NATIONAL CENTER For HEALTH STATISTICS | Series 2 Number 20

VITAL and HEALTH STATISTICS

DATA EVALUATION AND METHODS RESEARCH

Variance and Covariance of Life Table Functions Estimated From a Sample of Deaths

Formulas for the variance and covariance of functions of abridged and complete life tables based on a sample of deaths.

Washington, D.C. March 1967

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
John W. Gardner
Secretary

Public Health Service
William H. Stewart
Surgeon General

Public Health Service Publication No. 1000-Series 2-No. 20

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C., 20402 - Price 15 cents

Public Health Service Publication No. 1000-Series 2-No. 20

Library of Congress Catalog Card Number 66-62207

FOREWORD

Annually the National Center for Health Statistics prepares two sets of abridged life tables for the United States. In the regular abridged life tables the age-specific mortality rates are based on tabulations of all deaths occurring during the calendar year distributed by age, color, and sex. In the sample life tables the age-specific mortality rates are based on a 10-percent sample of deaths; these rates are available several months prior to the rates based on a complete count of deaths.

Formulas for the variance of the functions of life tables based on a complete count of deaths were derived several years ago. These formulas have appeared in several publications prepared by Dr. C. L. Chiang. However, formulas for the variance of the life table functions were not available for assessing the reliability of life tables based on a sample of deaths. Accordingly the National Center for Health Statistics invited Dr. Chiang to generalize his earlier work and make the results applicable to life tables based on a sample of deaths. This report presents his results and includes the formulas for variance and covariance of the life tables functions based on a sample of deaths.

As the Center's project director for this study, I identified the problem and worked with Dr. Chiang in the early stages of developing the stochastic model and in reviewing this manuscript for publication.

Monroe G. Sirken, Ph.D., Chief
Division of Health Records
Statistics

CONTENTS

IN THIS REPORT formulas for the variance and covariance of functions of abridged and complete life tables based on a sample of deaths are derived. It is well known that life table functions, as any statistical quantities, are random variables. When a life table is constructed on the basis of a sample of deaths rather than on the total count, there is a component of sampling *variation associated with the observed values. This component of variation must be assessed in making statistical inference regarding survival experience of a population as determined in such a table. In the formulas for the variance and covariance of estimates of the probability of dying* (q_i), *the survival rate* (p_{0i}), *and the expectation of life* (e_i) *presented in this paper, both the random variation and the sampling variation are taken into account.*

VARIANCE AND COVARIANCE OF LIFE TABLE FUNCTIONS ESTIMATED FROM A SAMPLE OF DEATHS

C. L. Chiang, Ph.D., *School of Public Health, University of California, Berkeley*

I. INTRODUCTION

The Federal Government has published annual abridged life tables for the United States since 1945.[1] These tables are based on age-specific mortality rates computed from the total number of deaths occurring during the calendar year and from the estimate of the midyear population. In 1958[2] the Federal vital statistics agency established another series of annual abridged life tables based on a 10-percent sample of deaths with the objective of publishing annual life tables on a more current schedule.

Using a 10-percent sample rather than the total number of deaths makes very little difference in the numerical values of a life table, but it does increase the amount of variation associated with these values. This is because the life table functions as determined by a sample are subject to sampling variation and to the random variation present when the total count is used. The main functions of general interest are the probability of dying (q_i), the survival rate (p_{0i}), and the expectation of life (e_i). The purpose of this paper is to derive formulas for the variances of the estimates of these functions, taking into account both random variation and sampling variation. In section II we shall reproduce the corresponding formulas for these functions for the case where only random variation is considered.

II. A LIFE TABLE BASED ON THE COMPLETE COUNT OF DEATHS

In earlier studies of life table and mortality rates[3-5] probability distributions of life table functions and formulas for the corresponding variances were derived. Some of these formulas are reproduced below for easy reference; the original publications can be consulted for details.

In an abridged life table the life span is divided into $w+1$ intervals, each of which is defined by two exact ages (x_i, x_{i+1}), for $i = 0, 1, ..., w$, except for the last interval which is usually a half-open interval such as "95 and over." The age x_0 may be taken as 0, the age at birth, and x_w as the age at the beginning of the last interval. For the interval (x_i, x_{i+1}) let $n_i = x_{i+1} - x_i$ be the length of the interval; when $n_i = 1$ for all i, we have the complete life table. Let D_i be the number dying in the age interval (x_i, x_{i+1}) during the calendar year and P_i be the corresponding midyear population, so that the sum

$$D_0 + D_1 + \ldots + D_w = D \tag{1}$$

is the total number of deaths during the year and

$$P_0 + P_1 + \ldots + P_w = P \tag{2}$$

is the total midyear population. The ratio

$$\frac{D_i}{P_i} = M_i \tag{3}$$

is the age-specific mortality rate.

