delta(t) \mu(t) dr. \tag{7}
\]

In equation 7, we let \( y = \ln(e_x^*/e_x) \) be the response variable and \( x = 1 - \text{SMR} \) be the independent variable in a simple linear regression equation. The value \( -\int_{t_x}^{t'} \delta(t) \mu(t) dr \) is the intercept and \( \int_{t_x}^{t'} \mu^*(r) dr \) is the slope. The intercept and slope can be estimated for different ages. Using the data described in Materials and Methods, separate equations are estimated at ages 25, 45, and 65 years.

Equation 7 is an extension of earlier results that assumed \( \delta(t) = 0 \) for all \( t \) and led to a regression similar to equation 7 with zero intercept. The intercept term in equation 7 provides more flexibility in fitting the regression line without complicating the model and provides a preferable method for the relation between life expectancy and SMR.

The Appendix extends the methodology to analyze the contribution of each cause of death to life expectancy.

**MATERIALS AND METHODS**

In occupational health studies in which the objective is to estimate the SMR, one may be interested in knowing how the SMR for a study population translates into a gain or loss in expectation of life. Separate equations are formed for the workforce at 25 years, the middle workforce at 45 years, and the retirement workforce at 65 years.

Using the white resident population of the United States including each of the 50 states and the District of Columbia by age and sex from the US Bureau of the Census Current Population Report for April 1, 1980 (4), and the deaths in 1980 from the US National Center for Health Statistics (5), we calculated the all-cause SMR for each of the 50 states and the District of Columbia for both men and women in age groups.
25, 45, and 65 years. The US age-specific mortality rates serve as the standard. The corresponding life expectancy is based on the published life table for each state (6) and the entire United States (7). Following equation 7, the $y$ variable is the logarithm of the ratio of the life expectancy of each state and the life expectancy of the United States. The $x$ variable is $1 - \text{SMR}$ of the corresponding state. These data are used to estimate the linear regression equation expressed as $y = \alpha + \beta x$, where $\alpha$ is the intercept and $\beta$ is the slope.

Due to the final form derived from equation 7, the regression relation of the life expectancy of the study population and the corresponding SMR is

$$e^*_x = e^{\alpha + \beta x} \times e_x.$$  \hspace{1cm} (8)

**RESULTS**

By using SAS Proc GLM, the parameters are estimated in the model for the selected three ages (25, 45, 65 years) and for both sexes. The results are summarized in table 1.

In table 1, $e_x$ is the life expectancy of the corresponding group for the total US white population (7). The comparison of the old model (zero intercept) and the new model (nonzero intercept) are illustrated in figures 1–6 for both sexes and for ages 25, 45, and 65, which are of interest to occupational health professionals. The solid line indicates the line with nonzero intercept (new model).

The two models are quite similar; however, the new model is more flexible and can capture the relation between life expectancy and SMR better than the old model.

Because the intercepts are positive and the slopes of the regression equations from the new model are flat-

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Life expectancy ($e_x$ years)</th>
<th>Parameter*</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>47.8</td>
<td>0.0011</td>
</tr>
<tr>
<td>45</td>
<td>29.4</td>
<td>0.0040</td>
</tr>
<tr>
<td>65</td>
<td>14.2</td>
<td>0.0045</td>
</tr>
<tr>
<td>25</td>
<td>54.5</td>
<td>0.0018</td>
</tr>
<tr>
<td>45</td>
<td>35.4</td>
<td>0.0024</td>
</tr>
<tr>
<td>65</td>
<td>18.5</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

* Parameters $\alpha$ and $\beta$ are defined in equation 8 of the text.

**FIGURE 1.** Standardized mortality ratio and life expectancy in 1980 for US white men at 25 years of age. Solid line, new model (nonzero intercept); dotted line, old model (zero intercept).
ter than those of the previous model, the estimated life expectancy is slightly greater than that of the old model in the observed range of SMR.

The statistical estimation based on equation 7 establishes the regression equations that allows one to translate the SMR into a life expectancy for each of the age-gender groups. The life expectancy increases gradually as the SMR decreases (figures 7 and 8).

The relation of changes in life expectancy for ages 25, 45, and 65 are presented in figure 9 for men and figure 10 for women.

For example, a 1 percent decrease of SMR increases the life expectancy for men who are 25, 45, and 65 years old by 0.1325 years (1.6 months), 0.1165 years (1.4 months), and 0.0756 years (0.9 months), respectively. A decrease in the SMR of 1 percent for 25-, 45-, and 65-year-old women will increase life expectancy by 0.1016 years (1.2 months), 0.0938 years (1.1 months), and 0.0692 years (0.8 months), respectively. The changes of life expectancy due to changes of SMR are greater in younger age groups than those of older groups, and the contribution of changes of SMR to the changes of life expectancy in the male population are greater than those of the female population. For changes not studied in this paper, interpolation techniques can be used.

**DISCUSSION**

A new relation between the summary statistics, life expectancy, and SMR has been developed that establishes a new equation for age 25 and extends the results to ages 45 and 65. Although SMR is widely used in the occupational mortality studies, it is not well understood by lay people. However, life expectancy is a commonly used summary statistic, and it is more easily understood as the average future years of life. The regression equations derived from the new theoretical relation between these two statistics can help nonspecialists understand and interpret the results of studies in which the SMR is the primary summary measure.

