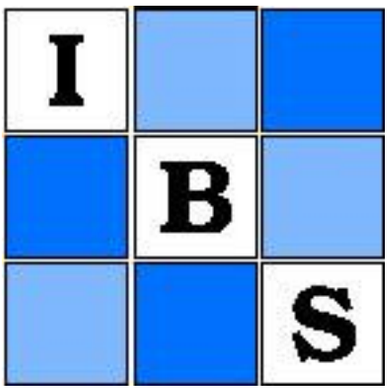

BIOMETRICS 34, 541–554
December, 1978

The Analysis of Failure Times in the Presence of Competing Risks

R. L. PRENTICE

Program in Epidemiology and Biostatistics, The Fred Hutchinson Cancer Research Center, Seattle, Washington 98104, U.S.A.

and

Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.

J. D. KALBFLEISCH

Department of Statistics, University of Waterloo, Waterloo, Ontario, Canada

A. V. PETERSON, JR., N. FLOURNOY, V. T. FAREWELL and N. E. BRESLOW

Program in Epidemiology and Biostatistics, The Fred Hutchinson Cancer Research Center, Seattle, Washington 98104, U.S.A.

and

Department of Biostatistics, University of Washington, Seattle, Washington, 98195, U.S.A.

Summary

Distinct problems in the analysis of failure times with competing causes of failure include the estimation of treatment or exposure effects on specific failure types, the study of interrelations among failure types, and the estimation of failure rates for some causes given the removal of certain other failure types. The usual formulation of these problems is in terms of conceptual or latent failure times for each failure type. This approach is criticized on the basis of unwarranted assumptions, lack of physical interpretation and identifiability problems. An alternative approach utilizing cause-specific hazard functions for observable quantities, including time-dependent covariates, is proposed. Cause-specific hazard functions are shown to be the basic estimable quantities in the competing risks framework. A method, involving the estimation of parameters that relate time-dependent risk indicators for some causes to cause-specific hazard functions for other causes, is proposed for the study of interrelations among failure types. Further, it is argued that the problem of estimation of failure rates under the removal of certain causes is not well posed until a mechanism for cause removal is specified. Following such a specification, one will sometimes be in a position to make sensible extrapolations from available data to situations involving cause removal. A clinical program in bone marrow transplantation for leukemia provides a setting for discussion and illustration of each of these ideas. Failure due to censoring in a survivorship study leads to further discussion.

1. Introduction

The term 'competing risk problem' has come to encompass the study of any failure process in which there is more than one distinct cause or type of failure. Frequent reference is

Key Words: Competing risks; Regression; Time-dependent covariates; Latent failure times; Independence of competing risks; Identifiability; Cause-specific hazard functions; Proportional hazard models; Censoring.

made to possible influences of competing failure types in studies reported in the clinical, epidemiologic, demographic, basic science and industrial literature. Available data on each study subject typically include $T \geq 0$, the time of failure, which may be right censored and $J \epsilon \{1, 2, \ldots, m\}$, the type of failure, which will be unknown if T is censored. A regression vector $\mathbf{z} = (z_1, \ldots, z_p)$ may also be available to record characteristics of the study subject as well as treatment allocations or exposure levels. Some components of $\mathbf{z}$ may be time-dependent, that is $\mathbf{z} = \mathbf{z}(t)$, as occurs when successive measurements are taken on study subjects as they are followed over time. As an example (Hoel 1972), T may be the age of death in a mouse radiation study, J may describe the cause of death as thymic lymphoma, reticulum cell sarcoma, or other, and the regression variable may indicate whether or not each particular mouse was kept in germ free isolation. Similarly, in the University Group Diabetes Project (e.g., Cornfield 1971, Gilbert, Saracci, Meier, Zelen, Rümke and White 1975) T may describe the time from entry into the study to death, J may describe the cause of death, and $\mathbf{z}$ may record the treatment assignment as well as other disease, personal or demographic characteristics of the patient.

Three distinct problems arise in the analysis of failure times with competing risks. These are (a) inference on the effects of treatment, exposure or other regression variables on specific types of failure, (b) the study of interrelations among failure types under specified study conditions and (c) the estimation of failure rates for some causes given the removal of some or all other causes. Problem (a) arose, for example, in the University Group Diabetes Project. The fact that one treatment, Tolbutamide, appeared to give rise to greater cardiovascular mortality was the primary point of interest and controversy in the study. Problems (b) and (c) both involve inference on the relationship between failure types. Problem (b) is concerned with such relationships under actual study conditions, while to address problem (c) it is necessary to extrapolate to an altered situation in which some failure types are no longer operative. Problem (c) is the classical problem of competing risks analysis. A historical example, reported in Kearns (1931), dates back to 1760 and Daniel Bernoulli's interest in estimating mortality rates should smallpox be eradicated.

Bone marrow transplantation in the treatment of acute leukemia patients (e.g., Thomas, Storb, Clift, Fefer, Johnson, Neiman, Lerner, Glucksberg and Buckner 1975a, 1975b) will provide a useful setting in which to discuss the ideas of this paper. For use in subsequent sections, a brief outline of this treatment procedure will be given. Acute leukemia patients, usually endstage, are conditioned by means of ordinarily lethal doses of radiation and chemotherapy in an attempt to eradicate the leukemia and to prepare the patient for a marrow graft. This is followed by the infusion of bone marrow cells from an HLA identical sibling donor. The donor cells repopulate the patient's marrow and hopefully lead to complete hematologic and immunologic reconstitution. Major causes of death for patients undergoing such a procedure are related to (i) the recurrence of leukemia, (ii) graft versus host disease (GVHD) and (iii) interstitial pneumonia. GVHD is believed to occur as an immunologic reaction of the new marrow graft against the patient. It is evident primarily through effects on the skin, liver and gut. Problem (a) arises, for example, in the study of the relationship of prognostic factors, such as type of conditioning regimen, to the recurrence of leukemia. Problem (b) arises in an attempt to relate the risk of recurrent leukemia to the corresponding risk of mortality from GVHD (Section 5). Problem (c) would arise in a discussion of protocol changes intended to eliminate GVHD mortality (Section 6).