Let N_i be the number of individuals alive at exact age x_i among whom the D_i deaths occur so that

$$\hat{q}_i = \frac{D_i}{N_i} \tag{4}$$

is an estimate of the probability that an individual alive at exact age x_i will die in the interval (x_i, x_{i+1}). The number N_i is not directly observed but is estimated from D_i and P_i. A meaningful concept of the relation between N_i and P_i from the standpoint of the life table is that P_i is an estimate of the total number of years lived in the interval (x_i, x_{i+1}) by N_i individuals. Let a_i be the average fraction of the interval (x_i, x_{i+1}) lived by each of the D_i individuals.[6] It can be seen that

$$P_i = n_i(N_i - D_i) + a_i n_i D_i \ . \tag{5}$$

The first term on the right side of (5) is the number of years lived by the $N_i - D_i$ survivors, and the second term is equal to the sum of the fractions of the interval lived by the D_i individuals who die during the interval. Equation (5) can be rewritten as

$$N_i = \frac{1}{n_i}\left[P_i + (1-a_i)n_i D_i\right]. \tag{6}$$

By substituting (6) in (4) we establish a basic relationship between the age-specific death rate (M_i) and the corresponding estimated probability of dying $(\hat{q}_i)$

$$\hat{q}_i = \frac{n_i M_i}{1 + (1-a_i)n_i M_i} \ . \tag{7}$$

This formula has been used for the construction of life tables[7] and has appeared in references 4 and 5.

To derive the formula for the variance of $\hat{q}_i$, we only need to observe that D_i is a binomial random variable in N_i "trials" with the binomial probability q_i. Accordingly, D_i has the expectation

$$E(D_i) = N_i q_i \tag{8}$$

and the variance

$$\sigma_{D_i}^2 = N_i q_i (1-q_i) . \tag{9}$$

Therefore, the sample variance of $\hat{q}_i$ is

$$S_{\hat{q}_i}^2 = \frac{1}{N_i}\hat{q}_i(1-\hat{q}_i) = \frac{1}{D_i}\hat{q}_i^2(1-\hat{q}_i). \tag{10}$$

Substituting (7) in (10) yields the required formula

$$S_{\hat{q}_i}^2 = \frac{n_i M_i (1 - a_i n_i M_i)}{P_i\left[1 + (1-a_i)n_i M_i\right]^3} \ . \tag{11}$$

The covariance between $\hat{q}_i$ and $\hat{q}_j$ for two age intervals is equal to zero as shown in reference 3. Formula (11) is very important in the statistical analysis of a life table since the formulas for the variances of other life table functions can all be expressed in terms of the variance of $\hat{q}_i$. Two important examples follow.

1. In a current life table, the proportion of survivors from age 0 to age x_α

$$\hat{p}_{0\alpha} = \frac{\ell_\alpha}{\ell_0} \tag{12}$$

is actually computed from

$$\begin{aligned}\hat{p}_{0\alpha} &= \hat{p}_0\hat{p}_1 \cdots \hat{p}_{\alpha-1} \\ &= (1-\hat{q}_0)(1-\hat{q}_1)\dots(1-\hat{q}_{\alpha-1}) ,\end{aligned} \tag{13}$$

where $\hat{p}_i = 1 - \hat{q}_i$ is an estimate of the probability of surviving the interval (x_i, x_{i+1}). The formula for the sample variance of $\hat{p}_{0\alpha}$ is

$$S_{\hat{p}_{0\alpha}}^2 = \hat{p}_{0\alpha}^2 \sum_{i=0}^{\alpha-1} \hat{p}_i^{-2} S_{\hat{q}_i}^2 \ . \tag{14}$$

2. The observed expectation of life at age x_α may be expressed as

$$\begin{aligned}\hat{e}_\alpha = a_\alpha n_\alpha &+ c_{\alpha+1}\hat{p}_{\alpha,\alpha+1} \\ &+ c_{\alpha+2}\hat{p}_{\alpha,\alpha+2} + \dots + c_w \hat{p}_{\alpha,w} ,\end{aligned} \tag{15}$$

where

$$c_i = (1-a_{i-1})\,n_{i-1} + a_i n_i . \qquad (16)$$

The sample variance of $\hat{e}_\alpha$ is given by

$$S^2_{\hat{e}_\alpha} = \sum_{i=\alpha}^{w-1} \hat{p}^2_{\alpha i}\,[\hat{e}_{i+1} + (1-a_i)\,n_i]^2\, S^2_{\hat{q}_i} . \qquad (17)$$

The derivation of formulas (14) and (17) is given in references 3 and 4 and will not be repeated here.

