ON THE PROBABILITY OF DEATH FROM SPECIFIC CAUSES IN THE PRESENCE OF COMPETING RISKS

CHIN LONG CHIANG
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

The characteristic of a mortality study is that the basic event, the death of an individual, is not repetitive. Suppose we wish to assess the congenital malformation as a cause of infant death; how shall we count the malformed child who dies of tuberculosis in his first year of life? The risk of death due to the congenital malformation no longer exists, but neither can he survive the condition until his first birthday. It is clear that the evaluation of congenital malformation as a cause of death must allow for the effect of all causes operating in the human population. To take a second example, perhaps the problem is to estimate the length of time between diagnosis and death due to coronary heart disease. A study group of diagnosed cases is formed, but a number of them will die from other diseases before the observation period is ended. How can we correct our estimate for the competing risks? In another study, we may be interested in the susceptibility of individuals with a certain chronic condition to other diseases. Is a diagnosed case of arteriosclerotic heart disease more likely to die from cancer than a person without the heart condition? How can we take into account the competition between arteriosclerotic heart disease and cancer for the life of the heart patient?

To answer these and similar questions, the investigator may explore three general types of probabilities of death with respect to a specific risk: (1) the crude probability, (2) the net probability, and (3) the partial crude probability. Symbolically, they are defined as follows. To describe death from a specific risk, say R_k, we have the crude probability

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

the net probabilities

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

$q_{x.k} = P\{$an individual alive at x will die in the interval $(x, x + 1)$ if R_k is eliminated as a risk of death from the population$\}$;

This investigation was carried out during the tenure of a Special Research Fellowship from the National Heart Institute, U.S. Public Health Service.

and the partial crude probabilities

$Q_{xk.1} = P\{$an individual alive at x will die in the interval $(x, x + 1)$ from R_k if R_1 is eliminated from the population$\}$,

$Q_{xk.12} = P\{$an individual alive at x will die in the interval $(x, x + 1)$ from R_k, if R_1 and R_2 are eliminated from the population$\}$.

When risk of death is not specified, we have the probabilities

$p_x = P\{$an individual alive at x will survive the interval $(x, x + 1)\}$;

$q_x = P\{$an individual alive at x will die in the interval $(x, x + 1)\}$;

and obviously $p_x + q_x = 1$. These probabilities are basic to the study of survival and analysis of the life table. In the human population the net and the partial crude probabilities are not observable except as they are related to the crude probability. A study of their relationships is part of the problem of competing risks.

The notion of net and crude probabilities is not new and their relationships have been discussed variously in the literature. An early theoretical approach to the problem was made by Bernoulli [9], and in 1874 Makeham explored some practical applications of the theory [8]. Interesting investigations have also been made by Karn [6], Fix and Neyman [4], Cornfield [3], Kimball [7], and Jordan [5]. Formulas expressing relations between net and crude probabilities were developed either under the assumption of a constant force of mortality (instantaneous death probability) for a given risk or under the assumption of a uniform distribution of deaths. We shall review these relations under a weaker assumption.

Partial crude probabilities have not received as much attention as they deserve in view of their usefulness in the study of cause-specific mortality, to which they are sometimes indispensable. Relations between the partial crude and the corresponding crude probabilities will be developed in this paper. The problem of estimation will also be treated. A brief account will be made of the medical follow-up study to illustrate an application of the theory of competing risks.

2. Relations between crude, net, and partial crude probabilities

Suppose there are c risks (or causes) of death acting simultaneously on each individual of a population and let these risks be denoted by $R_1, \cdots, R_c$. For each risk R_k there is a corresponding force of mortality, $\nu_{\tau k}$, with the probability $\nu_{\tau k}\Delta\tau + o(\Delta\tau)$ that an individual alive at time τ will die in the infinitesimal time interval $(\tau, \tau + \Delta\tau)$, for $k = 1, \cdots, c$. The sum

$$\nu_{\tau 1} + \cdots + \nu_{\tau c} = \mu_\tau \tag{1}$$

is the total force of mortality. While for each risk R_k, the force of mortality $\nu_{\tau k}$ is a function of time τ, we shall assume that, within the time interval $(x, x + 1)$, the ratio

$$\frac{\nu_{\tau k}}{\mu_\tau} = r_{xk}, \qquad x < \tau \leqq x + 1; \quad k = 1, \cdots, c, \tag{2}$$

is independent of time τ, but is a function of the interval $(x, x+1)$ and risk R_k only. Assumption (2) permits the absolute magnitude of the cause-specific force of mortality to vary at any instant, but requires that it remain a constant proportion of the total force of mortality throughout an interval.