A statistical model for competing risks data involves a specification of the distribution for the observable quantities $(T, J, \mathbf{z})$. Usually it will be sufficient to specify a model for (T, J) given $\mathbf{z}$. Such models, in terms of cause-specific hazard functions, will be discussed in Section 2. The usual formulation of competing risk problems, however, (e.g. Cox 1959, Berman 1963, Nelson 1970, Moeschberger and David 1971, and Gail 1975) introduces conceptual or latent

failure times $Y_1, \ldots, Y_m$ corresponding to each type of failure $J = 1, \ldots, m$. The observed T and J are then taken to be $T = \min(Y_1, \ldots, Y_m)$ and $J = \{j \mid Y_j \leq Y_k, k = 1, \ldots, m\}$. This formulation is discussed in Section 3. Sections 4–6 relate respectively to problems (a), (b) and (c) above. Section 7 discusses censoring as a competing cause of failure.

2. *Cause-Specific Hazard Functions and the Likelihood Function*

Suppose failure time T is continuous. The overall failure rate or hazard function for an individual with regression vector $\mathbf{z}$ is given by

$$\lambda(t; \mathbf{z}) = \lim_{\Delta t \to 0} P\{t \leq T < t + \Delta t \mid T \geq t; \mathbf{z}(t)\}/\Delta t$$

where, as above, $\mathbf{z}(t)$ denotes the value of the regression vector at time t (Cox 1972). Cause-specific hazard functions (Chiang 1968, p. 244, 1970, Altshuler 1970, Holt 1978, and Prentice and Breslow 1978) are defined by

$$\lambda_j(t; \mathbf{z}) = \lim_{\Delta t \to 0} P\{t \leq T < t + \Delta t, \quad J = j \mid T \geq t; \mathbf{z}(t)\}/\Delta t$$

for $j = 1, \ldots, m$. The function $\lambda_j(t; \mathbf{z})$ simply gives the instantaneous failure rate from cause j at time t, given the regression vector $\mathbf{z}(t)$, in the presence of the other failure types. Assuming distinct failure types, the overall hazard function can be expressed in terms of cause-specific hazard functions as

$$\lambda(t; \mathbf{z}) = \sum_1^m \lambda_j(t; \mathbf{z}).$$

The likelihood function can be specified in terms of the overall survivor function

$$F(t; \mathbf{z}^*) = \exp\left\{-\int_0^t \lambda(u; \mathbf{z})du\right\},$$

and the probability function for time to failure and cause of failure

$$f_j(t; \mathbf{z}^*) = \lambda_j(t; \mathbf{z})F(t; \mathbf{z}^*),$$

where $\mathbf{z}^* = \mathbf{z}^*(t)$ denotes $\{\mathbf{z}(u); u \leq t\}$.

Suppose now that n study subjects give rise to data $(t_i, j_i, \delta_i, \mathbf{z}_i^*)$, $i = 1, \ldots, n$, where t_i is the failure time, j_i is the cause of failure, δ_i is a censoring indicator, and $\mathbf{z}_i^* = \mathbf{z}_i^*(t_i)$ is a vector-valued regression function for the ith study subject. The censoring indicator takes value one if failure occurs and value zero otherwise. The cause of failure j_i may be specified arbitrarily if $\delta_i = 0$. As usual an independent censoring mechanism will be assumed. This means that at any fixed $\{t, \mathbf{z}(t)\}$ individuals are not selectively censored on the basis of a relatively good or relatively poor prognosis. This condition is met by the usual censoring schemes such as fixed time censoring (Type I), independent random censoring, order statistic censoring (Type II), as well as by more general censoring schemes in which censorship at $\{t; \mathbf{z}(t)\}$ depends arbitrarily on the previous number of failures and censorings. Further discussion of independent censoring mechanisms is given in Kalbfleisch and Mackay (1978a).

The likelihood function under an independent censoring mechanism is, up to proportionality,

$$\prod_{i=1}^n \{[\lambda_{j_i}(t_i; \mathbf{z}_i)]^{\delta_i} F(t_i; \mathbf{z}_i^*)\} = \left(\prod_{i=1}^n [\lambda_{j_i}(t_i; \mathbf{z}_i)]^{\delta_i} \prod_{j=1}^m \exp\left\{-\int_0^{t_i} \lambda_j[u; \mathbf{z}(u)]du\right\}\right). \quad (1)$$

Note that the likelihood function is completely specified by the cause-specific hazard functions $\lambda_j(t, \mathbf{z}), j = 1, \ldots, m$. Note also that upon rearrangement the likelihood factors into a component for each j. In fact, the likelihood factor for $\lambda_j(t; \mathbf{z})$ is precisely the same as would be obtained by regarding all failures from causes other than j as censored at their time of failure. This provides a formal justification, at least for the estimation of $\lambda_j(t; \mathbf{z})$, for the common procedure of regarding failures from other causes as censored when studying factors that effect a certain failure type. Also the likelihood factorization along with standard survival data techniques make it clear that the $\lambda_j(t; \mathbf{z})$ functions are identifiable; that is, the cause-specific hazard functions have the potential to be directly estimated from data of the form $(t, j, \delta, \mathbf{z}^*)$.

