



A Stochastic Study of the Life Table and Its Applications. III. The Follow-Up Study with the Consideration of Competing Risks

Author(s): Chin Long Chiang

Source: *Biometrics*, Vol. 17, No. 1 (Mar., 1961), pp. 57-78

Published by: International Biometric Society

Stable URL: <http://www.jstor.org/stable/2527496>

Accessed: 05/05/2009 18:08

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=ibs>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.



International Biometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Biometrics*.

<http://www.jstor.org>

A STOCHASTIC STUDY OF THE LIFE TABLE AND ITS APPLICATIONS

III. THE FOLLOW-UP STUDY WITH THE CONSIDERATION OF COMPETING RISKS^{1,2}

CHIN LONG CHIANG

University of California, Berkeley, California, U.S.A.

INTRODUCTION

Statistical studies falling into the general category of life testing and medical follow-up have as their common immediate objective the estimation of life expectation and survival rates for a defined population at risk. Usually such a study must be brought to a close before all the information on survival (of patients, electric bulbs, automobiles, etc.) is complete, and thus the study is said to be truncated. Whether the investigation is basically concerned with life testing or with medical follow-up, the nature of the problem is the same, although differences in sample size may call for different approaches. Thus methods developed for life testing may be applied to follow-up studies when the underlying conditions are met, and vice versa. In this study cancer survival data utilizing a large sample will be used as illustrative material, and we shall accordingly use the terminology of the medical follow-up study as a matter of convenience.

We are concerned then with a typical follow-up study in which a group of individuals with some common morbidity experience are followed from a well-defined zero point, such as date of hospital admission. Perhaps we wish to evaluate a certain therapeutic measure by comparing the expectation of life and survival rates of treated and untreated patients. Or we may wish to compare the expectation of life of treated and presumably cured patients with that of normal persons. When the period of observation is ended, there will usually remain a number of individuals on whom the mortality data in a typical study will be incomplete. Of first importance among these are the

¹This study was completed while the author was a Special Research Fellow of the National Heart Institute, Public Health Service, U. S. Department of Health, Education and Welfare.

²Parts of this paper were presented at the joint meeting of the American Statistical Association, the Institute of Mathematical Statistics, and the Biometric Society in Washington, D. C., December 29, 1959.

persons still alive at the close of the study. Secondly, if the investigation is concerned with mortality from a specific cause, the necessary information is incomplete and unavailable for patients who died from other causes. In addition, there will usually be a third group of patients who were "lost" to the study because of follow-up failure. These three groups present a number of statistical problems in the estimation of the expectation of life and survival rates. Many significant studies have been made along these lines, among them the early works of Greenwood [13] and Karn [17], the actuarial method of Berkson and Gage [1], a stochastic model of competing risks by Fix and Neyman [12], the parametric studies by Berkson and Gage [2] and Boag [3], the non-parametric approach of Kaplan and Meier [16], studies on life testing by Epstein and Sobel [10] and other interesting works by Dorn [8], Elveback [9], Fix [11], Harris, Meier and Tukey [14], and Littell [18].

The purpose of this paper is to adapt the biometric functions of the life table to the special conditions of the follow-up study. Part I considers the general type of study in which survival experience is investigated without specification as to the cause of death. An exact formula will be presented for the maximum likelihood estimator of the probability of death and its asymptotic variance. Special attention will be given to a method for computing the observed expectation of life in truncated studies and the corresponding variance. In Part II the discussion will be extended to apply to studies of mortality from a specific cause in the presence of competing risks. The relations between net, crude, and partial crude probabilities will be reviewed and formulas developed for their estimators and the corresponding variances and covariances. In the last part of the paper, data obtained from the California State Department of Public Health will be used to illustrate the application of the theoretical matter presented in Parts I and II.

Throughout this paper we shall assume that all individuals in a sample are subject to the same force of mortality (or, instantaneous death probability), and that the probability of dying for one individual is not influenced by the death of any other individual in the group. This is to say that the life-times of all individuals in a group are treated as independent and identically distributed random variables. We shall also assume in Part I that there will be no individuals lost to observation because of follow-up failure. The problem of lost cases will be considered in Part II (Remark 3).

For simplicity of presentation, a constant time interval (year) will be used. However, the methods developed in this paper apply equally well to cases where intervals are of different lengths; although the ob-

served expectation of life will have a slightly different form (Cf. [5] and [6]).

The probability symbols used in this paper are listed below for convenient reference. Considering death without specification to cause:

$$p_x = \text{Pr [an individual alive at time } x \text{ will survive the interval } (x, x + 1)],$$

$$q_x = \text{Pr [an individual alive at time } x \text{ will die in the interval } (x, x + 1)],$$

and obviously $p_x + q_x = 1$. When death is studied by cause, or risk, we have the net probabilities:

$$q_{xk} = \text{Pr [an individual alive at time } x \text{ will die in the interval } (x, x + 1) \text{ if risk } R_k \text{ is the only acting risk of death in the population],}$$

$$q_{x.k} = \text{Pr [an individual alive at time } x \text{ will die in the interval } (x, x + 1) \text{ if risk } R_k \text{ is eliminated from the population];}$$

the crude probability:

$$Q_{xk} = \text{Pr [an individual alive at time } x \text{ will die from cause } R_k \text{ in the interval } (x, x + 1), \text{ in the presence of all other risks in the population];}$$

and the partial crude probabilities:

$$Q_{xk.1} = \text{Pr [an individual alive at time } x \text{ will die from cause } R_k \text{ in the interval } (x, x + 1), \text{ when only risk } R_1 \text{ is eliminated from the population],}$$

$$Q_{xk.12} = \text{Pr [an individual alive at time } x \text{ will die from cause } R_k \text{ in the interval } (x, x + 1), \text{ when risks } R_1 \text{ and } R_2 \text{ are eliminated from the population].}$$

PART I. THE ESTIMATION OF THE PROBABILITY OF SURVIVAL AND EXPECTATION OF LIFE

1.1. *The basic random variables and their joint probability function.*

Consider a follow-up study conducted over a period of y years. A total of N_0 individuals are accepted into the study at any time prior to the closing date³ and are observed until death or until the study is terminated, whichever comes first. If we set the time of entrance into

³Although we have used the common closing date method to illustrate the techniques developed in this paper, it should be pointed out that these techniques are equally applicable to the date of last reporting method.

the study as the common point of origin for all N_0 individuals, then N_0 is taken to be the number with which the study began, or the number of individuals alive at time zero. Let x be the exact number of years since entrance into the study, and N_x the number of individuals who survive to the common point x . Clearly, N_x may also be defined as the number of survivors who entered the study at least x years before its closing date. The number of survivors will decrease as x increases, not only because of deaths but also because of withdrawals due to the closing of the study. We will describe this process of depletion systematically for the typical interval $(x, x + 1)$ with reference to Table 1.⁴