Consider death without specification of cause. Using the pure death process [1] it is easy to show that the probability that an individual alive at x will survive the interval $(x, x+1)$ is given by

$$p_x = \exp\left(-\int_x^{x+1} \mu_\tau \, d\tau\right), \tag{3}$$

and the probability of death in the interval is $q_x = 1 - p_x$. Similarly, when R_k is the only risk in effect, the net probability of death in the interval is given by

$$q_{xk} = 1 - \exp\left(-\int_x^{x+1} \nu_{\tau k} \, d\tau\right), \qquad k = 1, \cdots, c. \tag{4}$$

The crude probability of death from R_k may be written as

$$Q_{xk} = \int_x^{x+1} \exp\left(-\int_x^{\tau} \mu_T \, dT\right) \nu_{\tau k} \, d\tau, \qquad k = 1, \cdots, c, \tag{5}$$

where the first factor of the integrand is the probability of surviving from x to τ when all risks of death are acting and the second factor is the instantaneous death probability from risk R_k at time τ. Using assumption (2) of a constant relative risk, we have

$$\begin{aligned} Q_{xk} &= \frac{\nu_{\tau k}}{\mu_\tau} \int_x^{x+1} \exp\left(-\int_x^{\tau} \mu_T \, dT\right) \mu_\tau \, d\tau \\ &= \frac{\nu_{\tau k}}{\mu_\tau}\left[1 - \exp\left(-\int_x^{x+1} \mu_\tau \, d\tau\right)\right] = \frac{\nu_{\tau k}}{\mu_\tau} q_x, \end{aligned} \tag{6}$$

or

$$\frac{\nu_{\tau k}}{\mu_\tau} = \frac{Q_{xk}}{q_x}, \qquad x < \tau \leqq x+1; \quad k = 1, \cdots, c. \tag{7}$$

Equation (7) is obvious also from an intuitive viewpoint. If the ratio of the cause-specific force of mortality to the total force is a constant throughout an interval, this constant should also be equal to the ratio of the corresponding probabilities over the entire interval. Equations (6) and (1) imply also a trivial equality,

$$Q_{x1} + \cdots + Q_{xc} = q_x. \tag{8}$$

Now the net probability of death as defined in (4) is rewritten as

$$q_{xk} = 1 - \exp\left(-\frac{\nu_{\tau k}}{\mu_\tau}\int_x^{x+1} \mu_\tau \, d\tau\right). \tag{9}$$

Using equation (7), we have from (9) the relation between the net and the crude probabilities

$$q_{xk} = 1 - p_x^{Q_{xk}/q_x}, \qquad k = 1, \cdots, c. \tag{10}$$

The net probability of death when cause R_k is eliminated can be derived in the same way. By reason of symmetry we have

$$(11)\qquad \begin{aligned} q_{x.k} &= 1 - \exp\left(-\int_x^{x+1} (\mu_\tau - \nu_{\tau k})\, d\tau\right) \\ &= 1 - p_x^{(q_x - Q_{xk})/q_x}, \qquad k = 1, \cdots, c. \end{aligned}$$

Although relations (10) and (11) have been known for quite some time, their derivation is usually based on the assumption of a constant force of mortality or of a uniform distribution of deaths in the interval $(x, x + 1)$. These assumptions are often violated. In the first year of life, for example, the force of mortality tends to be a concave function of time; it starts with a maximum value at birth and decreases rapidly with increasing age. The assumption of a constant force of mortality is certainly not realistic in this case and the validity of any relationship based on such an assumption must also be questionable. It is comforting to note, therefore, that relations (10) and (11) still hold under the relatively weak assumption in (2).