The above likelihood development implicitly assumes that the covariate functions $\mathbf{z}^*(t)$ are deterministic or are generated by a stochastic mechanism external to the sample. In such circumstances a survivor function for t given $\mathbf{z}^*(t)$, for example, has clear meaning. More generally, however, it is necessary to consider the likelihood based on the joint distribution $\{T, J\}$ and $\mathbf{z}^*(T)$ which involves additional factors of the type $P\{\mathbf{z}(t) \mid T \geq t; \mathbf{z}(u), u < t\}$. It is still appropriate to use expression (1) for inference on the $\lambda_j(t, \mathbf{z})$ functions, though (1) is properly referred to as a partial rather than an ordinary likelihood (Cox 1975). This matter will be discussed in more detail elsewhere.

3. *Latent Failure Times*

As mentioned above, competing risk problems are most often formulated in terms of latent or potential failure times $Y_1, \cdots, Y_m$ corresponding to the m failure types, along with the assertion that the observed $T = \min(Y_1, \cdots, Y_m)$. A multiple decrement or joint survivor function

$$Q(y_1, \ldots, y_m; \mathbf{z}) = P(Y_1 > y_1, \ldots, Y_m > y_m; \mathbf{z}),$$

is postulated, where $\mathbf{z}$ will be restricted to be time-independent in this section. Problems of the type (a), (b) or (c) are then formulated in terms of Q. Recent literature has concentrated on identifiability aspects of the multiple decrement function. Before discussing these, the more basic question of the physical meaning of the latent failure times will be addressed.

Cox (1959) and Moeschberger and David (1971), among others, define Y_j to be the time of failure from cause j that would be observed if the possibility of failure from causes other than j were removed. They further assume that the observed $T = \min(Y_1, \cdots, Y_m)$. While this point of view ascribes a physical meaning to the latent failure times, it involves the very strong assumption that the time of failure from cause j under one set of study conditions in which all m causes are operative is precisely the same as under an altered set of conditions in which all causes except the *jth* have been removed. Such an assumption may be reasonable in very special situations, such as in an industrial study in which failure types occur in components of a system that are physically and functionally, as well as statistically, independent. More generally, however, the elimination of certain failure types may well alter the risks of other types of failure, as has long been recognized (Makeham 1874, Cornfield 1957). Cox (1959) comments that the latent failure time assumption may not be appropriate in certain examples he discusses, because different failure types do not arise in physically distinct components of the system under study. Evidently any assumption about the relationship between the observed T and times to failure for specific causes, given the removal of other causes will require detailed knowledge of the system under study and of the mechanism for cause removal. This interpretation of latent failure times is not considered further.

A second approach to latent failure time interpretation asserts the existence of times $Y_1, \cdots, Y_m$ on each study subject under the actual study conditions. The random variable Y_j is the observed time of failure if the individual fails of cause j, while no physical meaning is attached to the unobserved Y_k's. This point of view seems implicit in Gail (1975), for example. It does lead to

$$T = \min\{Y_1, \ldots, Y_m\}, \qquad J = \{j \mid Y_j \leq Y_k, \qquad k = 1, \ldots, m\}.$$

Problems concerning interrelation among failure types and of the effects of cause removal are then phrased in terms of the Y_j's and the multiple decrement function Q. For example, independence between causes j and k is identified with statistical independence of Y_j and Y_k. As elaborated below, the distribution of Y_j, given the removal of causes other than j, is usually assumed to be the marginal distribution of Y_j from Q. In view of a lack of physical meaning for unobserved latent failure times, in terms of observable quantities, it is perhaps not surprising that very serious identifiability problems arise when one attempts to use data on (T, J) to estimate Q or derivatives of Q.

Since the likelihood function (Section 2) can be expressed entirely in terms of the cause-specific hazard functions and the cause-specific hazard functions are identifiable it is clear that a function of Q can be estimated, without further assumption, if and only if it can be expressed in terms of the $\lambda_j(t; \mathbf{z})$'s. One can write

$$\begin{aligned}\lambda_j(t; \mathbf{z}) &= \lim_{\Delta t \to 0} P(t \leq Y_j < t + \Delta t \mid T \geq t; \mathbf{z})/\Delta t \\ &= [-\partial \log \quad Q(y_1, \ldots, y_m; \mathbf{z})/\partial y_j]_{y_k = t, \text{ all } k}, \qquad (2)\end{aligned}$$

for $j = 1, \cdots, m$. Consequently only functions of these 'diagonal derivatives' (2) of $\log Q$ will be estimable. For example the marginal distribution of the latent failure time Y_j is generally not identifiable. Let

$$Q_j(y_j; \mathbf{z}) = Q(0, \ldots, 0, y_j, \ldots, 0; \mathbf{z})$$

represent the marginal 'survivor' function for Y_j. Q_j is in 1–1 correspondence with a corresponding 'hazard' function

$$\begin{aligned}h_j(t; \mathbf{z}) &= -\partial \log \quad Q_j(t; \mathbf{z})/dt \\ &= [-\partial \log \quad Q(y_1, \ldots, y_m; \mathbf{z})/\partial y_j]_{y_j = t, y_k = 0, \, k \neq j}.\end{aligned}$$

As these quantities cannot, without additional assumption, be expressed in terms of the cause-specific hazard functions (2) the marginal distributions are non-identifiable. Aside from the inclusion of regression variables, a number of authors (Cox 1959, Berman 1963, Gail 1975, Tsiatis 1975, Peterson 1975, Miller 1977) have noted this non-identifiability problem. Furthermore, Peterson (1976) has shown that data of the type (T, J) will not usually even permit useful bounds to be developed for the Q_j functions.

Additional discussion of the latent failure time approach in relation to problems (a), (b) and (c) will be given in the next three sections.