At time x , the N_x survivors who begin the interval can be divided into two mutually exclusive groups according to their date of entrance into the study. A group of m patients entered the study more than $x + 1$ years before the closing date of the study. Out of these, δ patients will die in the interval and s will survive to begin the next interval. The second group of n patients entered the study less than $x + 1$ years before its termination, and hence are all counted as withdrawals

TABLE 1
DISTRIBUTION OF N_x PATIENTS ACCORDING TO WITHDRAWAL STATUS, SURVIVAL STATUS, AND CAUSE OF DEATH IN THE INTERVAL $(x, x + 1)^*$

Survival status and cause of death	Withdrawal status in the interval		
	Total number of patients	Number not due for withdrawal**	Number due for withdrawal***
Total Survivors	N_x $s + w$	m s	n w
Deaths, all causes	D	δ	ϵ
Cause of death			
R_1	D_1	δ_1	ϵ_1
R_2	D_2	δ_2	ϵ_2
.	.	.	.
.	.	.	.
R_c	D_c	δ_c	ϵ_c

*The subscript x , which should be added to each of the symbols in the table, is deleted to simplify formulas in the text.

**Patients admitted to the study more than $(x + 1)$ years before closing date.

***Patients admitted to the study less than $(x + 1)$ years but more than x years before closing date.

⁴The second part of Table 1, death by cause, was included for use in Part II of this paper.

in the interval $(x, x + 1)$, whether or not they survive, because for them the closing date precedes their $(x + 1)$ -anniversary date. Let us say that ϵ will die before the closing date and w will survive to withdraw alive.

Thus s , δ , w , and ϵ are the basic random variables (upper part of Table 1) whose distribution depends on the force of mortality. The values that these random variables take on will be used to estimate the probability p_x that a patient will survive the interval $(x, x + 1)$, and its complement q_x , the probability of death in the interval. The first step is to derive the joint probability function of these random variables.

Let μ_τ , a function of time τ , be the force of mortality acting on each individual in the study, such that

$$\mu_\tau \Delta\tau + o(\Delta\tau) = \Pr [\text{an individual alive at time } \tau \text{ will die in the interval } (\tau, \tau + \Delta\tau)], \text{ for } \tau \geq 0,$$

where $\Delta\tau$ stands for an infinitesimal time interval and $o(\Delta\tau)$ a quantity of smaller order of magnitude than $\Delta\tau$. It can be shown [5] that

$$p_x(t) = \exp \left\{ - \int_x^{x+t} \mu_\tau d\tau \right\}$$

is the probability that an individual alive at x will survive to $(x + t)$. If we assume a constant force of mortality within the interval $(x, x + 1)$, say $\mu_\tau = \mu_x$ depending only on x , for $x < \tau \leq x + 1$, then the probability of surviving the interval is given by⁵

$$p_x = e^{-\mu_x}.$$

For the subinterval $(x, x + t)$, with t between zero and one, we have

$$p_x(t) = e^{-t\mu_x} = p_x^t, \text{ for } 0 < t \leq 1.$$

Consider first the group of m individuals, each of whom has a constant probability p_x of surviving and a probability $q_x = 1 - p_x$ of dying in the interval $(x, x + 1)$. We have then a typical binomial case with the probability function:

$$f_1 = p_x^s (1 - p_x)^{\delta}. \quad (1)$$

The expected number of survivors and deaths are given, respectively, by

$$E(s | m) = mp_x, \text{ and } E(\delta | m) = m(1 - p_x). \quad (2)$$

The distribution of the random variables in the group due for withdrawal is not so straightforward. Making the assumption that, on

⁵When the assumption of a constant force of mortality is strong for an interval where the death rate is high, one may subdivide the interval and estimate the probability for each subinterval separately.

the average, each of the m individuals will withdraw at the point $x + \frac{1}{2}$, the probability of withdrawing alive is equal to $p_x^{1/2}$, and the probability of dying before the time of withdrawal $(1 - p_x^{1/2})$. Again we have a binomial case with the probability function⁶

$$f_2 = p_x^{w/2}(1 - p_x^{1/2})^\epsilon. \quad (3)$$

The expected number of survivors and deaths are given, respectively, by

$$E(w | n) = np_x^{1/2}, \quad \text{and} \quad E(\epsilon | n) = n(1 - p_x^{1/2}). \quad (4)$$

Since the N_x individuals are divided at time x into two distinctly different groups according to their withdrawal status, the joint probability of all the random variables is the product of the two probability functions (1) and (3),

$$f_1 f_2 = p_x^{(s+w/2)}(1 - p_x)^\delta(1 - p_x^{1/2})^\epsilon. \quad (5)$$

1.2. The maximum likelihood estimators and their asymptotic variances.

We are now in a position to use the maximum likelihood principle to obtain the estimator of the probability p_x and its complement q_x , and the asymptotic variance. Taking the logarithm of the joint probability function (5), we have the likelihood function

$$L = (s + \frac{1}{2}w) \ln p_x + \delta \ln(1 - p_x) + \epsilon \ln(1 - p_x^{1/2}). \quad (6)$$

Differentiating (6) with respect to p_x and setting the derivative equal to zero give the likelihood equation

$$(s + \frac{1}{2}w)\hat{p}_x^{-1} - \delta(1 - \hat{p}_x)^{-1} - \frac{\epsilon}{2}\hat{p}_x^{-1/2}(1 - \hat{p}_x^{1/2})^{-1} = 0,$$

which implies

$$\left(N_x - \frac{n}{2}\right)\hat{p}_x + \frac{\epsilon}{2}\hat{p}_x^{1/2} - (s + \frac{1}{2}w) = 0, \quad (7)$$

a quadratic equation in $\hat{p}_x^{1/2}$. Since $\hat{p}_x^{1/2}$ cannot take on negative values, we have the estimators,

$$\hat{p}_x = \left[\frac{-\frac{1}{2}\epsilon + \sqrt{\frac{1}{4}\epsilon^2 + 4(N_x - \frac{1}{2}n)(s + \frac{1}{2}w)}}{2(N_x - \frac{1}{2}n)} \right]^2 \quad (8)$$

⁶A more plausible assumption perhaps is that a withdrawal takes place randomly throughout the interval. However, under this assumption the probability of withdrawing alive is

$$\int_x^{x+1} e^{-(\tau-x)\mu_x} d\tau = -(1 - p_x)/\ln p_x,$$

and the resulting maximum likelihood equation is too unwieldy. When the probability of survival is not too low, the above expression and $p_x^{\frac{1}{2}}$ are close to each other. For $p_x = .70$, for example, $p_x^{\frac{1}{2}} = .837$ and $-(1 - p_x)/\ln p_x = .840$. In the case of extremely high mortality, one should subdivide the interval.