The following observations may be of interest. First, because of the absence of competing risks, the net probability q_{xk} is always greater than the corresponding crude probability Q_{xk}. The inequality $\nu_{\tau k} < \mu_\tau$ implies

$$(12)\qquad \exp\left(-\int_x^\tau \nu_{Tk}\, dT\right) > \exp\left(-\int_x^\tau \mu_T\, dT\right),$$

whatever may be $x < \tau \leqq x + 1$; hence

$$(13)\qquad \begin{aligned} q_{xk} &= \int_x^{x+1} \exp\left(-\int_x^\tau \nu_{Tk}\, dT\right) \nu_{\tau k}\, d\tau \\ &> \int_x^{x+1} \exp\left(-\int_x^\tau \mu_T\, dT\right) \nu_{\tau k}\, d\tau = Q_{xk}. \end{aligned}$$

Second, if two risks R_h and R_k are such that the corresponding crude probabilities have the relation $Q_{xh} > Q_{xk}$, it is easily seen from (10) and (11) that the corresponding net probabilities will have the relations $q_{xh} > q_{xk}$ and $q_{x.h} < q_{x.k}$. Third, the probability of surviving more than one interval is the product of the probabilities of surviving the subintervals. The probability of surviving the interval $(0, y)$ when risk R_k is eliminated, for example, may be computed from

$$(14)\qquad \exp\left[-\int_0^y (\mu_\tau - \nu_{\tau k})\, d\tau\right] = \prod_{x=0}^{y-1} (1 - q_{x.k}) = \prod_{x=0}^{y-1} p_x^{(q_x - Q_{xk})/q_x}.$$

And fourth, if a constant relative risk (2) is assumed over the entire interval $(0, y)$, then the net and crude probabilities over the interval $(0, y)$ will have relationships similar to (10) and (11). Thus the probability of surviving the interval $(0, y)$ when risk R_k is eliminated is given by

$$(15)\qquad \exp\left[-\int_0^y (\mu_\tau - \nu_{\tau k})\, d\tau\right] = p_{0y}^{(1 - p_{0y} - Q_{0yk})/(1 - p_{0y})},$$

where p_{0y}, the probability of surviving from 0 to y when all risks are acting, is given by

$$p_{0y} = \exp\left(-\int_0^y \mu_\tau \, d\tau\right), \tag{16}$$

and Q_{0yk} is the crude probability of death in the interval $(0, y)$ from risk R_k in the presence of competing risks in the population. Clearly (15) is a special case and may be rewritten in the form of (14). Since p_{0y} is equal to the product $p_0p_1 \cdots p_{y-1}$ and since in the present case (7) implies the equalities

$$\frac{1 - p_{0y} - Q_{0yk}}{1 - p_{0y}} = \frac{q_x - Q_{xk}}{q_x}, \qquad x = 0, \cdots, y - 1, \tag{17}$$

the right side of (15) becomes

$$p_{0y}^{(1-p_{0y}-Q_{0yk})/(1-p_{0y})} = \left(\prod_{x=0}^{y-1} p_x\right)^{(1-p_{0y}-Q_{0yk})/(1-p_{0y})} = \prod_{x=0}^{y-1} p_x^{(q_x - Q_{xk})/q_x}, \tag{18}$$

which is identical with the last expression in (14).

Suppose now that R_1 is eliminated from the population as a risk 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 risk R_k in the presence of all other risks, for $k = 2, \cdots, c$. We wish to express $Q_{xk.1}$ in terms of the probabilities p_x, q_x, and the crude probabilities Q_{x1} and Q_{xk}. Using the multiplication and addition theorems as in (5), we have

$$Q_{xk.1} = \int_x^{x+1} \exp\left[-\int_x^\tau (\mu_T - \nu_{T1})\, dT\right] \nu_{\tau k}\, d\tau. \tag{19}$$

To simplify (19), we note from (7) that the ratio $\nu_{\tau k}/(\mu_\tau - \nu_{\tau 1})$ is equal to $Q_{xk}/(q_x - Q_{x1})$ and is independent of time τ, for $x < \tau \leqq x + 1$. Now the partial crude probability may be rewritten as

$$\begin{aligned} Q_{xk.1} &= \frac{\nu_{\tau k}}{\mu_\tau - \nu_{\tau 1}} \int_x^{x+1} \exp\left[-\int_x^\tau (\mu_T - \nu_{T1})\, dT\right](\mu_\tau - \nu_{\tau 1})\, d\tau \\ &= \frac{Q_{xk}}{q_x - Q_{x1}}\left\{1 - \exp\left[-\int_x^{x+1} (\mu_\tau - \nu_{\tau 1})\, d\tau\right]\right\} = \frac{Q_{xk}}{q_x - Q_{x1}}\, q_{x.1}. \end{aligned} \tag{20}$$

Substituting (11) for $k = 1$ in (20) gives the final relation

$$Q_{xk.1} = \frac{Q_{xk}}{q_x - Q_{x1}}\left[1 - p_x^{(q_x - Q_{x1})/q_x}\right], \qquad k = 2, \cdots, c. \tag{21}$$

The sum of $Q_{xk.1}$, for $k = 2, \cdots, c$, is equal to the net probability of death when risk R_1 is eliminated from the population, for equation (8) implies

$$\sum_{k=2}^{c} Q_{xk} = q_x - Q_{x1}; \tag{22}$$

hence

$$\sum_{k=2}^{c} Q_{xk.1} = \sum_{k=2}^{c} \frac{Q_{xk}}{q_x - Q_{x1}}\left[1 - p_x^{(q_x - Q_{x1})/q_x}\right] = 1 - p_x^{(q_x - Q_{x1})/q_x} = q_{x.1}, \tag{23}$$

as one might have anticipated.