4. Inference on Cause-specific Regression Coefficients

As noted in Section 2 the likelihood function factors into a separate component for each $\lambda_j(t; \mathbf{z})$. The *jth* likelihood factor is precisely the likelihood that would be obtained if failures of types other than j were regarded as being censored. This implies that virtually all of the

usual survival data methods for a single failure type can be utilized for testing and estimation of $\lambda_j(t; \mathbf{z}), j = 1, \ldots, m$. For example, the proportiona hazards model of Cox (1972, 1975) may be utilized as in Holt (1978) and Prentice and Breslow (1978) to model effects of regression variables on cause-specific hazard functions. This involves a model

$$\lambda_j(t; \mathbf{z}) = \lambda_{0j}(t)\exp(\mathbf{z}\beta_j), \qquad j = 1, \ldots, m \tag{3}$$

where $\lambda_{0j}(\ ) \geq 0$ is arbitrary and β_j, $j = 1, \ldots, m$ are column vectors of cause-specific regression coefficients to be estimated from the data. This model leads to a partial likelihood (Holt 1978) for $\beta_1, \ldots, \beta_m$ that can be written

$$\prod_{j=1}^{m}\left[\prod_{i=1}^{d_j}\exp[\mathbf{z}_{(j)i}\beta_j]\Big/\sum_{l\in R(t_{j(i)})}\exp(\mathbf{z}_l\beta_j)\right], \tag{4}$$

where $t_{j(i)}$, $i = 1, \ldots, d_j$ denotes the d_j times of failure of type j, $\mathbf{z}_{j(i)}$ denotes the corresponding regressor variable, $R(t_{j(i)})$ is the set of study subjects known to be at risk just prior to $t_{j(i)}$ and $\mathbf{z}_{j(i)}$, $\mathbf{z}_l$ are evaluated at $t_{j(i)}$. Standard asymptotic likelihood methods can be applied to (4) for estimation of the β's. Note that no assumption is required concerning the interrelation among the causes of failure. Thus inference on the effects of treatment or exposure variables incorporated in $\mathbf{z}$ on specific types of failure can be made without introducing strong modelling assumptions. The interpretation of such effects is, however, restricted to actual study conditions and there is no implication that the same regression estimates would prevail under a new set of conditions in which, for example, certain causes of failure have been eliminated. Of course in circumstances in which different failure types arise in physically distinct components of a system, a stronger interpretation is possible in that $\lambda_j(t; \mathbf{z})$ is then precisely the hazard function for cause j given that other causes are inoperative. Note that specific β_j parameters can be estimated using the *jth* component of (4), without restricting the cause-specific hazard functions for other failure types to be of the proportional hazards form (3).

The $\lambda_{0j}(\cdot)$ functions may be estimated by the methods of Cox (1972), Kalbfleisch and Prentice (1973) or Breslow (1974). In particular, if there are no regressor variables present the Kaplan-Meier estimator may be utilized toward the estimation of $\lambda_j(t)$ (see Aalen 1976.)

Alternatively, the cause-specific hazard functions with time-independent covariates may be modelled using an accelerated failure time model

$$\lambda_j(t; \mathbf{z}) = \lambda_{0j}\{t\exp(\mathbf{z}\beta_j)\}\exp(\mathbf{z}\beta_j), \qquad j = 1, \ldots, m.$$

Again because of the factorization of the likelihood, inference on a particular β_j can proceed by any of the single failure type procedures used to analyze models linear in the logarithm of failure time. Possibilities include parametric approaches based on exponential, Weibull or log-normal distributions (e.g., Farewell and Prentice 1977) or rank procedures based on generalized Wilcoxon (Gehan 1965, Breslow 1970) or log-rank statistics (Mantel 1966, Peto and Peto 1972) or other generalized rank tests (Prentice 1978).

A specialization of (3) (Holt 1978) that, when appropriate, would be expected to improve the efficiency of β_j estimation, is given by

$$\lambda_j(t; \mathbf{z}) = \lambda_0(t)e^{\gamma_j}\exp(\mathbf{z}\beta_j), \qquad j = 1, \ldots, m. \tag{5}$$

In (5) the hazard functions are restricted to be proportional to each other with proportionality factor e^{γ_j} with, for uniqueness, $\gamma_1 = 0$. A partial likelihood for the β's and γ's can be written

$$\prod_{j=1}^{m} \prod_{i=1}^{d_j} \left\{ \exp(\gamma_j + \mathbf{z}_{(j)i}\beta_j) \Big/ \left[\sum_{k=1}^{m} \sum_{l \in R(t_{j(i)})} \exp(\gamma_k + \mathbf{z}_l\beta_k) \right] \right\}.$$

The proportional risk model (5) has some convenient properties, even though it would often be unacceptably restrictive. For example, the probability that a study subject with time-independent regression variable $\mathbf{z}$ fails of cause j follows a logistic model

$$P(J = j; \mathbf{z}) = \exp(\gamma_j + \mathbf{z}\beta_j) \Big/ \sum_{k=1}^{m} \exp(\gamma_k + \mathbf{z}\beta_k), \qquad j = 1, \ldots, m$$

regardless of $\lambda_0(\cdot)$. It follows also that T and J are statistically independent.

The latent failure time approach to the estimation of regression coefficients would involve specification of a regression model for $Q(y_1, \ldots, y_m; \mathbf{z})$. In that only functions of (2) enter the likelihood, it seems more direct and less restrictive to model the cause-specific hazard functions directly. Usually, for tractability, the latent failure times in such a regression model (e.g., David and Moeschberger 1978) would be assumed independent. Under this assumption the $\lambda_j(t; \mathbf{z})$ and $h_j(t; \mathbf{z})$ functions are easily seen to be identical so that models for the cause-specific hazard functions are in one-to-one correspondence with models for the multiple decrement function Q. It therefore seems important to concentrate on the $\lambda_j(t; \mathbf{z})$ functions for statistical modelling as they lead to procedures that have a clear interpretation regardless of the interrelation between causes of failure and yet are identical with the more traditional results, based on independent latent failure times, in circumstances in which an independence assumption is justifiable.