and

$$\hat{q}_x = 1 - \hat{p}_x. \quad (9)$$

The maximum likelihood estimator (8) is not unbiased; however, it is consistent in the sense that when the random variables s , w , and ϵ are replaced with their respective expectations as given by (2) and (4), the resulting expression is identical with the probability p_x . That is,

$$p_x \equiv \left[\frac{-\frac{1}{2}n(1 - p_x^{\frac{1}{2}}) + \sqrt{\frac{1}{4}n^2(1 - p_x^{\frac{1}{2}})^2 + 4(N_x - \frac{1}{2}n)(mp_x + \frac{1}{2}np_x^{\frac{1}{2}})}}{2(N_x - \frac{1}{2}n)} \right]^2.$$

To derive the formula for the asymptotic variance of the estimator \hat{p}_x , (or \hat{q}_x), we find the expectation of the second derivative of (6):

$$E\left(\frac{\partial^2 L}{\partial p_x^2}\right) = -\left[\frac{M_x}{p_x q_x} + \pi\right], \quad (10)$$

where

$$M_x = m + n(1 + p_x^{1/2})^{-1} \quad (11)$$

and

$$\pi = \frac{n}{4(1 + p_x^{1/2})^2 p_x^{3/2}} (1 - p_x^{1/2}). \quad (12)$$

According to the theorem on the asymptotic efficiency of an estimator, the asymptotic variance of \hat{p}_x (or \hat{q}_x) is given by the negative inverse of the expectation (10),

$$\sigma_{\hat{p}_x}^2 = 1/[M_x/(p_x q_x) + \pi]. \quad (13)$$

Usually the quantity π in the denominator of (13) will be small in comparison with the preceding term, and may be neglected to give the approximate formula

$$\sigma_{\hat{p}_x}^2 = p_x q_x / M_x. \quad (14)$$

The sample variance of \hat{p}_x (or \hat{q}_x) is obtained by substituting (8) and (9) in (13) or (14).

Remark 1: The problem discussed here relates to the study in which N_x is large. If N_x is small, one may use the exact time of death of each of the D patients and the exact time of withdrawal of each of the w patients to estimate the probability p_x . In this case, there will be N_x individual observations within the interval $(x, x + 1)$, and obviously it is unnecessary to consider the N_x patients as two distinct groups according to their withdrawal status. Let $t_i \leq 1$ be the time

of death within the interval $(x, x + 1)$ of the i -th death, for $i = 1, \dots, D$, with a probability

$$e^{-t_i \mu_x} \mu_x dt_i, \quad \text{for } i = 1, \dots, D;$$

let $T_j \leq 1$ be the time of withdrawal alive of the j -th withdrawal, for $j = 1, \dots, w$, with a probability

$$e^{-T_j \mu_x}, \quad \text{for } j = 1, \dots, w;$$

then the joint probability function of all the N_x observations becomes

$$p_x^s \prod_{i=1}^D (e^{-t_i \mu_x} \mu_x dt_i) \prod_{j=1}^w (e^{-T_j \mu_x}) = (-\ln p_x)^D p_x^{(s + \sum_{i=1}^D t_i + \sum_{j=1}^w T_j)} \prod_{i=1}^D dt_i.$$

Maximizing the last expression with respect to p_x gives the maximum-likelihood estimator (cf. [18]),

$$\hat{p}_x = \exp \left[-D / \left(s + \sum_{i=1}^D t_i + \sum_{j=1}^w T_j \right) \right].$$

1.3. Observed expectation of life.

A life table for the follow-up subjects can be readily constructed, once \hat{p}_x and \hat{q}_x have been determined from (8) and (9) for each interval of the study period. Let an arbitrary number l_0 denote the number of patients admitted to the study. The number l_x who survive to the exact time x is computed from the formula $l_x = l_0 \hat{p}_0 \hat{p}_1 \cdots \hat{p}_{x-1}$, and $l_x/l_0 = \hat{p}_0 \hat{p}_1 \cdots \hat{p}_{x-1}$ is the estimated x -year survival rate. For a patient alive at time x , the observed expectation of life can be expressed by the equation:

$$\hat{e}_x = a_x + c_{x+1} \hat{p}_x + c_{x+2} \hat{p}_x \hat{p}_{x+1} + \cdots \\ + c_y \hat{p}_x \hat{p}_{x+1} \cdots \hat{p}_{y-1} + c_{y+1} \hat{p}_x \hat{p}_{x+1} \cdots \hat{p}_y + \cdots, \quad (15)$$

where a_x is the average time lived in the interval $(x, x + 1)$ by the patients who die in that interval, and $c_x = 1 - a_{x-1} + a_x$. If, in a study covering a period of y years, there are no survivors remaining from the patients who entered the study in its first year, \hat{p}_{y-1} will be zero, and \hat{e}_x can be computed readily from the collected data. In the typical study, however, there will be w_{y-1} survivors who entered the study in its first year and withdraw alive in the final interval $(y - 1, y)$. In such cases, it is evident from (8) that \hat{p}_{y-1} is greater than zero and the values of $\hat{p}_y, \hat{p}_{y+1}, \dots$ are not observed within the limits of the study. Consequently, \hat{e}_x cannot be computed from equation (15).

It is nevertheless possible to estimate \hat{e}_x with a certain degree of accuracy if w_{y-1} is small. Suppose we rewrite equation (15) in the form

$$\hat{e}_x = a_x + c_{x+1}\hat{p}_x + c_{x+2}\hat{p}_x\hat{p}_{x+1} + \cdots + c_y\hat{p}_x\hat{p}_{x+1} \cdots \hat{p}_{y-1} + \frac{l_y}{l_x} (c_{y+1}\hat{p}_y + c_{y+2}\hat{p}_y\hat{p}_{y+1} + \cdots), \quad (16)$$

where l_y/l_x is written for $\hat{p}_x\hat{p}_{x+1} \cdots \hat{p}_{y-1}$. The problem is to estimate the values of $\hat{p}_y, \hat{p}_{y+1}, \cdots$ in the last term, since the preceding terms can be computed from the data available.

As a first approach, consider a typical interval $(z, z + 1)$ beyond time y with the probability of surviving the interval:

$$p_z = \exp\left(-\int_z^{z+1} \mu_\tau d\tau\right), \quad \text{for } z = y, y + 1, \cdots.$$

If the force of mortality is constant for $z \geq y$, the probability of survival is independent of z and we may write

$$p_z = e^{-\mu} = p, \quad \text{for } z = y, y + 1, \cdots.$$

Under this assumption $c_z = 1$, and we may replace the last term of (16) with $(l_y/l_x) (\hat{p} + \hat{p}^2 + \cdots)$, which converges to $(l_y/l_x)\hat{p}/(1 - \hat{p})$. As a result, we have

$$\hat{e}_x = a_x + c_{x+1}\hat{p}_x + c_{x+2}\hat{p}_x\hat{p}_{x+1} + \cdots + c_y\hat{p}_x\hat{p}_{x+1} \cdots \hat{p}_{y-1} + \frac{l_y}{l_x} [\hat{p}/(1 - \hat{p})].$$

Clearly, \hat{p} may be set equal to \hat{p}_{y-1} if the force of mortality is assumed to be constant beginning with time $(y - 1)$ instead of time y . From the point of view of sample variation, however, it is desirable to base the estimate \hat{p} on as large a sample size as possible. Suppose there exists a time T , for $T < y$, such that $\hat{p}_T, \hat{p}_{T+1}, \cdots$ are approximately equal, thus indicating a constant force of mortality after time T . Then \hat{p} may be set equal to \hat{p}_T , and the formula for the observed expectation of life becomes (cf. [4]):

$$\hat{e}_x = a_x + c_{x+1}\hat{p}_x + c_{x+2}\hat{p}_x\hat{p}_{x+1} + \cdots + c_y\hat{p}_x\hat{p}_{x+1} \cdots \hat{p}_{y-1} + \frac{l_y}{l_x} [\hat{p}_T/(1 - \hat{p}_T)], \quad (17)$$

for $x = 0, \cdots, y - 1$. When a_x is approximated with $\frac{1}{2}$, $c_x = 1$.