Generalization of formula (21) to apply to cases in which more than one risk is eliminated is straightforward. If, for example, both R_1 and R_2 are eliminated as risks of death, the partial crude probability that an individual alive at time x will die in the interval $(x, x+1)$ from risk R_k is given by

$$Q_{xk.12} = \frac{Q_{xk}}{q_x - Q_{x1} - Q_{x2}} [1 - p_x^{(q_x - Q_{x1} - Q_{x2})/q_x}], \qquad k = 3, \cdots, c. \tag{24}$$

In the discussion of these three types of probabilities, it is assumed that both q_x and p_x are greater than zero but less than unity. For if q_x were zero ($p_x = 1$), then Q_{xk} will also be zero, for $k = 1, \cdots, c$; therefore the ratios Q_{xk}/q_x, $Q_{xk}/(q_x - Q_{x1})$, and $(q_x - Q_{x1})/q_x$, and consequently formulas (10), (11), (21), and (24), will all cease to be meaningful. In other words, if an individual were certain to survive an interval, it would be meaningless to speak of his chances of dying from a specific risk. On the other hand, if p_x were zero ($q_x = 1$), we can see from (3) that the integral $\int_x^{x+1} \mu_\tau \, d\tau$ would approach infinity, which is fortunately an unrealistic outlook for the human population. Thus the crude, net, and partial crude probabilities are defined only for positive values of p_x and q_x.

REMARK 1. An important assumption made in this paper is that the forces of mortality are additive [equation (1)]. It is assumed that risks of death act independently of each other and that the elimination of one disease has no effect on the force of mortality of other risks. How true this assumption is must depend in part on the disease in question and its complex relationship with other diseases in the particular host population. While the proper experimental verification of the assumption is not practicable in the human population, the following discussion may indicate how this important aspect of the problem might be approached.

Consider two groups of individuals. Each individual in the first group is known to be free from disease R_1 at the beginning of the time interval $(x, x+1)$, although he may contract R_1 and die before the interval ends. The forces of mortality operating in the group are denoted by $\nu_{\tau k}$ and the corresponding probabilities of death by q_x, q_{xk}, $q_{x.k}$, and $Q_{xk.1}$, for $k = 1, \cdots, c$. These probabilities are similarly defined and have the same relations as in equations (1) through (21). The second group consists of individuals each of whom is known to be affected with disease R_1 at the beginning of the interval $(x, x+1)$, and each is subject to the forces of mortality $\nu_{\tau k}^{(1)}$, for $k = 1, \cdots, c$. The sum

$$\nu_{\tau 1}^{(1)} + \cdots + \nu_{\tau c}^{(1)} = \mu_\tau^{(1)} \tag{25}$$

is the total force of mortality. The net probabilities $q_{xk}^{(1)}$ and $q_{x.k}^{(1)}$, the crude probability $Q_{xk}^{(1)}$, the partial crude probability $Q_{xk.1}^{(1)}$, and the probabilities $q_x^{(1)}$ and $p_x^{(1)}$ have the same relations as those in the first group; for example, the net probability of death if risk R_k acts alone is given by

$$q_{xk}^{(1)} = 1 - \exp\left[-\int_x^{x+1} \nu_{\tau k}^{(1)} \, d\tau\right] = 1 - [p_x^{(1)}]^{Q_{xk}^{(1)}/q_x^{(1)}}, \qquad k = 1, \cdots, c, \tag{26}$$

where

$$q_x^{(1)} = 1 - p_x^{(1)} = 1 - \exp\left[-\int_x^{x+1} \mu_\tau^{(1)}\, d\tau\right], \tag{27}$$

and $Q_{xk}^{(1)}$ is the corresponding crude probability.