5. *The Study of Interrelations among Failure Types*

5.1 *Latent Failure Time Approach*

As mentioned above, in the latent failure time approach interrelations among failure types are identified with interrelations among the latent failure times. For example, independence between failure types j and k is taken to mean statistical independence between Y_j and Y_k. This point of view is questionable unless the Y_j's have a clear physical meaning. Furthermore, because of the identifiability problems mentioned above both the marginal and joint survivor functions of Y_j and Y_k are nonidentifiable, even with time-independent regression variables available. It follows that the hypothesis of independence is wholly untestable without introducing further assumptions. For this reason parametric restrictions are sometimes placed on the multiple decrement function Q in order to lead to a test for independence or to study interrelations more generally (e.g., Marshall and Olkin 1967, Moeschberger and David 1971, Nádas 1971, David 1974, Moeschberger 1974). Within such a parametric model it may well be possible to estimate parameters that describe possible dependencies among latent failure times. The difficulty with this approach is that data of the type $(T, J; \mathbf{z})$ do not allow one to distinguish between the assumed model and a model with independent risks but the same cause-specific hazard functions. As a simple illustration suppose one postulates a model

$$Q(y_1, y_2) = \exp[1 - \alpha_1 y_1 - \alpha_2 y_2 - \exp\{\alpha_{12}(\alpha_1 y_1 + \alpha_2 y_2)\}] \tag{6}$$

where $\alpha_1, \alpha_2 > 0$ and $\alpha_{12} > -1$. The parameter α_{12} measures the dependence between Y_1 and Y_2 assuming (6). The corresponding cause-specific hazard functions are

$$\lambda_j(t) = \alpha_j[1 + \alpha_{12} \exp\{\alpha_{12}(\alpha_1 + \alpha_2)t\}], \qquad j = 1, 2.$$

The likelihood function can be written in terms of the $\lambda_j(t)$'s and it is clear that all three parameters are estimable. The likelihood obtained, however, is identical to that arising from an independent risks model with the same $\lambda_j(t)$'s and multiple decrement function

$$Q^*(y_1, y_2) = \exp[1 - \alpha_1 y_1 - \alpha_2 y_2 - \sum_{j=1}^{2} \alpha_j \exp\{\alpha_{12}(\alpha_1 + \alpha_2)y_j\}(\alpha_1 + \alpha_2)^{-1}].$$

It follows that an estimated value of $\alpha_{12} \neq 0$ should not be taken as an indication of dependence between Y_1 and Y_2 unless there is external evidence to support (6). The estimate of the degree of dependence between Y_1 and Y_2 arises from a model assumption that cannot be tested by the data.

5.2 Time-Dependent Risk Indicators

Since there are fundamental problems in studying interrelations among failure times and since one can argue that such interrelations should not, in any case, be identified with interrelations among corresponding failure types, it is natural to attempt to formulate such problems in terms of the observable quantities $(T, J, \mathbf{z})$, rather than the latent failure times.

Failure types j_1 and j_2 will be said to be related if study subjects at high risk for a failure of type j_1, say, are systematically at high, or low, risk for a failure of type j_2. Clearly data in addition to (T, J) are required to examine such associations. One interesting possibility in this regard involves the use of multiple pathologic entities or multiple equipment faults at failure (Breslow, Day, Tomatis and Turusov 1974, Wong 1977). Suppose that j_1 refers to the primary cause of death listed on a death certificate for a human population; for example, j_1 may indicate death due to lung cancer. The frequency with which, say, cardiovascular disease (j_2) is listed as a 'contributing' cause, relative to the frequency of such a listing with other primary causes, may provide some information on the relationship between j_1 and j_2. This approach is clearly worth pursuing though there are some substantial difficulties to be overcome. For example, the pathologic data may primarily reflect developments that take place very close to death, brought on by the presence of advanced disease or by the treatment of the primary disease. For example, at face value, persons at high risk for Hodgkin's disease would also appear to be a high risk for bone marrow deficiency because of the conventional chemotherapeutic approach to the treatment of the disease. Also, death certificate data are likely to be subject to spurious associations between a 'primary' cause and certain contributory causes that are most likely to be discovered in diagnostic procedures related to the primary disease.

A second and promising approach to the use of observable quantities $(T, J, \mathbf{z})$ to study the relationship among failure types, involves the definition of risk-indicator variables for some failure types which can be related to cause-specific hazard functions for other failure types, as time-dependent covariates. Ideally a risk-indicator function for cause j would give an individual's propensity to fail from cause j at time t. Suppose a positive relationship is detected between such a time-dependent risk variable for cause j and the cause-specific hazard function for cause k. This would indicate that individuals at high risk for failure of type j (relative to other individuals at the same follow-up time and with similar characteristics) are simultaneously at high risk for failure of type k.

For a progressive fatal disease a risk-indicator function would be expected to be defined as a monotone function of time which reaches a limiting value at death from that disease. In chronic diseases a major component of a risk-indicator definition would involve the presence of early disease itself. In other situations only less direct 'risk factor' data may be available.

The idea of using time-dependent risk-indicator variables is similar to the illness-death

process of Neyman (1950), Fix and Neyman (1951) and Chiang (1968, p. 73). Risk-indicator variables would be included in the regression vector $\mathbf{z}(t)$. The proportional hazards model (3) and partial likelihood (4) provide a very flexible method for estimating the relationship between such time-dependent risk variables and cause-specific hazard functions. Supplementary analyses may examine the association between the risk indicators themselves.