Although formula (17) holds for $x = 0, \cdots, y - 1$, it will be apparent to the reader that the smaller the x , the larger the value of l_x ,

and the smaller will be the contribution of the last term. If the ratio $1_y/l_x$ is small, the error in assuming a constant force of mortality beyond y and in the choice of \hat{p}_T will have but little effect on the value of \hat{e}_x .

1.4. *The sample variance of the observed expectation of life.*

To avoid confusion in notation let us denote by α a fixed number and consider the observed expectation of life \hat{e}_α as given in formula (17). It was proven in [5] that the estimated probabilities of surviving any two non-overlapping intervals have a zero covariance, and hence the sample variance of the observed expectation of life may be computed from

$$S_{\hat{e}_\alpha}^2 = \sum_{x \geq \alpha} [\partial \hat{e}_\alpha / \partial \hat{p}_x]^2 S_{\hat{q}_x}^2, \quad (18)$$

where the derivatives are taken at the observed point \hat{p}_x , for $x \geq \alpha$. In the present case, we have

$$[\partial \hat{e}_\alpha / \partial \hat{p}_x] = \hat{p}_{\alpha x} [\hat{e}_{x+1} + (1 - a_x)], \quad \text{for } x \neq T, \quad (19)$$

and

$$\begin{aligned} [\partial \hat{e}_\alpha / \partial \hat{p}_T] &= \hat{p}_{\alpha T} [\hat{e}_{T+1} + (1 - a_T) + \{\hat{p}_{TY} / (1 - \hat{p}_T)^2\}], \quad \text{for } \alpha \leq T, \quad (20) \\ &= \hat{p}_{\alpha T} / (1 - \hat{p}_T)^2, \quad \text{for } \alpha > T, \end{aligned}$$

where $\hat{p}_{\alpha x} = \hat{p}_\alpha \hat{p}_{\alpha+1} \cdots \hat{p}_{x-1}$. Substituting (19) and (20) in (18) gives the sample variance of \hat{e}_α ,

$$\begin{aligned} S_{\hat{e}_\alpha}^2 &= \sum_{\substack{x=\alpha \\ x \neq T}}^{y-1} \hat{p}_{\alpha x}^2 [\hat{e}_{x+1} + (1 - a_x)]^2 S_{\hat{q}_x}^2 \\ &\quad + \hat{p}_{\alpha T}^2 [\hat{e}_{T+1} + (1 - a_T) + \{\hat{p}_{TY} / (1 - \hat{p}_T)^2\}]^2 S_{\hat{q}_T}^2 \end{aligned} \quad (21)$$

for $\alpha \leq T$, and

$$\begin{aligned} S_{\hat{e}_\alpha}^2 &= \sum_{x=\alpha}^{y-1} \hat{p}_{\alpha x}^2 [\hat{e}_{x+1} + (1 - a_x)]^2 S_{\hat{q}_x}^2 \\ &\quad + \{\hat{p}_{\alpha T}^2 / (1 - \hat{p}_T)^4\} S_{\hat{q}_T}^2, \quad \text{for } \alpha > T. \end{aligned} \quad (21a)$$

The value of \hat{p}_x and the sample variance of \hat{q}_x are obtained from formulas (8) and (13), respectively. When a_x is approximated with $\frac{1}{2}$, the quantity $(1 - a_x)$ in formulas (21) and (21a) may be replaced by $\frac{1}{2}$.

PART II. CONSIDERATION OF COMPETING RISKS

2.1. *Relations between net, crude, and partial crude probabilities.*

In a follow-up study, as in general mortality analysis, one may be interested in death due to a specific cause, or to a group of causes.

Depending upon the questions to be answered, the investigator may explore three general types of probabilities of death with respect to a specific cause, or risk:

1. *The crude probability.* The probability of death from a specific cause in the presence of all other risks in a population.
2. *The net probability.* The probability of death if a specific cause were the only cause in effect in the population or, conversely, the probability of death if a specific risk were eliminated from the population.
3. *The partial crude probability.* The probability of death from a specific cause in the presence of all other risks but with a second risk eliminated from the population.

Obviously, in the human population, the net and partial crude probabilities usually cannot be estimated directly except through their relations with the crude probability. The study of such relations is part of the problem of "competing risks", or "multiple-decrement". The subject has been variously discussed in the literature (see, for example, [12], [15], and [19]) and will be reviewed here only by way of introducing notation.

Assume c risks of death (or causes) acting simultaneously on each individual of a population (that is, competing for the life of the individual), and let these risks be denoted by R_1, \dots, R_c . For each risk there is a corresponding force of mortality, $\nu_{r_1}, \dots, \nu_{r_c}$, each of which is a function of time τ , and the sum of these

$$\nu_{r_1} + \dots + \nu_{r_c} = \mu_\tau \quad (22)$$

is then the total force of mortality. Within the time interval $(x, x + 1)$, we shall assume a constant force of mortality for each risk, say $\nu_{rk} = \nu_{xk}$, depending only on x and k , for $x < \tau \leq x + 1$. For all risks, we have $\mu_\tau = \mu_x$, for $x < \tau \leq x + 1$.