Our interest is in the relative changes in the forces of mortality due to some other risk operating in the two groups, say in a comparison of $\nu_{\tau 2}$ and $\nu_{\tau 2}^{(1)}$, or q_{x2} and $q_{x2}^{(1)}$ due to risk R_2. From a practical viewpoint, this may determine whether an individual affected with one disease (R_1, in this case) is more likely to die from a second disease (R_2) than an individual not affected with R_1. Theoretically, if $q_{x2}^{(1)}$ is shown to be equal to q_{x2}, then it may indicate that elimination of R_1 from the population will have little effect on the force of mortality of R_2. If the equality $q_{xk}^{(h)} = q_{xk}$ holds whatever may be $h \neq k$, for $h, k = 1, \cdots, c$, then the assumption of additivity as given in (1) probably is not too strong.

Actually, for a particular risk in question, R_1 say, assumption (1) is not necessary. In order that relations (10) and (11) hold for $k = 1$, it is only necessary that the total force of mortality be partitioned into two additive components: the component associated with R_1 and the component not associated with R_1. Symbolically,

$$\mu_x = \nu_{x1} + (\mu_x - \nu_{x1}). \tag{28}$$

Similarly, the necessary assumption for relation (21) to hold for $k = 2$ is that

$$\mu_x = \nu_{x1} + \nu_{x2} + (\mu_x - \nu_{x1} - \nu_{x2}). \tag{29}$$

Some indication for the validity of these assumptions may be obtained from the comparison between $q_{x.1}$ and $q_{x.1}^{(1)}$, and between $Q_{x2.1}$ and $Q_{x2.1}^{(1)}$.

3. Estimation of the crude, net, and partial crude probabilities

Consider a group of N_x individuals alive at time x, who are to be observed over the interval $(x, x + 1)$. By the end of the interval there will be s_x survivors and D_{xk} deaths from cause R_k, for $k = 1, \cdots, c$; or a total of

$$D_{x1} + \cdots + D_{xc} = D_x \tag{30}$$

deaths from all causes. Obviously

$$s_x + D_{x1} + \cdots + D_{xc} = N_x, \tag{31}$$

and the sum of the corresponding probabilities [compare equation (8)]

$$p_x + Q_{x1} + \cdots + Q_{xc} = 1. \tag{32}$$

It follows that the number of survivors and the number of deaths from each cause have a multinomial distribution, and the probabilities in (32) will have the estimators

$$\hat{p}_x = \frac{s_x}{N_x}, \qquad \hat{q}_x = \frac{D_x}{N_x}, \tag{33}$$

and

$$(34)\qquad \hat{Q}_{xk} = \frac{D_{xk}}{N_x}, \qquad k = 1, \cdots, c,$$

with variances

$$(35)\qquad \sigma^2_{\hat{p}_x} = \sigma^2_{\hat{q}_x} = \frac{1}{N_x} p_x q_x$$

and

$$(36)\qquad \sigma^2_{\hat{Q}_{xk}} = \frac{1}{N_x} Q_{xk}(1 - Q_{xk}), \qquad k = 1, \cdots, c;$$

and covariances

$$(37)\qquad \sigma_{\hat{p}_x, \hat{Q}_{xk}} = -\frac{1}{N_x} p_x Q_{xk}, \qquad k = 1, \cdots, c,$$

and

$$(38)\qquad \sigma_{\hat{Q}_{xh}\hat{Q},_{xk}} = -\frac{1}{N_x} Q_{xh} Q_{xk}, \qquad h \neq k; \quad h, k = 1, \cdots, c.$$

REMARK 2. In the above discussion N_x is treated as a constant. If the problem is visualized from a point in time prior to x, then the number of individuals surviving to x is a random variable. In this case, the estimators in (33) and (34) remain unchanged, but the quantity $1/N_x$ in formulas (35) through (38) for the variances and covariances of the estimators should be replaced by its expectation $E(1/N_x)$ as shown in [1]. Further, if two or more time intervals are considered simultaneously, it will be more convenient to have an explicit formula for the generating function of all the random variables, $s_x, D_{x1}, \cdots, D_{xc}$, for $x = 0, \cdots, w$, say,

$$(39)\qquad G = E\left[\prod_{x=0}^{w} \left\{ t_x^{s_x} \prod_{k=1}^{c} t_{xk}^{D_{xk}} \right\} | N_0 \text{ at } x = 0 \right],$$

where $s_x = N_{x+1}$, the number of individuals alive at the beginning of the interval $(x + 1, x + 2)$. It is easy to show by induction (see also [1]) that the generating function (39) is given by

$$(40)\qquad G = \left\{ \sum_{k=1}^{c} Q_{0k} t_{0k} + \sum_{y=1}^{w} \left[\prod_{x=0}^{y-1} p_x t_x \left(\sum_{k=1}^{c} Q_{yk} t_{yk} \right) \right] + \prod_{x=0}^{w} p_x t_x \right\}^{N_0}.$$