The inverse relation of SMR and life expectancy can be captured from a set of regression equations. The regression equations allow one to assess the impact of changes of SMR on the life expectancy of the study population. The impact is presented in visual plots of curves that provide clear illustration of the relation.

The age-specific mortality ratio in occupational cohort studies usually increases with increasing age due to the latent period associated with exposure. Because the age-specific ratio is assumed to be constant in the previous model for all ages, the estimated life expect-
FIGURE 3. Standardized mortality ratio and life expectancy in 1980 for US white men at 65 years of age. ———, new model (nonzero intercept); ———, old model (zero intercept).

FIGURE 4. Standardized mortality ratio and life expectancy in 1980 for US white women at 25 years of age. ———, new model (nonzero intercept); ———, old model (zero intercept).
FIGURE 5. Standardized mortality ratio and life expectancy in 1980 for US white women at 45 years of age. ——, new model (nonzero intercept); ———, old model (zero intercept).

FIGURE 6. Standardized mortality ratio and life expectancy in 1980 for US white women at 65 years of age. ———, new model (nonzero intercept); ———, old model (zero intercept).
FIGURE 7. Standardized mortality ratio and life expectancy in 1980 for US white men at 25, 45, and 65 years of age.

FIGURE 8. Standardized mortality ratio and life expectancy in 1980 for US white women at 25, 45, and 65 years of age.

FIGURE 10. Life expectancy in relation to changes in standardized mortality ratio in 1980 for US white women at 25, 45, and 65 years of age.
ancies tend to underestimate the corresponding values calculated on the basis of life table techniques (1). In the new model, an intercept is introduced in the new regression equations. It provides better estimation of life expectancies.

It should be noted that the theoretical equation 7 is an approximation. Our empirical analyses of the relation of life expectancy and SMR are based on the US white population in 1980. The model proposed assumed that stable age-specific mortality rates derive estimates of the life expectancy. Hence, the coefficients should be updated periodically. Calculations using these equations will be approximations for other countries; however, the relation of life expectancy and SMR established in this paper can be used to generate similar regression equations for study populations from different countries.

REFERENCES


APPENDIX

In this appendix, a relation similar to equation 7 is developed for life expectancy and cause-specific SMR. The total mortality intensity for the study population can be partitioned into \( r \) component cause-specific mortality intensities or

\[
\mu^*(t) = \mu^*(t; 1) + \cdots + \mu^*(t; r).
\]

The total mortality intensity for a standard population can be partitioned in a similar way,

\[
\mu(t) = \mu(t; 1) + \cdots + \mu(t; r),
\]

where \( \mu^*(t; j) \) and \( \mu(t; j) \) for \( j = 1, 2, \ldots, r \) are the cause-specific mortality intensity functions for the study and comparison population at age \( t \). The ratio of the two forces of mortality can be written as

\[
\frac{\mu^*(t)}{\mu(t)} = \frac{[\mu^*(t; 1) + \cdots + \mu^*(t; r)]}{\mu(t)} = \sum \theta_j(t) \frac{\mu(t; j)}{\mu(t)} = \sum \theta_j(t) \pi_j(t),
\]

where \( \theta_j(t) = \mu^*(t; j)/\mu(t; j) \) is the cause-specific mortality intensity ratio and \( \pi_j(t) = \mu(t; j)/\mu(t) \) is the proportional mortality for cause \( j \) of the comparison population.

Under the homogeneity of cause-specific mortality intensity ratio and constant proportional mortality assumption, \( \theta_j(t) \) is the cause-specific SMR (denoted as \( SMR_j \)) and a relation of life expectancy, and \( SMR_j \) is derived in (1). We can relax homogeneity of cause-specific mortality assumption as

\[
\theta_j(t) = SMR_j + \epsilon_j(t),
\]

and the constant proportional mortality assumption as

\[
\pi_j(t) = \pi_j + \delta_j(t),
\]

where \( \pi_j \) is constant, \( \sum \pi_j = 1 \), and \( \sum \delta_j(t) = 0 \). Then, by the same method in the theoretical relation section, one has

\[
e^* = e^{-\int_0^t \theta_j(\tau) \mu(\tau) d\tau - \sum \int_0^t \epsilon_j(\tau) \pi_j(\tau) d\tau} \times e^{-\int_0^t \pi_j - SMR_j \int_0^t \mu(\tau) d\tau} \times e^x.
\]

If we let

\[
y = \ln(e^*/e_x),
\]

\[
x_j = (1 - SMR_j), \quad \text{or}\ (1 - SMR) \pi_j,
\]

\[
\beta_j = \int_0^t \pi_j \mu(\tau) d\tau, \quad \text{or} \int_0^t \mu(\tau) d\tau,
\]

and

\[
\alpha = -\sum \int_0^t \epsilon_j(\tau) \delta_j(\tau) \mu(\tau) \ d\tau - \sum \int_0^t \epsilon_j(\tau) \pi_j \mu(\tau) \ d\tau.
\]

Then, one has a relation of life expectancy and the cause-specific SMR as

\[
y = \alpha + \sum \beta_j x_j.
\]