III. A LIFE TABLE BASED ON A SAMPLE OF DEATHS

In the construction of the new series of life tables, the actual number (D_i) of deaths is not observed directly but is estimated by a sampling procedure. The estimated value of D_i is then used as the basis for the computation of the mortality rate and the life table functions. We have then a sample of size d taken without replacement from the total of D death certificates such that

$$\frac{d}{D} = f \qquad (18)$$

is a preassigned sampling fraction. In the new series described here, $f=.10$. Depending upon the distribution by age at death, a number d_i of deaths in the sample will fall into the age interval (x_i, x_{i+1}) with

$$d_0 + d_1 + \ldots + d_w = d . \qquad (19)$$

The number D_i is estimated from

$$\hat{D}_i = \frac{d_i}{f} \qquad (20)$$

and the age-specific mortality rate from

$$M_i = \frac{\hat{D}_i}{P_i} = \frac{d_i}{fP_i} . \qquad (21)$$

It is clear that d_i (or $\hat{D}_i$) and hence M_i are subject to both random variation and sampling variation. As a result, the formula for the variance of $\hat{q}_i$ as presented in section II must be revised and the covariance between $\hat{q}_i$ and $\hat{q}_j$ will need to be evaluated. Our first step is to derive the formulas for the variance of d_i and the covariance between d_i and d_j.

The variance of d_i.—The probability distribution of d_i depends upon the total number of deaths D_i in the age interval (x_i, x_{i+1}) and the total number of deaths D at all ages. Relative to D_i and D, the variance of d_i may be written as

$$\sigma^2_{d_i} = \sigma^2_{E(d_i|D_i,D)} + E\sigma^2_{d_i|D_i,D} . \qquad (22)$$

The first term on the right side of (22) is the variance of the conditional expectation of d_i given D_i and D, and the second term is the expectation of the variance of the conditional distribution of d_i given D_i and D. We shall discuss them separately.

Since the sample is taken without replacement, the conditional distribution of d_i given D_i and D is hypergeometric with the expectation

$$E(d_i|D_i, D) = d\,\frac{D_i}{D} = fD_i . \qquad (23)$$

Using formulas (23) and (9), we compute the variance of $E(d_i|D_i,D)$,

$$\sigma^2_{E(d_i|D_i,D)} = f^2\sigma^2_{D_i} = f^2 N_i q_i (1-q_i) . \qquad (24)$$

For the last term in equation (22), we make use of the well-known theorem in the hypergeometric distribution to write

$$\begin{aligned}\sigma^2_{d_i|D_i,D} &= d\frac{D_i}{D}\left(1-\frac{D_i}{D}\right)\left(\frac{D-d}{D-1}\right)\\ &= d\frac{D_i}{D}\left(1-\frac{D_i}{D}\right)\left(\frac{D-d}{D}\right)\end{aligned} \qquad (25)$$

since the number D is usually large. We now rewrite (25)

$$\begin{aligned}\sigma^2_{d_i|D_i,D} &= \frac{d}{D}\left(1-\frac{d}{D}\right)\left(D_i - \frac{D_i^2}{D}\right)\\ &= f(1-f)\left(D_i - \frac{D_i^2}{D}\right)\end{aligned} \qquad (26)$$

and its expectation

$$E\sigma^2_{d_i|D_i,D} = f(1-f)\left[E(D_i) - E\left(\frac{D_i^2}{D}\right)\right] . \qquad (27)$$

The first expectation inside the brackets was given in (8), and the second expectation may be rewritten as

$$E\left(\frac{1}{D}D_i^2\right) = E\left[\frac{1}{D}E(D_i^2|D)\right] . \qquad (28)$$

Our problem is to find the conditional expectation $E(D_i^2/D)$.

To avoid confusion in notation, let us consider the particular case in which $i=0$ and write the conditional expectation

$$E(D_0^2|D) = \sum_{\delta=0}^{D} \delta_0^2 Pr\{D_0 = \delta_0 | D\} \tag{29}$$

where δ_0 is the value that the random variable D_0 takes on and $Pr\{D_0 = \delta_0|D\}$ is the corresponding conditional probability. Using Bayes' law,

$$Pr\{B|A\} = \frac{Pr\{B\}\cdot Pr\{A|B\}}{Pr\{A\}} \tag{30}$$

we may rewrite the conditional probability as

$$Pr\{D_0 = \delta_0|D\} = \frac{Pr\{D_0 = \delta_0\}\cdot Pr\{D_1 + \cdots + D_w = D - \delta_0\}}{Pr\{D_0 + D_1 + \cdots + D_w = D\}}$$