Let $Q_{xk}(t)$ be the crude probability that an individual alive at time x will die in the interval $(x, x + t)$, for $0 < t \leq 1$, from cause R_k in the presence of all other risks. It follows directly from addition and multiplication theorems that

$$Q_{xk}(t) = \int_x^{x+t} e^{-(\tau-x)\mu_x} \nu_{xk} d\tau, \quad \text{for } 0 < t \leq 1; \quad k = 1, \dots, c. \quad (23)$$

The first factor of the integrand is the probability of surviving from x to τ when all risks of death are acting, while the second factor is

the instantaneous probability of death from cause R_k . Integrating (23) gives

$$Q_{xk}(t) = \frac{\nu_{xk}}{\mu_x} [1 - e^{-t\mu_x}] = \frac{\nu_{xk}}{\mu_x} [1 - p_x(t)],$$

for $0 < t \leq 1$; $k = 1, \dots, c$. (24)

It is clear from (22) that the sum of the crude probabilities in (24) is equal to the complement of $p_x(t)$, or

$$Q_{x1}(t) + \dots + Q_{xc}(t) + p_x(t) = 1, \text{ for } 0 < t \leq 1. \quad (25)$$

For $t = 1$, we shall abbreviate $Q_{xk}(1)$ to Q_{xk} , etc. For the purpose of this study we are particularly interested in the subinterval $(x, x + \frac{1}{2})$, with

$$Q_{xk}(\frac{1}{2}) = \frac{\nu_{xk}}{\mu_x} [1 - e^{-\mu_x/2}] = Q_{xk}[1 + p_x^{1/2}]^{-1}, \text{ for } k = 1, \dots, c. \quad (26)$$

In this case the sum of the crude probabilities (26) is the complement of $p_x^{1/2}$, the probability of surviving half the interval, and

$$Q_{x1}[1 + p_x^{1/2}]^{-1} + \dots + Q_{xc}[1 + p_x^{1/2}]^{-1} + p_x^{1/2} = 1. \quad (27)$$

When risk R_k acts alone, the net probability that an individual alive at time x will die in the interval $(x, x + 1)$ is

$$q_{xk} = 1 - e^{-\nu_{xk}} = 1 - [e^{-\mu_x}]^{\nu_{xk}/\mu_x}. \quad (28)$$

From formulas (24) and (28) we obtain the relation between the net and crude probabilities

$$q_{xk} = 1 - p_x^{Q_{xk}/q_x}, \text{ for } k = 1, \dots, c. \quad (29)$$

By analogy we can write the net probability of death in the interval $(x, x + 1)$ when risk R_k is eliminated,

$$q_{x..k} = 1 - p_x^{(q_x - Q_{xk})/q_x}, \text{ for } k = 1, \dots, c. \quad (30)$$

Now suppose that R_1 is eliminated as a cause of death, and let $Q_{xk..1}$ be the partial crude probability that an individual alive at time x will die in the interval $(x, x + 1)$ from cause R_k in the presence of all other risks, for $k = 2, \dots, c$. Using a similar reasoning as in the crude probability [eq. (23)], we can write

$$Q_{xk..1} = \int_x^{x+1} e^{-(\tau-x)(\mu_x - \nu_{x1})} \nu_{xk} d\tau$$

$$= \frac{Q_{xk}}{q_x - Q_{x1}} [1 - p_x^{(q_x - Q_{x1})/q_x}], \text{ for } k = 2, \dots, c. \quad (31)$$

Similarly, if risks R_1 and R_2 are eliminated, the partial crude probability that an individual alive at x will die in the interval $(x, x + 1)$ from cause R_k in the presence of all other causes is given by

$$Q_{xk \ 12} = \frac{Q_{xk}}{q_x - Q_{x1} - Q_{x2}} [1 - p_x^{(q_x - Q_{x1} - Q_{x2})/q_x}], \text{ for } k = 3, \dots, c. \quad (32)$$

A detailed discussion on the partial crude probabilities is given in [7].

2.2. *The basic random variables and their joint probability function.*

The identification of the random variables in a follow-up study in the presence of competing risks and the derivation of their joint probability function follows directly from the discussion in Section 1 of Part I. The deaths in each of the two groups according to withdrawal status are further divided by cause of death as shown in Table 1, Part I.

Each of the m individuals not due for withdrawal in the interval $(x, x + 1)$ will fall into one of the $c + 1$ mutually exclusive groups, depending upon whether he survives the interval or dies from cause R_1, \dots, R_c , with the sum of the corresponding probabilities $p_x, Q_{x1}, \dots, Q_{xc}$, equal to unity [eq. (25)]. Thus we have a multinomial case with the probability function

$$f_1 = p_x^s Q_{x1}^{\delta_1} \dots Q_{xc}^{\delta_c}, \quad (33)$$

where s is the number of survivors and δ_k is the number of deaths from cause R_k , for $k = 1, \dots, c$. Their mathematical expectations are given respectively, by

$$E(s \mid m) = mp_x, \text{ and } E(\delta_k \mid m) = mQ_{xk}, \text{ for } k = 1, \dots, c. \quad (34)$$

Each individual in the group of n due for withdrawal in the interval $(x, x + 1)$ has the probability $p_x^{1/2}$ of withdrawing alive and the probability

$$Q_{xk}(\frac{1}{2}) = Q_{xk}(1 + p_x^{1/2})^{-1}, \text{ for } k = 1, \dots, c, \quad (26)$$

of dying from cause R_k . Since $p_x^{1/2}$ and the probabilities in (26) again add up to unity [eq. (27)], the n observations also constitute a multinomial case with the probability function (cf. footnote 6)

$$f_2 = p_x^{w/2} \prod_{k=1}^c [Q_{xk}(1 + p_x^{1/2})^{-1}]^{\epsilon_k}, \quad (35)$$

where w is the number of individuals withdrawing alive and ϵ_k the number of individuals who die from cause R_k before the time of withdrawal. The mathematical expectations are, respectively,

$$E(w \mid n) = np_x^{1/2}, \text{ and } E(\epsilon_k \mid n) = nQ_{xk}(1 + p_x^{1/2})^{-1},$$

for $k = 1, \dots, c. \quad (36)$

Because of the separation of the individuals into two distinct groups at time x according to their withdrawal status, the joint probability of all the random variables in Table 1 is the product of the two joint probabilities (33) and (35):

$$f_1 f_2 = p_x^{s+w/2} \prod_{k=1}^c Q_{xk}^{\delta_k} \prod_{k=1}^c [Q_{xk}(1 + p_x^{1/2})^{-1}]^{\epsilon_k}. \quad (37)$$

Formula (37) may be simplified by rearranging terms and using the relations, $D_k = \delta_k + \epsilon_k$, and $\epsilon = \epsilon_1 + \dots + \epsilon_c$, to give the final form of the joint probability function

$$f_1 f_2 = p_x^{s+w/2} (1 + p_x^{1/2})^{-\epsilon} \prod_{k=1}^c Q_{xk}^{D_k}. \quad (38)$$

2.3. Maximum-likelihood estimators of crude, net, and partial crude probabilities.

We will again use the maximum likelihood principle to obtain the estimators of the probabilities: p_x , Q_{x1} , \dots , Q_{xc} . In this case the likelihood function obtained from (38) is

$$L = (s + \frac{1}{2}w) \ln p_x - \epsilon \ln (1 + p_x^{1/2}) + D_1 \ln \left(1 - p_x - \sum_{k=2}^c Q_{xk} \right) + \sum_{k=2}^c D_k \ln Q_{xk}, \quad (39)$$

where the substitution

$$Q_{x1} = 1 - p_x - \sum_{k=2}^c Q_{xk} \quad (40)$$

has been made. Differentiating the likelihood function (39) with respect to p_x , Q_{x2} , \dots , Q_{xc} , respectively, and setting the derivatives equal to zero, we obtain a system of c simultaneous equations:

$$\partial L / \partial p_x = [(s + \frac{1}{2}w) / \hat{p}_x] - \epsilon / [2\hat{p}_x^{1/2}(1 + \hat{p}_x^{1/2})] - (D_1 / \hat{Q}_{x1}) = 0, \quad (41a)$$

$$\partial L / \partial Q_{xk} = (D_k / \hat{Q}_{xk}) - (D_1 / \hat{Q}_{x1}) = 0, \quad \text{for } k = 2, \dots, c. \quad (41b)$$