Direct computation from (40) gives the joint probability function

$$(41)\qquad \prod_{x=0}^{w} \frac{N_x!}{s_x! D_{x1}! \cdots D_{xc}!} p_x^{S_x} Q_{x1}^{D_{x1}} \cdots Q_{xc}^{D_{xc}}$$

and the joint moments; for example,

$$(42)\qquad \sigma_{N_x N_y | N_0} = N_0 p_{0y}(1 - p_{0x}), \qquad 0 < x \leqq y,$$

where p_{0z}, the probability of surviving from 0 to z, is given by (16). Although the random variables N_x and N_y are linearly correlated, the estimators of the probabilities over two time intervals are not. For example, the estimators $\hat{p}_x$ and $\hat{p}_y$ have a zero covariance, but they are not independently distributed. A detailed discussion on this point is given in [1].

To derive formulas for the estimators of the net and the partial crude probabilities, we substitute (33) and (34) in formulas (10), (11), (21), and (24). As a result we obtain

$$\hat{q}_{xk} = 1 - \left(\frac{s_x}{N_x}\right)^{D_{xk}/D_x}, \qquad k = 1, \cdots, c, \tag{43}$$

$$\hat{q}_{x.k} = 1 - \left(\frac{s_x}{N_x}\right)^{(D_x - D_{xk})/D_x}, \qquad k = 1, \cdots, c, \tag{44}$$

$$\hat{Q}_{xk.1} = \frac{D_{xk}}{D_x - D_{x1}}\left[1 - \left(\frac{s_x}{N_x}\right)^{(D_x - D_{x1})/D_x}\right], \qquad k = 2, \cdots, c, \tag{45}$$

and

$$\hat{Q}_{xk.12} = \frac{D_{xk}}{D_x - D_{x1} - D_{x2}}\left[1 - \left(\frac{s_x}{N_x}\right)^{(D_x - D_{x1} - D_{x2})/D_x}\right], \qquad k = 3, \cdots, c. \tag{46}$$

Here again, the sum of the estimators $\hat{Q}_{xk.1}$, for $k = 2, \cdots, c$, is equal to the estimator $\hat{q}_{x.1}$,

$$\begin{aligned}\sum_{k=2}^{c} \hat{Q}_{xk.1} &= \sum_{k=2}^{c} \frac{D_{xk}}{D_x - D_{x1}}\left[1 - \left(\frac{s_x}{N_x}\right)^{(D_x - D_{x1})/D_x}\right] \\ &= 1 - \left(\frac{s_x}{N_x}\right)^{(D_x - D_{x1})/D_x} = \hat{q}_{x.1}.\end{aligned} \tag{47}$$

Although formulas (43) through (46) for the estimator of the net and the partial crude probabilities are quite simple, the exact formulas for their variance and covariance are difficult to derive. Approximate formulas, however, can be developed by the delta method; for example, for the variance of the estimator $\hat{q}_{xk}$, we first write

$$\sigma^2_{\hat{q}_{xk}} = \left\{\frac{\partial}{\partial \hat{Q}_{xk}} \hat{q}_{xk}\right\}^2 \sigma^2_{\hat{Q}_{xk}} + \left\{\frac{\partial}{\partial \hat{p}_x} \hat{q}_{xk}\right\}^2 \sigma^2_{\hat{p}_x} + 2\left\{\frac{\partial}{\partial \hat{Q}_{xk}} \hat{q}_{xk}\right\}\left\{\frac{\partial}{\partial \hat{p}_x} \hat{q}_{xk}\right\} \sigma_{\hat{Q}_{xk}, \hat{p}_x} \tag{48}$$

where the partial derivatives are taken at the true point (Q_{xk}, p_x). Using formulas (35), (36), and (37) for the variance and covariance, we have after simplification

$$\sigma^2_{\hat{q}_{xk}} = \frac{(1 - q_{xk})^2}{N_x p_x q_x}\left[p_x \log(1 - q_{xk}) \log(1 - q_{x.k}) + Q^2_{xk}\right], \qquad k = 1, \cdots, c. \tag{49}$$