$$= \frac{Pr\{D_0 = \delta_0\}\cdot \sum_{\delta_1}\cdots\sum_{\delta_w}\prod_{k=1}^{w} Pr\{D_k = \delta_k\}}{\sum_{\delta_0}\sum_{\delta_1}\cdots\sum_{\delta_w}\prod_{k=0}^{w} Pr\{D_k = \delta_k\}}, \tag{31}$$

where the sum in the numerator is taken over all possible values of δ so that

$$\delta_1 + \cdots + \delta_w = D - \delta_0 \tag{32}$$

and in the denominator

$$\delta_0 + \delta_1 + \cdots + \delta_w = D. \tag{33}$$

The probability distribution of each D_i is binomial with the probability distribution

$$Pr\{D_i = \delta_i\} = \frac{N_i!}{\delta_i!(N_i - \delta_i)!} q_i^{\delta_i}(1-q_i)^{N_i - \delta_i}. \tag{34}$$

When (34) is substituted, formula (31) becomes very unwieldy. Ordinarily, however, the probability q_i is small and N_i is large, so that the Poisson distribution should be a good approximation; this will give us

$$Pr\{D_i = \delta_i\} = \frac{e^{-\lambda_i}\lambda_i^{\delta_i}}{\delta_i!} \tag{35}$$

where, for simplicity,

$$\lambda_i = N_i q_i = E(D_i) \tag{36}$$

is the expected number of deaths in the age interval (x_i, x_{i+1}) as given in (8). Now formula (35) is substituted in the last expression of (31) to give the sum in the numerator

$$\sum_{\delta_1}\cdots\sum_{\delta_w}\prod_{k=1}^{w} Pr\{D_k = \delta_k\} = \sum_{\delta_1}\cdots\sum_{\delta_w}\frac{e^{-\lambda_1}\lambda_1^{\delta_1}}{\delta_1!}\cdots\frac{e^{-\lambda_w}\lambda_w^{\delta_w}}{\delta_w!}$$

$$= e^{-(\lambda_1+\cdots+\lambda_w)}\frac{1}{(D-\delta_0)!}\sum_{\delta_1}\cdots\sum_{\delta_w}\frac{(D-\delta_0)!}{\delta_1!\ldots\delta_w!}\lambda_1^{\delta_1}\cdots\lambda_w^{\delta_w}$$

$$= e^{-(\lambda_1+\cdots+\lambda_w)}\frac{1}{(D-\delta_0)!}(\lambda_1+\cdots+\lambda_w)^{D-\delta_0} \tag{37}$$

the denominator

$$\sum_{\delta_0}\sum_{\delta_1}\cdots\sum_{\delta_w}\prod_{k=0}^{w} Pr\{D_k = \delta_k\} \tag{38}$$

$$= e^{-(\lambda_0+\lambda_1+\cdots+\lambda_w)}\frac{1}{D!}(\lambda_0+\lambda_1+\cdots+\lambda_w)^D,$$

and, finally, after simplification, the conditional probability

$$Pr\{D_0 = \delta_0|D\} = \frac{D!}{\delta_0!(D-\delta_0)!}\left[\frac{\lambda_0}{\lambda_0+\lambda_1+\cdots+\lambda_w}\right]^{\delta_0}\left[1-\frac{\lambda_0}{\lambda_0+\lambda_1+\cdots+\lambda_w}\right]^{D-\delta_0}. \tag{39}$$

Formula (39) shows that the conditional distribution of D_0 given D is binomial with the proportion of the expected number of deaths in the age interval (x_0, x_1)

$$\frac{\lambda_0}{\sum_{k=0}^{w}\lambda_k} = \frac{N_0 q_0}{\sum_{k=0}^{w} N_k q_k} \tag{40}$$

as the binomial probability. It follows that

$$E(D_0^2|D) = D\frac{N_0 q_0}{\sum_{k=0}^{w} N_k q_k}\left(1 - \frac{N_0 q_0}{\sum_{k=0}^{w} N_k q_k}\right) + D^2\frac{(N_0 q_0)^2}{\left(\sum_{k=0}^{w} N_k q_k\right)^2}. \tag{41}$$

In general we have

$$E(D_i^2|D)=D\frac{N_i q_i}{\sum_{k=0}^{w} N_k q_k}\left(1-\frac{N_i q_i}{\sum_{k=0}^{w} N_k q_k}\right) + D^2\frac{(N_i q_i)^2}{\left(\sum_{k=0}^{w} N_k q_k\right)^2}. \qquad (42)$$