5.3 Example

Consider now an illustration from the Seattle marrow transplantation program mentioned in the introduction. We are indebted to Dr. E. D. Thomas for permission to discuss these data and to list some results from a recent analysis. These results are, however, intended for illustration only and a more comprehensive analysis will appear elsewhere. A question of significant biologic implication concerns the relationship between GVHD and leukemia relapse. For example, it may be that a graft versus host (GVH) reaction of a certain degree of severity is useful in the eradication of residual or new leukemia cells. Alternatively, a severe GVH reaction may simply be destructive to the patient's organs and be associated with immunologic abnormalities.

Regular measurements are taken over the patient's post-transplant course that can be used to define a GVHD risk-indicator to be related as a time-dependent regression variable to leukemia relapse. For example the date of onset of GVHD is recorded and a GVHD grade is assigned on a scale of zero to four. Other details of the effect of the GVHD on specific organs are also noted as is a designation of acute or chronic GVHD. Here, for the purpose of illustration, a GVHD risk-indicator variable is defined simply as a variable, $z(t)$, that takes value zero between the time of transplant and the diagnosis of GVHD and value one thereafter. This variable is then related to the cause-specific hazard function for leukemia relapse using the proportional hazards model (3). A partial likelihood (4) and Newton-Raphson iteration applied to data on 135 Seattle marrow transplant recipients gives a maximum partial likelihood estimate for the corresponding regression coefficient of $\hat{\beta} = -.792$. The estimated standard error of $\hat{\beta}$ from the 'observed' information matrix is .320, giving a standard normal value of -2.47 which is significant at the .02 level. This suggests that the leukemia relapse rate is reduced by an estimated multiplicative factor $\exp(\hat{\beta}) = .45$ upon the onset of GVHD. Many refinements in the analysis are, of course, necessary before such association can be claimed. For example the 135 patients included 31 syngeneic (identical twin) transplants, which do not give rise to GVHD, as well as allogeneic (HLA matched sibling) transplants. The patient group was quite heterogeneous in terms of conditioning regimen, age and leukemia risk factors. The proportional hazards framework is convenient for adjusting for possible confounding effects of such other variables. As a simple illustration, Table 1 extends the analysis given above to include an indicator variable for type of transplant (0-syngeneic, 1-allogeneic) as well as a variable giving patient age. The estimated effect of GVHD on leukemic relapse is virtually unchanged by these inclusions and is still significant at the .05 level.

6. Failure Rate Estimation Following Cause Removal

6.1 General

The estimation of failure probabilities given the removal of some or all other causes has been a central and long standing problem in competing risk methodology. Such quantities

TABLE 1
Proportional Hazards Analysis of Leukemia Marrow Transplant Data

Factor (z)	Coefficient (β)	Standard Error	Normal Deviate
GVHD Risk Indicator	−.764	.370	−2.06
Syngeneic vs. Allogeneic	.054	.340	0.16
Age in Years/10	.127	.098	1.29

are often referred to as net (all other causes removed) or partial crude (some but not all other causes removed) probabilities. Unlike the problems of Sections 4 and 5 this estimation involves extrapolation (Cornfield 1957) from one set of study conditions in which m causes of failure are operative to another set of conditions in which only a subset are active. The principal point to be made is that this problem is not, in general, well defined until the mechanism for cause removal is clearly specified. Such a specification in a sense prescribes the 'direction' or 'axis' for the extrapolation. Further, in order to make a valid extrapolation, it will usually be necessary to be in possession of detailed knowledge of the biological or physical mechanism giving rise to failures. Before illustrating these points some frequently used approaches to net and partial crude probability estimation will be reviewed.

6.2 Statistical Definitions of Cause Removal

Chiang (1968, p. 246) asserts that probability statements for cause j given that it is the only cause of failure that is operative should be based on the cause-specific hazard function $\lambda_j(t; \mathbf{z})$ (Chiang considered only homogeneous populations). This actually involves a very strong additional assumption that the instantaneous failure rate for cause j under actual study conditions, with all m causes acting, is identical to that under new conditions with only cause j possible. Similar strong assumptions attend Chiang's procedures for estimation of partial crude probabilities, which again simply involve setting equal to zero the cause-specific hazard functions for all 'removed' causes.

The more prevalent latent failure time approach to cause removal (reviewed in Gail 1975) also involves strong additional assumptions. The usual assumption is that a realization of the latent failure times $(y_1, \ldots, y_m)$ is unchanged by cause removal and that the observed time is no longer $t = \min(y_1, \ldots, y_m)$ but is rather the minimum of latent times for causes that have not been removed. In effect, the stochastic mechanism generating failures is assumed to continue beyond latent failure times for causes that have been removed until the smallest operative failure time is reached. This leads to the marginal survivor function for the remaining causes as the basis for failure probability calculations. Unfortunately, as noted in Section 3 for the Q_j functions, these quantities are not estimable from data of the type $(T, J, \mathbf{z})$, with $\mathbf{z}$ time-independent. This point of view then not only involves strong additional assumptions but it does not lead to useful inference techniques. It is perhaps surprising that this approach has received so much attention in the literature.

Other 'statistical' definitions of the meaning of cause removal would be possible within the multiple decrement framework. For example, the effect of removal of causes other than j could be assumed to leave Y_j unchanged but to condition probability statements on the region in which $Y_k > y_k^0$, $k \neq j$ for suitable chosen large values y_k^0. This would lead to other non-identifiable derivatives of the multiple decrement function.

6.3 Illustration

As suggested above, except in circumstances of complete biologic or physical independence among system components giving rise to the various failure types, it is unrealistic to suppose that general statistical methods can be put forward that will encompass all possible mechanisms for cause removal. Consider again the marrow transplantation setting. One can envisage two distinct mechanisms for the removal of GVHD as a cause of death. The two would be expected to have rather different effects on failure rates for the remaining causes of death such as leukemia relapse.