Using the same approach, we obtain the covariances

$$\sigma_{\hat{q}_{xh}, \hat{q}_{xk}} = -\frac{(1 - q_{xh})(1 - q_{xk})}{N_x p_x q_x}\left[p_x \log(1 - q_{xh}) \log(1 - q_{xk}) - Q_{xh} Q_{xk}\right], \qquad h \neq k; \quad h, k = 1, \cdots, c; \tag{50}$$

and

$$(51)\qquad \sigma_{\hat{q}_{xk},\hat{p}_x} = -\frac{(1-q_{xk})Q_{xk}}{N_x}, \qquad k = 1, \cdots, c.$$

Approximate formulas for the variances and covariances of the estimators $\hat{q}_{x.k}$ and $\hat{Q}_{xk.1}$ can be obtained in the same way. For the sake of completeness, they are listed below.

$$(52)\qquad \sigma^2_{\hat{q}_{x.k}} = \frac{(1-q_{x.k})^2}{N_x p_x q_x}[p_x \log(1-q_{xk}) \log(1-q_{x.k}) + (q_x - Q_{xk})^2], \qquad k = 1, \cdots, c,$$

$$(53)\qquad \sigma_{\hat{q}_{x.h},\hat{q}_{x.k}} = -\frac{(1-q_{x.h})(1-q_{x.k})}{N_x p_x q_x} [p_x \log(1-q_{xh}) \log(1-q_{xk}) - (q_x - Q_{xh})(q_x - Q_{xk})], \qquad h \neq k;\quad h, k = 1, \cdots, c,$$

$$(54)\qquad \sigma_{\hat{q}_{x.k},\hat{p}_x} = -\frac{(1-q_{x.k})(q_x - Q_{xk})}{N_x}, \qquad k = 1, \cdots, c,$$

$$(55)\qquad \sigma^2_{\hat{Q}_{xk.1}} = \frac{q_x - Q_{x1} - Q_{xk}}{N_x(q_x - Q_{x1})Q_{xk}} Q^2_{xk.1} + \frac{[Q_{xk.1}(q_x - Q_{x1}) - Q_{xk}]^2}{N_x p_x q_x (q_x - Q_{x1})}\left[(q_x - Q_{x1}) + Q_{x1} p_x \left\{\frac{\log p_x}{q_x}\right\}^2\right], \qquad k = 1, \cdots, c,$$

$$(56)\qquad \sigma_{\hat{Q}_{xh.1},\hat{Q}_{xk.1}} = -\frac{Q_{xh.1}Q_{xk.1}}{N_x(q_x - Q_{x1})} + \frac{[Q_{xh.1}(q_x - Q_{x1}) - Q_{xh}][Q_{xk.1}(q_x - Q_{x1}) - Q_{xk}]}{N_x p_x q_x (q_x - Q_{x1})}\left[(q_x - Q_{x1}) + Q_{x1} p_x \left\{\frac{\log p_x}{q_x}\right\}^2\right], \qquad k = 1, \cdots, c,$$

and

$$(57)\qquad \sigma_{\hat{q}_{xk.1},\hat{p}_x} = -\frac{[Q_{xk} - Q_{xk.1}(q_x - Q_{x1})]}{N_x}, \qquad h \neq k;\quad h, k = 2, \cdots, c.$$

Since the sum of the estimators $\hat{Q}_{xk.1}$, for $k = 2, \cdots, c$, is equal to $\hat{q}_{x.1}$ [equation (47)], the following relations should hold for the variances and covariances

$$(58)\qquad \sum_{k=2}^{c} \sigma^2_{\hat{Q}_{xk.1}} + 2 \sum_{h=2}^{c-1} \sum_{k=h+1}^{c} \sigma_{\hat{Q}_{xh.1},\hat{Q}_{xk.1}} = \sigma^2_{\hat{q}_{x.1}}$$

and

$$(59)\qquad \sum_{k=2}^{c} \sigma_{\hat{Q}_{xk.1},\hat{p}_x} = \sigma_{\hat{q}_{x.1},\hat{p}_x}.$$

When (52), (54), (55), (56), and (57) are substituted into (58) and (59), we have two identities which can be proven by direct computation.

4. An application to the follow-up study

The application of the theory of competing risks to a medical follow-up study was illustrated with actual data taken from the files of the Tumor Registry of the California State Department of Public Health [2]. A brief account of the problem is given below.