Using (42) and the relation

$$E(D)=E\left(\sum_{k=0}^{w} D_k\right)=\sum_{k=0}^{w} N_k q_k \qquad (43)$$

we compute the expectation in (28)

$$E\left\{\frac{1}{D}E(D_i^2|D)\right\}=\frac{N_i q_i}{\Sigma N_k q_k}\left(1-\frac{N_i q_i}{\Sigma N_k q_k}\right) + E(D)\frac{(N_i q_i)^2}{(\Sigma N_k q_k)^2} \qquad (44)$$

$$=\frac{N_i q_i}{\Sigma N_k q_k}\left(1-\frac{N_i q_i}{\Sigma N_k q_k}\right)+\frac{(N_i q_i)^2}{\Sigma N_k q_k},$$

where the summation in each denominator is taken from $k=0$ to $k=w$. The expectation of the variance in (27) thus becomes

$$E\sigma^2_{d_i|D_i,D}=f(1-f)\left\{N_i q_i-\frac{N_i q_i}{\Sigma N_k q_k}\left(1-\frac{N_i q_i}{\Sigma N_k q_k}\right)-\frac{(N_i q_i)^2}{\Sigma N_k q_k}\right\} \qquad (45)$$

$$=f(1-f)N_i q_i\left(1-\frac{1}{\Sigma N_k q_k}\right)\left(1-\frac{N_i q_i}{\Sigma N_k q_k}\right).$$

Substituting (24) and (45) in (22) gives the required formula for the variance of d_i,

$$\sigma^2_{d_i}=f^2 N_i q_i(1-q_i) + f(1-f)N_i q_i\left(1-\frac{1}{\Sigma N_k q_k}\right)\left(1-\frac{N_i q_i}{\Sigma N_k q_k}\right). \qquad (46)$$

It may be noted that since

$$D_0+D_1+\ldots+D_w=D \qquad (47)$$

and

$$\frac{N_0 q_0}{\Sigma N_k q_k}+\frac{N_1 q_1}{\Sigma N_k q_k}+\ldots+\frac{N_w q_w}{\Sigma N_k q_k}=1, \qquad (48)$$

the conditional joint distribution of $D_0,\ldots,D_w$ given D is multinomial with the expectation of the product $D_i\ D_j$

$$E(D_i D_j|D)=D(D-1)\left(\frac{N_i q_i}{\Sigma N_k q_k}\right)\left(\frac{N_j q_j}{\Sigma N_k q_k}\right). \qquad (49)$$

The covariance between d_i *and* d_j.—Again we write the covariance between d_i and d_j relative to D_i, D_j, and D as follows:

$$\sigma_{d_i d_j}=\sigma_{E(d_i|D_i,D),E(d_j|D_j,D)}+E\sigma_{d_i,d_j|D_i,D_j,D}. \qquad (50)$$

We have the conditional expectation as given in (23),

$$E(d_i|D_i,D)=fD_i, \qquad (23)$$

and hence

$$\sigma_{E(d_i|D_i,D),E(d_j|D_j,D)}=\sigma_{fD_i,fD_j}=f^2\sigma_{D_i,D_j}. \qquad (51)$$

Since the numbers of deaths at different ages are independently distributed, $\sigma_{D_i D_j}=0$, the covariance of the conditional expectations vanishes,

$$\sigma_{E(d_i|D_i,D_j,D),E(d_j|D_i,D_j,D)}=0 \qquad (52)$$

and the covariance between d_i and d_j is reduced to

$$\sigma_{d_i d_j}=E\sigma_{d_i,d_j|D_i,D_j,D}. \qquad (53)$$

The joint distribution of $d_0,d_1,\ldots,d_w$ given $D_0,D_1,\ldots,D_w$ and D is the generalized hypergeometric, and the covariance between d_i and d_j is

$$\sigma_{d_i,d_j|D_i,D_j,D}=-d\frac{D_i}{D}\frac{D_j}{D}\left(1-\frac{d}{D}\right) = -f(1-f)\frac{D_i D_j}{D}. \qquad (54)$$

Therefore (53) becomes

$$\sigma_{d_i,d_j}=-f(1-f)E\left(\frac{D_i D_j}{D}\right) = -f(1-f)E\left[\frac{1}{D}E(D_i D_j|D)\right]. \qquad (55)$$

Introducing (49) in (55) and using (43) we have the formula for the covariance between d_i and d_j ,

$$\sigma_{d_i,d_j} = -f(1-f)\left[1-\frac{1}{\Sigma N_k q_k}\right]\frac{(N_i q_i)(N_j q_j)}{\Sigma N_k q_k}. \tag{56}$$

