Proc. Nat. Acad. Sci. USA
Vol. 72, No. 1, pp. 20–22, January 1975

A Nonidentifiability Aspect of the Problem of Competing Risks

(crude survival probabilities/net survival probabilities)

ANASTASIOS TSIATIS*

Department of Statistics, University of California, Berkeley, Calif. 94720

Communicated by Jerzy Neyman, October 17, 1974

ABSTRACT **For an experimental animal exposed to $k > 1$ possible risks of death $R_1, R_2, \cdots, R_k$, the term i-th potential survival time designates a random variable Y_i supposed to represent the age at death of the animal in hypothetical conditions in which R_i is the only possible risk. The probability that Y_i will exceed a preassigned t is called the i-th net survival probability. The results of a survival experiment are represented by k "crude" survival functions, empirical counterparts of the probabilities $Q_i(t)$ that an animal will survive at least up to the age t and eventually die from R_i. The analysis of a survival experiment aims at estimating the k net survival probabilities using the empirical data on those termed crude. Theorems 1 and 2 establish the relationship between the net and the crude probabilities of survival. In particular, Theorem 2 shows that, without the not directly verifiable assumption that in their joint distribution the variables $Y_1, Y_2, \cdots, Y_k$ are mutually independent, a given set of crude survival probabilities $Q_i(t)$ does not identify the corresponding net probabilities. An example shows that the results of a customary method of analysis, based on the assumption that $Y_1, Y_2, \cdots, Y_k$ are independent, may have no resemblance to reality.**

As recently summarized by David (1), the customary treatment of competing risks is based on the model that we shall term the model of potential survival times. Consider an individual living organism born at time $t = 0$, and assume that through its lifetime it is exposed to $k > 1$ different "risks" or possible causes of death $R_1, R_2, \cdots, R_k$. For $i = 1, 2, \cdots, k$ let Y_i denote a random variable described as the "potential survival time" of the individual in hypothetical conditions in which R_i is the only risk of death, and let $H_i(t) = P\{Y_i > t\}$. The function H_i is described as the i-th net survival probability or the i-th net "decrement" function. The potential survival times Y_i are contrasted with the actual survival time, say X, when the individual in question is exposed to all the $k > 1$ competing risks, so that $X = \min(Y_1, Y_2, \cdots, Y_k)$. The function $Q_i(t)$, described as the i-th crude survival function, is defined as the probability that the individual considered will survive up to age t and then die from cause R_i. Obviously, for $i = 1, 2, \cdots, k$ and $t \geq 0$,

$$Q_i(t) = P\{(Y_i > t) \bigcap_{j \neq i} (Y_j > Y_i)\}. \qquad [1]$$

Ordinarily, the studies of competing risks are based on empirical counterparts of the crude survival functions $Q_i(t)$, perhaps derived from observations of a cohort of experimental animals. The purpose of such studies is to estimate the net survival probabilities and to predict the patterns of mortality to be expected in hypothetical conditions when certain causes of death are either eliminated or modified in their importance. Here the joint distribution of potential survival times Y_i is of great importance. As summarized by David (1), recent studies are based either on assumptions specifying the functional form of this joint distribution with a few adjustable parameters or, predominantly, on the qualitative hypothesis, say A, that the potential survival times $Y_1, Y_2, \cdots, Y_k$ are mutually independent.

The purpose of the present paper is to show that without the hypothesis A, the model of potential survival times is unidentifiable: the set of crude survival functions $Q_i(t)$ is consistent with an infinity of joint distributions of potential survival times. Thus, a fully realistic treatment of the problem of competing risks depends on properly validated detailed hypotheses on the joint distribution of the Y's or, indeed, on a straight stochastic model of competition of risks in the spirit of the following quotation from Chiang (ref. 2, p. 242), "Are people suffering from arteriosclerotic heart disease more likely to die from pneumonia than people without a heart condition?" Naturally, the details of such an approach must be properly validated.

Without the assumption A of independence of the Y's, their joint distribution may be characterized by the function

$$H^{(k)}(t_1, t_2, \cdots, t_k) = P\left\{\bigcap_{i=1}^{k} (Y_i > t_i)\right\}, \qquad [2]$$

to be described as the multiple decrement function. We assume that this function has continuous partial derivatives with respect to all of its arguments. Obviously, the i-th net probability of surviving up to age t is obtained from [2] by substituting $t_i = t$ and $t_j = 0$ for all $j \neq i$. We now establish the relationship between [2] and the i-th crude survival function $Q_i(t)$.