GVHD is presumed to arise through minor genetic differences between the marrow donor and the patient. One possible mechanism to remove GVHD as a cause of death would relate to a strengthening of the donor-recipient matching criteria for entry into the treatment program. Such a change would be expected to give rise to a substantially altered immunologic response of the marrow graft to residual or new leukemia cells. In fact, one could view patients (approximately half) that do not experience clinically detectable GVHD as having fortuitously experienced greater genetic similarity than that required for entry into the program. This line of thought, together with the analysis of Table 1, suggests that recurrent leukemia relapse rates would increase upon removal of GVHD by this mechanism. This illustration points out that, upon specification of a mechanism for cause removal, the data at hand may well be useful for extrapolating failure rates under cause removal. In the current setting the identical twin (syngeneic) transplants, which involve a total genetic match between donor and recipient, provide rather direct information on recurrent leukemia mortality if matching criteria were altered to the extent that absolutely no GVHD arose. Note also that the cause removal assumptions of both Chiang's approach and the latent failure time approach conflict with the graft versus leukemia effect that is suggested by Table 1.

A second conceivable mechanism for the removal of GVHD as a cause of death would involve treatment by an agent that is able to control the severity of the GVH reaction to the extent that patients survive through the acute phase of GVHD (one agent, antithymocyte globulin, appears to be somewhat effective in this respect). Such a mechanism would in no way alter the donor recipient matching criteria, the conditioning treatment prior to marrow transplantation or the marrow grafting procedure itself, As such, any graft versus leukemia effect (Table 1) would presumably remain. Also if the agent had antileukemic potential itself one might expect a reduction in the recurrent leukemia rates. On the other hand, an agent with severe immunosuppressive potential may simply lead to a more conducive environment for leukemia relapse and thereby higher recurrent leukemia rates. Once again knowledge of the mechanism for cause removal and of the interaction of this mechanism with the biological system giving rise to the recurrences is required for sensible extrapolations to be made.

7. Censoring as a Cause of Failure

Throughout the previous sections withdrawal or removal from a study was assumed to arise from an 'independent censoring mechanism'. This condition includes independent random censorship as well as many other censoring mechanisms. Suppose now that attention is restricted to random censorship but that the independence assumption is relaxed. This probability structure allows censoring to be included as one of the failure types. For simplicity suppose there are only two failure types, death and censoring. As a competing risk problem this formulation is unique in that there is an underlying death mechanism that is unaffected by the presence of the censoring. The system has, in this respect, more structure than has been assumed above. This structure gives physical meaning to a latent failure time

for death and the marginal distribution that arises from the elimination of censoring is clearly the relevant target of estimation. In fact, it is suspected that attempts to model censorship as a competing risk have provided a primary impetus for the emphasis on latent failure times. As in Section 3, however, this marginal distribution is non-identifiable without additional assumptions, such as statistical independence of death and censorship. Williams and Lagakos (1978) and Kalbfleisch and Mackay (1978b) have provided a discussion of certain death and censoring systems, which permit the marginal distribution for death to be estimated.

It is of interest to examine whether the time-dependent risk-indicator approach of Section 5 can provide evidence for or against independence of censoring and death. This approach would require the determination of the relationship between a time-dependent measure of the risk of dying with the instantaneous censoring rate. One possibility in the clinical trial setting would utilize performance status measurements recorded over the patient's disease course. Such measures (e.g., Karnofsky scale) range from an upper level, at no clinical evidence of disease, to intermediate levels depending on the patient's ability for self-care and need for hospitalization, to a lower boundary at death. One could define a risk indicator $\mathbf{z}(t)$ for death as the difference between an individual's performance status and that of the average performance status for study subjects at risk at the same follow-up time, t. A test for a zero corresponding coefficient, in the presence of other regression effects, would then examine whether individuals are being selectively censored when they have a relatively poor, or relatively good, prognosis in comparison to other study subjects with similar characteristics at the same follow-up time.

Acknowledgment

This work was supported by Grant Nos. CA-15704, CA-21193 and GM-24472 and contract N01-ES-8-2125 awarded by the National Institutes of Health, DHEW.

Résumé

Les différents problèmes posés par l'analyse du nombre d'insuccés quand plusieurs causes possibles interfèrent font intervenir l'estimation des effets du traitement sur des types d'insuccés spécifiques, l'étude des interelations entre types d'insuccés, et l'estimation des taux d'insuccés pour certaines causes, certains types d'insuccés n'étant pas considérés. La formulation usuelle de ces problèmes, se fait pour chaque type d'insuccés en terme de nombre d'insuccés réel ou potentiel. Cette approche est critiquée, absence d'interprétation physique, problèmes d'indentification. On propose une autre approche utilisant pour chaque cause spécifique des fonctions aléatoires des quantités observables contenant des covariables dépendant du temps. On montre que ces fonctions aléatoires sont les quantités de base estimables. Pour étudier les interelations entre types d'insuccés, on propose une méthode qui permet d'estimer les paramètres qui relient les indicateurs de risque dépendant du temps pour certaines causes aux fonctions aléatoires relatives à d'autres causes. En outre, on montre que le problème de l'estimation du taux d'insuccés, un certain nombre de causes étant supprimées, n'est pas bien posé tant le mècanisme de la suppression des causes n'est pas spécifié. Cette précision étant donnée, on sera parfois en mesure de faire des extrapolations à partir des données disponibles à d'autres situations tenant compte des suppressions. Un programme clinique de transplantation de moelle osseuse dans le cas de la leucémie fournit une base de discussion et une illustration de chacune de ces idées. L'insuccés due à l'arrêt (du traitement) dans une étude de survie permet une discussion ultérieure.