The variance and covariance of $\hat{q}_i$.—In this case the estimated probability $\hat{q}_i$ is computed from

$$\hat{q}_i = \frac{\hat{D}_i}{N_i} = \frac{d_i}{fN_i} \tag{57}$$

where N_i as before is defined as the number of individuals alive at x_i among whom D_i deaths occur in the age interval (x_i, x_{i+1}). Since both f and N_i are constant, the variance of $\hat{q}_i$ can be obtained from

$$\sigma^2_{\hat{q}_i} = \frac{1}{f^2N_i^2}\sigma^2_{d_i}. \tag{58}$$

Substituting (46) in (58) we have the variance

$$\sigma^2_{\hat{q}_i} = \frac{1}{N_i} q_i(1-q_i) + \left(\frac{1}{f}-1\right)\frac{1}{N_i} q_i \times \left(1-\frac{1}{\Sigma N_k q_k}\right)\left(1-\frac{N_i q_i}{\Sigma N_k q_k}\right). \tag{59}$$

Formula (59) shows that the variance of $\hat{q}_i$ consists of two components. The first term on the right side of (59) is the component of random variation, and the second term is the component of sampling variation. The second component decreases as the sampling fraction f increases. When the total count of deaths is used so that $f = 1$, the second term vanishes and (59) is reduced to a form of which formula (10) is an estimate.

The covariance of $\hat{q}_i$ and $\hat{q}_j$ can also be derived by using the covariance of d_i and d_j. From (57) we can write

$$\sigma_{\hat{q}_i,\hat{q}_j} = \frac{1}{f^2N_iN_j}\sigma_{d_i d_j}. \tag{60}$$

In view of (56),

$$\sigma_{\hat{q}_i,\hat{q}_j} = -\left(\frac{1}{f}-1\right)\left(1-\frac{1}{\Sigma N_k q_k}\right)\frac{q_i q_j}{\Sigma N_k q_k}. \tag{61}$$

Thus the covariance between $\hat{q}_i$ and $\hat{q}_j$ is entirely due to sampling variation and vanishes when $f=1$.

The variance of $\hat{p}_{0\alpha}$ *and* $\hat{e}_\alpha$.—The formulas for the variance of the proportion of survivors and the observed expectation of life are derived by using an approximate method (see p. 232 of reference 8). For the proportion of survivors

$$\hat{p}_0 = \hat{p}_0\hat{p}_1 \cdots \hat{p}_{\alpha-1}, \tag{13}$$

we take the derivatives

$$\frac{\partial}{\partial \hat{p}_i}\hat{p}_{0\alpha}\Big|_{\hat{p}=p} = \frac{p_{0\alpha}}{p_i} \tag{62}$$

and write

$$\sigma_{\hat{p}_{0\alpha}} = \sum_{i=0}^{\alpha-1}\left\{\frac{\partial}{\partial\hat{p}_i}\hat{p}_{0\alpha}\right\}^2\sigma^2_{\hat{q}_i} + \sum_{\substack{i=0\\ i\neq j}}^{\alpha-1}\sum_{j=0}^{\alpha-1}\left\{\frac{\partial}{\partial\hat{p}_i}\hat{p}_{0\alpha}\right\} \times \left\{\frac{\partial}{\partial\hat{p}_j}\hat{p}_{0\alpha}\right\}\sigma_{\hat{q}_i\hat{q}_j} \tag{63}$$

since $\sigma^2_{\hat{p}_i} = \sigma^2_{\hat{q}_i}$ and $\sigma_{\hat{p}_i,\hat{p}_j} = \sigma_{\hat{q}_i,\hat{q}_j}$. Substituting (62) in (63) gives

$$\sigma^2_{\hat{p}_{0\alpha}} = p^2_{0\alpha}\left[\sum_{i=0}^{\alpha-1}\frac{1}{p_i^2}\sigma^2_{\hat{q}_i} + \sum_{\substack{i=0\\ i\neq j}}^{\alpha-1}\sum_{j=0}^{\alpha-1}\frac{1}{p_ip_j}\sigma_{\hat{q}_i,\hat{q}_j}\right], \tag{64}$$

where the variance and covariance of $\hat{q}_i$ and $\hat{q}_j$ are given in (59) and (61), respectively.

To derive the formula for the variance of $\hat{e}_\alpha$, the observed expectation of life at age x_α, we recall expression (15)

$$\hat{e}_\alpha = a_\alpha n_\alpha + c_{\alpha+1}\hat{p}_{\alpha,\alpha+1} + c_{\alpha+2}\hat{p}_{\alpha,\alpha+2} + \cdots + c_w\hat{p}_{\alpha,w} \tag{15}$$

and write

$$\sigma^2_{\hat{e}_\alpha} = \sum_{i=\alpha}^{w-1} \left\{\frac{\partial}{\partial \hat{p}_i} \hat{e}_\alpha\right\}^2 \sigma^2_{\hat{q}_i} \tag{65}$$

$$+ \sum_{i=\alpha}^{w-1} \sum_{\substack{j=\alpha \\ i \neq j}}^{w-1} \left\{\frac{\partial}{\partial \hat{p}_i} \hat{e}_\alpha\right\} \left\{\frac{\partial}{\partial \hat{p}_j} \hat{e}_\alpha\right\} \sigma_{\hat{q}_i, \hat{q}_j}.$$