Net and crude survival probabilities

THEOREM 1. *Whatever be the joint distribution of potential survival times, characterized by the multiple decrement function* [2], *the derivative $Q_i'(t)$ of the i-th crude survival function is equal to the partial derivative of* [2] *with respect to t_i evaluated at $t_1 = t_2 = \cdots = t_k = t$.*

Proof. Because the numbering of the k competing risks is arbitrary, it will be sufficient to prove the theorem assuming $i = 1$, which will simplify the notation somewhat. Let t and h^* be arbitrary positive numbers and $0 < h < h^*$. The definition of $Q_i(t)$ implies that the difference

$$Q_1(t) - Q_1(t+h) = P\{(t < Y_1 \leq t + h) \bigcap_{j>1} (Y_j > Y_1)\} \qquad [3]$$

* Present address: Department of Statistics, University of Wisconsin, Madison, Wisc. 53706.

has a lower bound

$$P\{(t < Y_1 \le t + h) \bigcap_{j>1} (Y_j > t + h^*)\} = H^{(k)}(t, t + h^*, \cdots, t + h^*) - H^{(k)}(t + h, t + h^*, \cdots, t + h^*) \quad [4]$$

Similarly, the following formula represents an upper bound of [3]

$$P\{(t < Y_1 \le t + h) \bigcap_{j>1} (Y_j > t)\} = H^{(k)}(t, t, \cdots, t) - H^{(k)}(t + h, t, \cdots, t). \quad [5]$$

By dividing [3], [4], and [5] by h and by passing to the limit as $h \to 0$, one obtains, for all $h^* > 0$,

$$\left.\frac{\partial H^{(k)}}{\partial t_1}\right|_{t_1 = t_2 \cdots = t_k = t} \le Q_1'(t) \le \left.\frac{\partial H^{(k)}}{\partial t_1}\right|_{\substack{t_1 = t \\ t_j = t + h^*}}, \text{ for } j > 1 \quad [6]$$

Another passage to the limit as $h^* \to 0$ yields the desired result

$$Q_1'(t) = \left.\frac{\partial H^{(k)}}{\partial t_1}\right|_{t_1 = t_2 \cdots = t_k = t} = H_1^{(k)}(t) \text{ (say)} \quad [7]$$

Q.E.D.

In particular, if the potential survival times are mutually independent, so that

$$H^{(k)}(t_1, \cdots, t_k) = \prod_{i=1}^{k} H_i(t_i), \quad [8]$$

the derivative $Q_j'(t)$ can be written as

$$Q_j'(t) = -r_j(t) \prod_{i=1}^{k} H_i(t) \quad [9]$$

where, as in David (1),

$$r_j(t) = -\frac{d}{dt} \log H_j(t) \quad [10]$$

designates the "force of mortality," from R_j, and we have

$$H_j(t) = \exp\left\{-\int_0^t r_j(x)dx\right\}. \quad [11]$$

Using the same notation and putting $r(x) = \Sigma_j r_j(x)$, formula [9] can be rewritten as

$$Q_j'(t) = -r_j(t) \exp\left\{-\int_0^t r(x)dx\right\}. \quad [12]$$

Theorem 1 indicates that any given multiple decrement function $H^{(k)}$ determines uniquely the crude survival functions

$$Q_i(t) = -\int_t^\infty H_i^{(k)}(x)dx.$$

Now we place ourselves in the position of not knowing the multiple decrement function and, even, of not knowing whether the potential survival times Y_i are mutually independent. On the other hand, the crude survival functions $Q_i(t)$ can be estimated and, in fact, we shall now assume that they are known precisely. The crucial question is how much information about the joint distribution of potential survival times does the set of crude survival functions provide.

THEOREM 2. *Whatever be the set of crude survival functions $Q_i(t)$, for $i = 1, 2, \cdots, k$, there exists a system of net survival probabilities, say $H_j^*(t)$ for $j = 1, 2 \cdots, k$, which, combined with the assumption A that the potential survival times are independent, implies the crude survival functions $Q_i^*(t)$ that coincide with the given $Q_i(t)$.*

Proof. The proof consists in using the k given functions $Q_j(t)$ to make appropriate substitutions in the left sides of [12] and in solving the resulting equations

$$Q_j'(t) = -r_j^*(t) \exp\left\{-\int_0^t r^*(x)dx\right\} \quad [13]$$

with respect to the $r_j^*(t)$, where $r^*(x) = \Sigma_j r_j^*(x)$. Summing [12] for $j = 1, 2, \cdots, k$ yields

$$\sum_{j=1}^{k} Q_j'(t) = -r^*(t) \exp\left\{-\int_0^t r^*(x)dx\right\} = \frac{d}{dt} \exp\left\{-\int_0^t r^*(x)dx\right\} \quad [14]$$

This implies

$$\sum_{j=1}^{k} Q_j(t) = \exp\left\{-\int_0^t r^*(x)dx\right\}, \quad [15]$$

which, in connection with [13], yields

$$r_j^*(t) = -Q_j'(t) \Big/ \sum_{i=1}^{k} Q_i(t). \quad [16]$$

Finally, formula [11] gives

$$H_j^*(t) = \exp\left\{\int_0^t [Q_j'(x)/\Sigma_i Q_i(x)]dx\right\}. \quad [17]$$

Naturally, the substitution of H_j^* and r_j^* into [9] will yield the derivative of Q_j^* coinciding with that of Q_j. Q.E.D.