From (41b) it can be deduced that

$$D_k / \hat{Q}_{xk} = D / (1 - \hat{p}_x), \quad \text{for } k = 1, \dots, c, \quad (42)$$

and therefore the ratio D_1 / \hat{Q}_{x1} in equation (41a) can be replaced with $D / (1 - \hat{p}_x)$. When this substitution is made and the terms in (41a) are rearranged we have a quadratic equation in $\hat{p}_x^{1/2}$ that is identical to equation (7) in Part I. Hence the estimators \hat{p}_x , in (8), and \hat{q}_x for all causes of death will have the same value as in the simple case where death is investigated without specification to cause, as one would anticipate. Substituting (8) in (42) gives the estimators

$$\hat{Q}_{xk} = (D_k/D)\hat{q}_x, \quad \text{for } k = 1, \dots, c. \quad (43)$$

To obtain the estimators of the net and partial crude probabilities, we substitute (43) in formula (29), (30), (31), and (32), and after simplification,

$$\hat{q}_{xk} = 1 - \hat{p}_x^{D_k/D}, \quad \text{for } k = 1, \dots, c; \quad (44)$$

$$\hat{q}_{x.k} = 1 - \hat{p}_x^{(D-D_k)/D}, \quad \text{for } k = 1, \dots, c; \quad (45)$$

$$\hat{Q}_{xk.1} = [D_k/(D - D_1)][1 - \hat{p}_x^{(D-D_1)/D}], \quad \text{for } k = 2, \dots, c; \quad (46)$$

and

$$\hat{Q}_{xk.12} = [D_k/(D - D_1 - D_2)][1 - \hat{p}_x^{(D-D_1-D_2)/D}], \quad \text{for } k = 3, \dots, c. \quad (47)$$

The estimators given in formulas (44), (45), (46), and (47) are also maximum likelihood estimators because of the invariance property of maximum likelihood estimators.

Remark 2: If there were no withdrawals in the interval $(x, x + 1)$, i.e., if $n = 0$, the problem is reduced to the classical multiple-decrement problem, with s survivors and $D_k = \delta_k$ deaths from cause R_k , for $k = 1, \dots, c$. These random variables will still have a multinomial distribution [eq. (33)], and the formulas for the estimators of p_x , q_x , and Q_{xk} are reduced to

$$\hat{p}_x = s/N_x, \quad (8a)$$

$$\hat{q}_x = D/N_x, \quad (9a)$$

and

$$\hat{Q}_{xk} = D_k/N_x, \quad \text{for } k = 1, \dots, c. \quad (43a)$$

Formulas (44) through (47) may still be used for the estimators of the net and partial crude probabilities, but with \hat{p}_x given by (8a).

Remark 3: The problem of cases lost to the study due to failure of follow-up is still unsolved, and perhaps it has no unique solution. Since the probability that a patient will be lost to follow-up is in part dependent upon the type of a study, assumptions with respect to lost cases may be valid for one study but not for another. If the number of lost cases is small, depending upon the type of study, one of the following assumptions may be made and the data handled accordingly: (1) patients lost will have the same probability of surviving as patients not lost, and may be deleted from the study; (2) all lost cases survive to the close of the study; (3) all die at the time of becoming lost; and (4) becoming lost is another competing risk. If sufficient knowledge of follow-up is unavailable, the fourth alternative is preferred.

2.4. *Asymptotic variance and covariance of the estimators.*

Formulas for the variance and covariance of the estimators may be determined by using the asymptotic property of maximum-likelihood estimators. The inverse of the asymptotic covariance matrix of the estimators, $\hat{p}_x, \hat{Q}_{x2}, \dots, \hat{Q}_{xc}$, is given by

$$\|\Lambda\| = \left[\begin{array}{c|c} \left\| -E\left(\frac{\partial^2 L}{\partial p_x^2}\right) \right\| & \left\| -E\left(\frac{\partial^2 L}{\partial p_x \partial Q_{xk}}\right) \right\| \\ \hline 1 \times 1 & 1 \times (c-1) \\ \hline \left\| -E\left(\frac{\partial^2 L}{\partial p_x \partial Q_{xk}}\right) \right\|' & \left\| -E\left(\frac{\partial^2 L}{\partial Q_{xh} \partial Q_{xk}}\right) \right\| \\ \hline (c-1) \times 1 & (c-1) \times (c-1) \end{array} \right] \quad (48)$$

in which the elements are obtained by differentiating formula (41a) and (41b). Direct calculation gives the following mathematical expectations

$$-E(\partial^2 L / \partial p_x^2) = M_x[(1/p_x) + (1/Q_{x1})] + \pi, \quad (49)$$

$$-E(\partial^2 L / \partial p_x \partial Q_{xk}) = M_x / Q_{x1}, \quad \text{for } k = 2, \dots, c, \quad (50)$$

$$-E(\partial^2 L / \partial Q_{xh} \partial Q_{xk}) = M_x / Q_{x1}, \quad \text{for } h \neq k; h, k = 2, \dots, c, \quad (51)$$

and

$$-E(\partial^2 L / \partial Q_{xk}^2) = M_x[(1/Q_{x1}) + (1/Q_{xk})], \quad \text{for } k = 2, \dots, c, \quad (52)$$

where M_x and π are defined by equation (11) and (12), respectively, of Part I. Substituting the respective expectations into (48), we have

$$\|\Lambda\| = \left[\begin{array}{c|cccc} M_x\left(\frac{1}{p_x} + \frac{1}{Q_{x1}}\right) + \pi & M_x/Q_{x1} & M_x/Q_{x1} & \cdots & M_x/Q_{x1} \\ \hline M_x/Q_{x1} & M_x\left(\frac{1}{Q_{x1}} + \frac{1}{Q_{x2}}\right) & M_x/Q_{x1} & \cdots & M_x/Q_{x1} \\ M_x/Q_{x1} & M_x/Q_{x1} & M_x\left(\frac{1}{Q_{x1}} + \frac{1}{Q_{x3}}\right) & \cdots & M_x/Q_{x1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ M_x/Q_{x1} & M_x/Q_{x1} & M_x/Q_{x1} & \cdots & M_x\left(\frac{1}{Q_{x1}} + \frac{1}{Q_{xc}}\right) \end{array} \right] \quad (53)$$

with its determinant

$$\Lambda = (M_x^c/Q_{x1} \cdots Q_{xc}p_x) + \pi(M_x^{c-1}/Q_{x1} \cdots Q_{xc})(1 - p_x). \quad (54)$$