It is easy to see that for each k,

$$\frac{\partial}{\partial \hat{p}_i} \hat{p}_{\alpha k} = p_{\alpha i}\, p_{i+1,k} \qquad \text{for } \alpha \le i < k$$

$$\frac{\partial}{\partial \hat{p}_i} \hat{p}_{\alpha k} = 0, \qquad \text{for } i \ge k \tag{66}$$

and hence from (15) we have

$$\frac{\partial}{\partial \hat{p}_i} \hat{e}_\alpha = \sum_{k=i+1}^{w} c_k\, p_{\alpha i}\, p_{i+1,k}$$

$$= p_{\alpha i}\left[c_{i+1} + \sum_{k=i+2}^{w} c_k\, p_{i+1,k}\right]. \tag{67}$$

Since $c_{i+1} = (1-a_i)n_i + a_{i+1}\, n_{i+1}$,

as defined in (16), (67) may be rewritten as

$$\frac{\partial}{\partial \hat{p}_i} \hat{e}_\alpha = p_{\alpha i}\left[(1-a_i)\,n_i + a_{i+1}\, n_{i+1} + \sum_{k=i+2} c_k p_{i+1,k}\right]$$

$$= p_{\alpha i}\left[(1-a_i)\, n_i + c_{i+1}\right]. \tag{68}$$

Now we substitute (68) in (65) to obtain the required formula for the variance of $\hat{e}_\alpha$,

$$\sigma^2_{\hat{e}_\alpha} = \sum_{i=\alpha}^{w-1} p^2_{\alpha i}\left[(1-a_i)\, n_i + e_{i+1}\right]^2 \sigma^2_{\hat{q}_i}$$

$$+ \sum_{i=\alpha}^{w-1} \sum_{\substack{j=\alpha \\ i \neq j}}^{w-1} p_{\alpha i}\, p_{\alpha j}\left[(1-a_i)\, n_i + e_{i+1}\right]$$

$$\left[(1-a_j)\, n_j + e_{j+1}\right]\sigma_{\hat{q}_i, \hat{q}_j}$$

$$\alpha = 0, 1, \ldots, \tag{69}$$

where the variance and covariance of $\hat{q}_i$ and $\hat{q}_j$ are given by (59) and (61), respectively.

Sample variances of the life table functions.— Estimates of the variances of the life table functions derived in the preceding sections can be made by substituting observed values for the corresponding unknown quantities in the respective formulas. Thus we use

$$\hat{D}_i = \frac{d_i}{f} \tag{20}$$

as an estimate of the expected number of deaths $N_i\, q_i$ and compute the age-specific death rate from

$$M_i = \frac{\hat{D}_i}{P_i}, \tag{21}$$

N_i from (57)

$$N_i = \frac{\hat{D}_i}{\hat{q}_i},$$

and the estimated probability of dying from

$$\hat{q}_i = \frac{n_i\, M_i}{1 + (1-a_i)\, n_i M_i}. \tag{7}$$

When these substitutions are made in (59) and (61), we have the sample variance of $\hat{q}_i$

$$S^2_{\hat{q}_i} = \frac{1}{\hat{D}_i}\hat{q}_i^2\,(1-\hat{q}_i) + \left(\frac{1}{f}-1\right)\frac{1}{\hat{D}_i}\hat{q}_i^2\left(1-\frac{1}{D}\right)\left(1-\frac{\hat{D}_i}{D}\right) \tag{70}$$

and the covariance between $\hat{q}_i$ and $\hat{q}_j$

$$S_{\hat{q}_i, \hat{q}_j} = -\left(\frac{1}{f}-1\right)\left(1-\frac{1}{D}\right)\frac{\hat{q}_i \hat{q}_j}{D}. \tag{71}$$

It is interesting to note that although it was necessary to introduce the quantity N_i in the derivation of the formulas for the variance and covariance, the final expressions in (70) and (71) hold whatever may be the relation between $\hat{q}_i$ and M_i. The variances of $\hat{p}_{0\alpha}$ and $\hat{e}_\alpha$ can now be estimated from

$$S^2_{\hat{p}_{0\alpha}} = \hat{p}^2_{0\alpha}\left[\sum_{i=0}^{\alpha-1}\frac{1}{\hat{p}_i^2} S^2_{\hat{q}_i} + \sum_{i=0}^{\alpha-1}\sum_{\substack{j=0 \\ i\neq j}}^{\alpha-1}\frac{1}{\hat{p}_i \hat{p}_j} S_{\hat{q}_i, \hat{q}_j}\right] \tag{72}$$

and

$$S^2_{\hat{e}_\alpha} = \sum_{i=\alpha}^{w-1} \hat{p}^2_{\alpha i} \left[(1-\alpha_i)n_i + \hat{e}_{i+1}\right]^2 S^2_{\hat{q}_i} + \sum_{i=\alpha}^{w-1} \sum_{\substack{j=\alpha \\ i \neq j}}^{w-1} \hat{p}_{\alpha i} \hat{p}_{\alpha j} \left[(1-\alpha_i)n_i + \hat{e}_{i+1}\right] \left[(1-a_j)n_j + \hat{e}_{j+1}\right] S_{\hat{q}_i, \hat{q}_j}. \tag{73}$$