Not infrequently, empirical studies of competing risks, conducted on the tentative assumption that the contemplated risks $R_1, R_2, \cdots, R_k$ are independent, end with expressions of satisfaction that the computed crude survival functions Q_i show a reasonable agreement with their empirical counterparts. *Theorem 2*, just proved, indicates that this agreement cannot be considered as any kind of confirmation of the hypothesis of independence of the potential survival times Y_i. Neither is it an encouragement to think that the estimates of the net survival probabilities of the various risks are necessarily realistic.

An illustrative example

The following example has been selected for its simplicity, with no effort to approach any real mechanism of interaction between some two diseases. Assuming $k = 2$, we consider the function

$$H^{(2)}(t_1, t_2) = \exp\{-\lambda t_1 - \mu t_2 - \vartheta t_1 t_2\} \quad [18]$$

as representing the two-fold decrement function, with some positive values of the three parameters λ, μ, and ϑ. In other words, [18] is supposed to characterize the true distribution of potential survival times Y_1 and Y_2 of individuals exposed to the competing risks of death from some $k = 2$ causes R_1 and R_2. The true net probabilities of surviving up to age t in conditions when either R_1 or R_2 are the sole possible causes of death are

$$H_1(t) = \exp\{-\lambda t\} \quad [19]$$

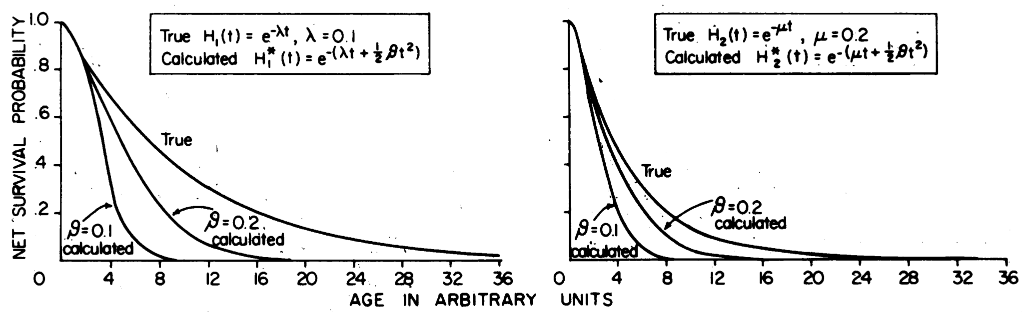

FIG. 1. Comparison of true net survival functions $H_1(t)$ and $H_2(t)$ and their counterparts $H_1^*(t)$ and $H_2^*(t)$ calculable on the unjustified assumption that the potential survival times are mutually independent.

and

$$H_2(t) = \exp\{-\mu t\}, \quad [20]$$

respectively. Formula [7] gives the derivatives of the corresponding crude probabilities of survival and subsequent death from R_1 or R_2 in condition of competition between these two causes:

$$Q_1'(t) = -(\lambda + \vartheta t)\exp\{-\lambda t - \mu t - \vartheta t^2\}$$
$$Q_2'(t) = -(\mu + \vartheta t)\exp\{-\lambda t - \mu t - \vartheta t^2\}. \quad [21]$$

The integration of these formulas yields the true crude survival functions with deaths due to R_1 and R_2, respectively. Then, formula [17] provides what would have been the results of calculations of the net survival probabilities under the exposure to just one of the two causes, of calculations done so to speak routinely, with the presumption that the two causes "act independently." Clearly, the results must depend upon the value of $\vartheta > 0$, whereas the true net survival probabilities of [19] and [20] are independent of this parameter. The two panels of Fig. 1 illustrate the relationship between the true net survival functions $H_1(t)$ and $H_2(t)$ and their counterparts $H_1^*(t)$ and $H_2^*(t)$, which would result from faultless calculations based on the unjustified presumption that the potential survival times Y_1 and Y_2 are independent. The values of the first two parameters chosen are $\lambda = 0.1$, $\mu = 0.2$. These are combined with two alternative values of $\vartheta = 0.1$ and $\vartheta = 0.02$. It is seen that, depending upon the value of ϑ, the net survival probabilities estimated through the unverified and not directly verifiable assumption of independence of potential survival times may have little resemblance to those "true."

Concluding remarks

The foregoing theory and the example seem to indicate that a realistic treatment of the problem of competing risks depends upon an analysis of biological circumstances more delicate than the model of potential survival times can provide.

The contents of the present paper are a part of the author's Ph.D. dissertation, submitted at the Department of Statistics, University of California, Berkeley. I thank Prof. Herbert A. David for informing me of a paper on a similar subject with apparently similar results presented by David M. Rose at the meeting of the American Statistical Association held in August 1974 in St. Louis, Mo. The title of the paper Dr. Rose presented at that meeting is "Dependent competing risks."

This investigation was partially supported by NIH Research Grant no. GM-10525, National Institutes of Health, Public Health Service.

1. David, H. A. (1974) "Parametric approaches to the theory of competing risks," in *Reliability and Biometry, Statistical Analysis of Lifelength,*" eds. Proschan, F. & Serfling, R. J. (SIAM, Philadelphia), pp. 275–290.
2. Chiang, C. L. (1968) *Introduction to Stochastic Processes in Biostatistics* (Wiley, New York).