References

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scandinavian Journal of Statistics 3,* 15–27.

Altshuler, B. (1970). Theory for the measurement of competing risks in animal experiments. *Mathematical Biosciences 6,* 1–11.

Berman, S. M. (1963). Notes on extreme values, competing risks and semi-Markov processes. *Annals of Mathematical Statistics 34,* 1104–1106.
Breslow, N. E. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika 57,* 579–94.
Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics 30,* 89–99.
Breslow, N. E., Day, N. E., Tomatis, L. and Turusov, V. S. (1974). Associations between tumor types in a large-scale carcinogenesis study of CF-1 mice. *Journal of the National Cancer Institute 52,* 233–239.
Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics.* Wiley, New York.
Chiang, C. L. (1970). Competing risks and conditional probabilities. *Biometrics 26,* 767–776.
Cornfield, J. (1957). The estimation of the probability of developing a disease in the presence of competing risks. *American Journal of Public Health 47,* 601–607.
Cornfield, J. (1971). The university group diabetes program. A further statistical analysis of the mortality findings. *Journal of the American Medical Association 217,* 1676–1687.
Cox, D. R. (1959). The analysis of exponentially distributed lifetimes with two types of failure. *Journal of the Royal Statistical Society, Series B 21,* 411–421.
Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B 34,* 187–220.
Cox, D. R. (1975). Partial likelihood. *Biometrika 62,* 269–276.
David, H. A. (1974). Parametric approaches to the theory of competing risks. In *Reliability and Biometry,* Statistical Analysis of Lifelength. F. Prochan and R. J. Serfling (eds.) 275–290. SIAM, Philadelphia.
David, H. A. and Moeschberger, M. L. (1978). *The Theory of Competing Risks.* Griffin, London.
Farewell, V. T. and Prentice, R. L. (1977). A study of distributional shape in life testing. *Technometrics 19,* 69–75.
Fix, E. and Neyman, J. (1951). A simple stochastic model to recovery, relapse, death and loss of patients. *Human Biology 23,* 204–241.
Gail, M. (1975). A review and critique of some models used in competing risk analysis. *Biometrics 31,* 209–222.
Gehan, E. (1965). A generalized Wilcoxon test for comparing arbitrarily singly censored samples. *Biometrika 52,* 203–223.
Gilbert, J. P., Saracci, R., Meier, P., Zelen, M., Rümke, C. L. and White, C. (1975). Report of the committee for the assessment of biometric aspects of controlled trials of hypoglycemic agents. *Journal of the American Medical Association 231,* 583–608.
Hoel, D. G. (1972). A representation of mortality data by competing risks. *Biometrics 28,* 475–488.
Holt, J. D. (1978). Competing risk analyses with special reference to matched pair experiments. *Biometrika 65,* 159–166.
Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika 60,* 267–278.
Kalbfleisch, J. D. and Mackay, R. J. (1978a). Censoring and the immutable likelihood. Technical Report No. 5, Dept. of Statistics, University of Waterloo.
Kalbfleisch, J. D. and Mackay, R. J. (1978b). A note on constant sum models. To appear, *Biometrika.*
Kearns, M. H. (1931). An inquiry into various death rates and the comparative influence of certain diseases on the duration of life. *Annals of Human Eugenics 4,* 279–326.
Makeham, W. M. (1874). On an application of the theory of the composition of decremental forces. *Journal of the Institute of Actuaries 18,* 317–322.
Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports 50,* 163–170.
Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association 62,* 30–44.
Miller, D. R. (1977). A note on independence of multivariate lifetimes in competing risk models. *Annals of Statistics 5,* 576–579.
Moeschberger, M. L. (1974). Life tests under dependent competing causes of failure. *Technometrics 16,* 39–47.
Moeschberger, M. L. and David, H. A. (1971). Life tests under competing causes of failure and the theory of competing risks. *Biometrics 27,* 909–923.
Nádas, A. (1971). The distribution of the identified minimum of a normal pair determines the distribution of the pair. *Technometrics 13,* 201–202.
Nelson, W. (1970). Hazard plotting methods for analysis of life data with different failure modes.

Journal of Qualitative Technology 2, 126–149.
Neyman, J. (1950). *First Course in Probability and Statistics.* Holt, New York.
Peterson, A. V. (1975). Nonparametric estimation in the competing risk problem. Technical Report 73, Department of Statistics, Stanford University.
Peterson, A. V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. *Proceedings of the National Academy of Science 73,* 11–13.
Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). *Journal of the Royal Statistical Society, Series A 135,* 185–206.
Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika 65,* 167–180.
Prentice, R. L. and Breslow, N. E. (1978). Retrospective studies and failure time models. *Biometrika 65,* 153–158.
Thomas, E. D., Storb, R., Clift, R. A., Fefer, A., Johnson, F. L., Neiman, P. E., Lerner, K. G., Glucksberg, H. and Buckner, C. D. (1975a). Bone marrow transplantation (first of two parts). *New England Journal of Medicine 292,* 832–843.
Thomas, E. D., Storb, R., Clift, R. A., Fefer, A., Johnson, F. L., Neiman, P. E., Lerner, K. G., Glucksberg, H. and Buckner, C. D. (1975b). Bone marrow transplantation (second of two parts). *New England Journal of Medicine 292,* 895–902.
Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Science 72,* 20–22.
Williams, J. S. and Lagakos, S. W. (1977). Models for censored survival analysis: constant sum and variable-sum models. *Biometrika 64,* 215–224.
Wong, O. (1977). A competing-risk model based on the life table procedure in epidemiologic studies. *International Journal of Epidemiology 6,* 153–159.

Received February 1978; *Revised July* 1978.