In a medical follow-up study conducted over a number of years, patients are admitted continuously over the entire study period and are observed until death or until the study is terminated, whichever comes first. The time of admission is taken as the common point of origin for all patients; x is the exact number of years since admission and N_x the number of patients who survive to the end of x years of follow-up. The number of survivors will decrease as x increases, not only because of deaths, but also because of withdrawals consequent to the closing of the study.

Consider a typical interval $(x, x + 1)$. The N_x survivors who begin the interval can be divided into two mutually exclusive groups according to their date of admission into the study. First we have a group of m_x patients who entered the study more than $x + 1$ years before its closing date. Out of these, δ_x will die during the interval and s_x will survive to begin the next interval as N_{x+1}. The δ_x deaths will be further divided into c groups according to the cause of death. The second group of n_x patients entered the study less than $x + 1$ years before its termination, and hence are all counted as withdrawals in the interval $(x, x + 1)$, because for them the closing date of the study precedes their $x + 1$ anniversary date. Let us say that w_x will survive to withdraw alive and ϵ_x will die before the closing date. The ϵ_x deaths will again be divided into c groups by cause.

Clearly, the estimation of the probabilities discussed in the present paper should be based on information regarding both the cause of death and the withdrawal status of the patient. This was done in [2]. Under the assumption of a constant force of mortality for each risk of death and uniform distribution of the n_x withdrawals within the interval $(x, x + 1)$, the estimators of p_x, q_x, and the crude probabilities Q_{xk} turned out to be

$$(60) \qquad \hat{p}_x = \left\{ \frac{-\frac{\epsilon_x}{2} + \left[\left(\frac{\epsilon_x}{2}\right)^2 + 4\left(N_x - \frac{n_x}{2}\right)\left(s_x + \frac{w_x}{2}\right)\right]^{1/2}}{2\left(N_x - \frac{n_x}{2}\right)} \right\}^2, \qquad \hat{q}_x = 1 - \hat{p}_x,$$

and

$$(61) \qquad {}_{xk}\hat{Q} = \frac{D_{xk}}{D_x}\,\hat{q}_x, \qquad k = 1, \cdots, c.$$

Here D_{xk} is the number of deaths from cause R_k in the entire group N_x and the total number of deaths

$$(62) \qquad D_x = D_{x1} + \cdots + D_{xc} = \delta_x + \epsilon_x.$$

Substituting (60) and (61) into (10), (11), (21), and (24) gives the estimators of the net and partial crude probabilities. They are similar to (43) through (46) except that s_x/N_x will be replaced with $\hat{p}_x$ as given in (60).

Approximate formulas for the variances and covariances of these estimators are similar in form to equations (35) through (38) and (49) through (57) in the ordinary case, with the difference that N_x is replaced by M_x where

$$M_x = m_x + \frac{n_x}{1 + p_x^{1/2}}; \tag{63}$$

for example, the approximate formula for the variance of $\hat{p}_x$ is given by

$$\sigma^2_{\hat{p}_x} = \frac{1}{M_x} p_x q_x. \tag{64}$$

REFERENCES

[1] C. L. Chiang, "A stochastic study of the life table and its applications: I. Probability distributions of the biometric functions," *Biometrics*, Vol. 16 (1960), pp. 618–635.
[2] ———, "A stochastic study of the life table and its applications: III. The follow-up study with the consideration of competing risks," *Biometrics*, Vol. 17 (1961), pp. 57–78.
[3] J. Cornfield, "The estimation of the probability of developing a disease in the presence of competing risks," *J. Amer. Public Health Assoc.*, Vol. 47 (1957), pp. 601–607.
[4] E. Fix and J. Neyman, "A simple stochastic model of recovery, relapse, death and loss of patients," *Human Biology*, Vol. 23 (1951), pp. 205–241.
[5] C. W. Jordan, *Life Contingencies*, Chicago, Society of Actuaries, 1952.
[6] M. N. Karn, "A further study of methods of constructing life tables when certain causes of deaths are eliminated," *Biometrika*, Vol. 25 (1933), pp. 91–101.
[7] A. W. Kimball, "Disease incidence estimation in populations subject to multiple causes of death," *ISI Bull.*, Vol. 36 (1958), pp. 193–204.
[8] W. M. Makeham, "On an application of the theory of the composition of decremental forces," *J. Inst. Actuar.*, Vol. 18 (1874), pp. 317–322.
[9] I. Todhunter, *A History of the Mathematical Theory of Probability*, New York, Chelsea, 1949.