Denoting the cofactors of the determinant by Λ_{hk} , for $h, k = 1, \dots, c$, the formulas for the asymptotic variance and covariance are given by

$$\sigma_{\hat{p}_x}^2 = \Lambda_{11}/\Lambda = (p_x q_x/M_x)[1/\{1 + (\pi p_x q_x/M_x)\}], \quad (55)$$

$$\sigma_{\hat{Q}_{xk}}^2 = \Lambda_{kk}/\Lambda = [Q_{xk}(1 - Q_{xk})/M_x] \left[\frac{1 + \{\pi p_x(q_x - Q_{xk})/M_x(1 - Q_{xk})\}}{1 + \{\pi p_x q_x/M_x\}} \right],$$

for $k = 2, \dots, c$, (56)

$$\sigma_{\hat{p}_x, \hat{Q}_{xk}} = \Lambda_{1k}/\Lambda = -(p_x Q_{xk}/M_x)[1/\{1 + (\pi p_x q_x/M_x)\}],$$

for $k = 2, \dots, c$, (57)

and

$$\sigma_{\hat{Q}_{xh}, \hat{Q}_{xk}} = \Lambda_{xk}/\Lambda = -(Q_{xh}Q_{xk}/M_x)[\{1 + (\pi p_x/M_x)\}/\{1 + (\pi p_x q_x/M_x)\}],$$

for $h \neq k; \quad h, k = 2, 3, \dots, c$. (58)

Since the term Q_{x1} was not explicitly included in the likelihood function (39), the formulas for the variance of \hat{Q}_{x1} and the covariances between \hat{Q}_{x1} and other estimators were not presented. It is obvious by reason of symmetry, however, that expressions for \hat{Q}_{xk} do start from \hat{Q}_{x1} , which is to say that formulas (56), (57), and (58) hold also for $k = 1$.

The quantities inside the square brackets in formulas (55) through (58) may be approximated with unity when M_x is moderately large. These formulas then reduce to familiar expressions of the multinomial case:

$$\sigma_{\hat{p}_x}^2 = p_x q_x/M_x, \quad (59)$$

$$\sigma_{\hat{Q}_{xk}}^2 = Q_{xk}(1 - Q_{xk})/M_x, \quad \text{for } k = 1, \dots, c, \quad (60)$$

$$\sigma_{\hat{p}_x, \hat{Q}_{xk}} = -p_x Q_{xk}/M_x, \quad \text{for } k = 1, \dots, c, \quad (61)$$

and

$$\sigma_{\hat{Q}_{xh}, \hat{Q}_{xk}} = -Q_{xh}Q_{xk}/M_x, \quad \text{for } h \neq k; \quad h, k = 1, 2, \dots, c. \quad (62)$$

Formulas for the asymptotic variance and covariance of the estimators of the net and partial crude probabilities can be obtained with the same approach as employed in [7]. To save space, only two formulas are presented below.

$$\sigma_{\hat{q}_{x,k}}^2 = [(1 - q_{x,k})^2/M_x p_x q_x] \cdot [p_x \ln(1 - q_{xk}) \ln(1 - q_{x,k}) + (q_x - Q_{xk})^2], \quad (63)$$

for $k = 1, \dots, c$, for the net probability of death when risk R_k is eliminated, and

$$\begin{aligned} \sigma_{\hat{q}_{xk.1}}^2 = & [(q_x - Q_{x1} - Q_{xk}) / \{M_x(q_x - Q_{x1})Q_{xk}\}] Q_{xk.1}^2 \\ & + [\{Q_{xk.1}(q_x - Q_{x1}) - Q_{xk}\}^2 / \{M_x p_x q_x (q_x - Q_{x1})\}] \\ & \cdot [(q_x - Q_{x1}) + Q_{x1} p_x (\ln p_x / q_x)^2], \end{aligned} \quad (64)$$

for $k = 2, \dots, c$, for the partial crude probability.

AN EXAMPLE OF LIFE TABLE CONSTRUCTION FOR THE FOLLOW-UP POPULATION

The application of the methods developed in Parts I and II will be illustrated with data collected by the Tumor Registry of the California State Department of Public Health. The material selected consists of 5982 patients⁷ admitted to certain California hospitals and clinics between January 1, 1942, and December 31, 1954, with a diagnosis of cancer of the cervix uteri. For the purpose of this illustration, the latter date is taken as the common closing date of the study; the date of entrance to follow-up for each patient is the date of hospital admission.

The first step is to construct a table similar to Table 2, showing the survival experience of the patients grouped according to their withdrawal status. The interval length selected (column 1) will depend upon the nature of the investigation; in this case a fixed length of one year was convenient and satisfactory. The total number of patients admitted to the study is entered as N_0 in the first line of column 2, which in this example is 5982. To determine their withdrawal status in the first interval (0, 1) the patients were separated into two groups: admissions before 1954, and consequently at least one year before the close of the study; and admissions during the year 1954, all due for withdrawal since the study was terminated before their first anniversary. Of the patients admitted prior to 1954, s_0 (4030 in column 3) survived to their first anniversary and δ_0 (1287) died during the first year. The deaths were further divided by cause into δ_{01} (1105) deaths due to cancer of the cervix uteri and δ_{02} (182) deaths from all other causes. The survival status of the 1954 admissions is determined at the close of the study, as it is for patients due for withdrawal in any interval. In this study, w_0 (576) patients withdrew alive in the first interval, and ϵ_0 (89) patients died before the closing date. These deaths were again divided by cause into ϵ_{01} (70, column 9) and ϵ_{02} (19, column 10).

⁷An additional 251 cases of uncertain survival status were deleted from this illustration.

TABLE 3
SURVIVAL EXPERIENCE AFTER DIAGNOSIS OF CANCER OF THE CERVIX UTERI:
THE MAIN LIFE TABLE FUNCTIONS AND THEIR STANDARD ERRORS

Interval since diagnosis (years)	x -year survival rate \hat{p}_{0x}		Estimated probability of death in interval $(x, x + 1)$		Observed expectation of life at x	
	1000 \hat{p}_{0x}	1000 $S_{\hat{p}_{0x}}$	1000 \hat{q}_x	1000 $S_{\hat{q}_x}$	\hat{e}_x	$S_{\hat{e}_x}$
$x - x + 1$	(2)	(3)	(4)	(5)	(6)	(7)
0-1	1000.00	0.00	242.54	5.69	12.90	2.83
1-2	757.46	5.80	181.43	6.26	15.86	3.74
2-3	620.03	6.65	103.03	5.95	18.27	4.57
3-4	556.15	7.01	85.76	6.38	19.31	5.09
4-5	508.46	7.33	64.13	6.50	20.08	5.56
5-6	475.85	7.61	58.20	7.23	20.42	5.94
6-7	448.15	7.95	43.76	7.34	20.65	6.31
7-8	428.54	8.29	43.20	8.45	20.57	6.60
8-9	410.03	8.71	33.69	8.85	20.48	6.89
9-10	396.22	9.17	46.55	12.15	20.17	7.13
10-11	377.77	9.98	43.85	14.30	20.13	7.47
11-12	361.21	10.97	51.06	20.30	20.03	7.81
12-13	342.77	12.73	00.00	00.00	20.08	7.79
13	342.77	12.73	—	—	19.08	7.79

Source: California Tumor Registry, Department of Public Health, State of California.