Since every quantity in equations (70) through (73) can be determined from observed data, the sample variances of the life table functions can all be computed.

REFERENCES

[1]National Center for Health Statistics: *Guide to United States Life Tables, 1900-1959,* by M. G. Sirkin, L. L. Bachrach, and G. A. Carlson. PHS Pub. No. 1086, Public Health Bibliographic Series No. 42. Public Health Service. Washington. U.S. Government Printing Office, 1963.

[2]National Center for Health Statistics: Comparison of two methods of constructing abridged life tables by reference to a "standard" table. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 4. Public Health Service. Washington. U.S. Government Printing Office, Mar. 1966.

[3]Chiang, C. L.: A stochastic study of the life table and its applications, I, probability distributions of the biometric functions. *Biometrics* 16(4):618-635, Dec. 1960.

[4]Chiang, C. L.: A stochastic study of the life table and its applications, II, sample variance of the observed expectation of life and other biometric functions. *Human Biology* 32(3):221-238, Sept. 1960.

[5]National Vital Statistics Division: Standard error of the age-adjusted death rate, by C. L. Chiang. *Vital Statistics–Special Reports,* Vol. 47, No. 9. Public Health Service. Washington. U.S. Government Printing Office, Aug. 1961.

[6]Chiang, C. L., Norris, F. D., Olson, F. E., Shipley, P. W., Supplee, H. E.: Determination of the Fraction of Last Year of Life and Its Variation by Age, Race, and Sex (unpublished manuscript).

[7]Norris, F. D., and Riese, D. C.: *Abridged Life Tables, California, 1959-61.* California State Department of Public Health, Bureau of Vital Statistics and Data Processing, Sept. 1963.

[8]Kendall, M. G., and Stuart, A.: *The Advanced Theory of Statistics,* Vol. I. London. Charles Griffin and Co., Ltd., 1963.

— o o o —

I wish to express my appreciation to Miss Helen E. Supplee for a critical reading of the paper and valuable suggestions.

U. S. GOVERNMENT PRINTING OFFICE : 1967 O - 246-085

OUTLINE OF REPORT SERIES FOR VITAL AND HEALTH STATISTICS

Public Health Service Publication No. 1000

Series 1. *Programs and collection procedures.*—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions, data collection methods used, definitions, and other material necessary for understanding the data.

Series 2. *Data evaluation and methods research.*—Studies of new statistical methodology including: experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, contributions to statistical theory.

Series 3. *Analytical studies.*—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

Series 4. *Documents and committee reports.*—Final reports of major committees concerned with vital and health statistics, and documents such as recommended model vital registration laws and revised birth and death certificates.

Series 10. *Data from the Health Interview Survey.*—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, based on data collected in a continuing national household interview survey.

Series 11. *Data from the Health Examination Survey.*—Data from direct examination, testing, and measurement of national samples of the population provide the basis for two types of reports: (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics; and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

Series 12. *Data from the Institutional Population Surveys.*—Statistics relating to the health characteristics of persons in institutions, and on medical, nursing, and personal care received, based on national samples of establishments providing these services and samples of the residents or patients.

Series 13. *Data from the Hospital Discharge Survey.*—Statistics relating to discharged patients in short-stay hospitals, based on a sample of patient records in a national sample of hospitals.

Series 20. *Data on mortality.*—Various statistics on mortality other than as included in annual or monthly reports—special analyses by cause of death, age, and other demographic variables, also geographic and time series analyses.

Series 21. *Data on natality, marriage, and divorce.*—Various statistics on natality, marriage, and divorce other than as included in annual or monthly reports—special analyses by demographic variables, also geographic and time series analyses, studies of fertility.

Series 22. *Data from the National Natality and Mortality Surveys.*—Statistics on characteristics of births and deaths not available from the vital records, based on sample surveys stemming from these records, including such topics as mortality by socioeconomic class, medical experience in the last year of life, characteristics of pregnancy, etc.

For a list of titles of reports published in these series, write to:

Office of Information and Publications
National Center for Health Statistics
U.S. Public Health Service
Washington, D.C. 20201