The second interval began with the 4030 survivors from the first interval, which is entered as N_1 in line 2 of column 2. All 1953 admissions included in N_1 were due for withdrawal in the second interval.

The main life table functions and the corresponding sample standard errors as shown in Table 3 are determined from the data given in Table 2. The x -year survival rate \hat{p}_{0x} is by definition equal to l_x divided by the radix l_0 , or $\hat{p}_0\hat{p}_1, \dots, \hat{p}_{x-1}$. The sample variance of \hat{p}_{0x} is computed from a formula given in a previous publication [6] (see also [13]):

$$S_{\hat{p}_{0x}}^2 = \hat{p}_{0x}^2 \sum_{u=0}^{x-1} \hat{p}_u^{-2} S_{\hat{q}_u}^2.$$

Formulas (13) and (14) were both used to compute the sample standard error of \hat{q}_x , with numerical results that were almost identical to the fourth decimal place. The figures appearing in column 5 of Table 3 were obtained by formula (14). The observed expectation of life was determined from formula (17), for which \hat{p}_T was set equal to \hat{p}_{11} .

TABLE 4
SURVIVAL EXPERIENCE AFTER DIAGNOSIS OF CANCER OF THE CERVIX UTERI:
ESTIMATED CRUDE AND NET PROBABILITIES OF DEATH FROM CANCER
OF THE CERVIX UTERI AND FROM OTHER CAUSES

Interval since diagnosis (years)	Estimated probability of surviving interval ($x, x + 1$)	Estimated crude probabilities of death in interval ($x, x + 1$) from		Estimated net probabilities of death in interval ($x, x + 1$) when	
		Cervix cancer	Other causes	Cervix cancer Acting alone	Cervix cancer Eliminated
$x - x + 1$	$1000 \hat{p}_x$	$1000 \hat{Q}_{x1}$	$1000 \hat{Q}_{x2}$	$1000 \hat{q}_{x1}$	$1000 \hat{q}_{x2}$
(1)	(2)	(3)	(4)	(5)	(6)
0-1	757.46	207.11	35.43	211.17	39.77
1-2	818.57	155.97	25.46	158.11	27.71
2-3	896.97	84.65	18.38	85.46	19.22
3-4	914.24	62.89	22.87	63.63	23.63
4-5	835.87	44.40	19.73	44.85	20.19
5-6	941.80	25.76	32.44	26.19	32.87
6-7	956.24	23.17	20.59	23.41	20.84
7-8	956.80	22.47	20.73	22.70	20.97
8-9	966.31	14.44	19.25	14.58	19.39
9-10	953.45	29.93	16.62	30.18	16.88
10-11	956.15	24.36	19.49	24.60	19.73
11-12	948.94	17.02	34.04	17.32	34.34
12-13	1000.00	—	—	—	—

Source: California Tumor Registry, Department of Public Health, State of California.

Table 4 shows the estimated probability of surviving each interval and the estimated crude and net probabilities of death from R_1 , cancer of the cervix uteri and R_2 , all other causes of death. Since only two risks are studied, the probability q_{x2} is equal to q_{x1} , the net probability of death when cancer of the cervix uteri is eliminated as a risk of death from the population. For each interval the sum of \hat{p}_x , \hat{Q}_{x1} , and \hat{Q}_{x2} is unity, and the estimated net probability \hat{Q}_{xk} is always greater than the corresponding crude probability \hat{Q}_{xk} .

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to George Linden and his staff at the California Tumor Registry, Department of Public

Health, State of California, for their cooperation and assistance in making their follow-up data available for our use. My thanks are also due to William F. Taylor for reading the paper.

REFERENCES

- [1] Berkson, J. and Gage, R. P. [1950]. Calculation of survival rates for cancer. *Proc. Staff Meetings Mayo Clinic* 25, 270-86.
- [2] Berkson, J. and Gage, R. P. [1952]. Survival curve for cancer patients following treatment. *J. Amer. Stat. Assoc.* 47, 501-15.
- [3] Boag, J. W. [1949]. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. Roy. Stat. Soc.* 11, 15-53.
- [4] Chiang, C. L. [1958]. An application of stochastic processes to the life table and standard error of age-adjusted rates (abstract). *Biometrics* 14, 133-4.
- [5] Chiang, C. L. [1960]. A stochastic study of the life table and its applications: I. Probability distributions of the biometric functions. *Biometrics* 16, 618-35.
- [6] Chiang, C. L. [1960]. 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, 221-38.
- [7] Chiang, C. L. (unpublished). On the probability of death from specific causes in the presence of competing risks. (To be published in the *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*).
- [8] Dorn, H. [1950]. Methods of analysis for follow-up studies. *Human Biology* 22, 238-48.
- [9] Elveback, L. [1958]. Estimation of survivorship in chronic disease: the "Actuarial" method. *J. Amer. Stat. Assoc.* 53, 420-40.
- [10] Epstein, B. and Sobel, M. [1953]. Life testing. *J. Amer. Stat. Assoc.* 48, 486-502.
- [11] Fix, E. [1951]. Practical implications of certain stochastic models on different methods of follow-up studies. (Presented at annual meeting of the Western Branch, A. P. H. A., Nov., 1951).
- [12] Fix, E. and Neyman, J. [1951]. A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology* 23, 205-41.
- [13] Greenwood, M. [1926]. A report on the natural duration of cancer. *Reports on Public Health and Medical Subjects, No. 33*, 1-26. His Majesty's Stationary Office.
- [14] Harris, T. E., Meier, P. and Tukey, J. W. [1950]. Timing of the distribution of events between observations. *Human Biology* 22, 249-70.
- [15] Jorden, C. W. [1952]. *Life Contingencies*. Society of Actuaries, Chicago.
- [16] Kaplan, E. L. and Meier, P. [1958]. Nonparametric estimation from incomplete observations. *J. Amer. Stat. Assoc.* 53, 457-81.
- [17] Karn, M. N. [1933]. A further study of methods of constructing life tables when certain causes of death are eliminated. *Biometrika* 25, 91.
- [18] Littell, A. S. [1952]. Estimation of the T-year survival rate from follow-up studies over a limited period of time. *Human Biology* 24, 87-116.
- [19] Neyman, J. [1950]. *First course in probability and statistics*. Henry Holt, New York.