

An Analytic Method for Randomized Trials with Informative Censoring: Part 1

JAMES M. ROBINS

Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115

Keywords: dependent censoring; informative censoring; causal inference; counterfactual causality; repeated measures.

Received October 20, 1994; accepted June 13, 1995

Abstract. Consider a randomized trial in which time to the occurrence of a particular disease, say pneumocystis pneumonia in an AIDS trial or breast cancer in a mammographic screening trial, is the failure time of primary interest. Suppose that time to disease is subject to informative censoring by the minimum of time to death, loss to and end of follow-up. In such a trial, the censoring time is observed for all study subjects, including failures. In the presence of informative censoring, it is not possible to consistently estimate the effect of treatment on time to disease without imposing additional non-identifiable assumptions. The goals of this paper are to specify two non-identifiable assumptions that allow one to test for and estimate an effect of treatment on time to disease in the presence of informative censoring. In a companion paper (Robins, 1995), we provide consistent and reasonably efficient semiparametric estimators for the treatment effect under these assumptions. In this paper we largely restrict attention to testing. We propose tests that, like standard weighted-log-rank tests, are asymptotically distribution-free α -level tests under the null hypothesis of no causal effect of treatment on time to disease whenever the censoring and failure distributions are conditionally independent given treatment arm. However, our tests remain asymptotically distribution-free α -level tests in the presence of informative censoring provided either of our assumptions are true. In contrast, a weighted log-rank test will be an α -level test in the presence of informative censoring only if (1) one of our two non-identifiable assumptions hold, and (2) the distribution of time to censoring is the same in the two treatment arms. We also extend our methods to studies of the effect of a treatment on the evolution over time of the mean of a repeated measures outcome, such as CD-4 count.

1. Introduction

Consider a two-armed randomized trial in which time to the occurrence of a particular disease, say pneumocystis pneumonia in an AIDS trial or breast cancer in a mammographic screening trial, is the outcome of primary interest. Suppose that time to disease is subject to informative censoring by the minimum of time to death, loss to and end of follow-up. Our goal is both to test the null hypothesis of no effect of treatment on time to disease, and to estimate the magnitude of the treatment effect. Robins (1987a, Theorem AD1 and Sections 5 and 6) showed that the treatment-arm specific marginal distribution of time to disease occurrence is identified if data are available on a time-varying marker-process and if, the cause-specific hazard of censoring does not depend on time to disease conditional on the history of the marker process. Treatment arm specific marginal survival curves for time to disease can then be compared using the inverse probability of censoring weighted log-rank tests described in Robins and Rotnitzky (1992) and Robins (1993) or by using the G-computation algorithm described in Robins (1986, 1987b, 1989, p. 129).

When, as will be assumed in this paper, data on such a marker process are not available,

Lin, Robins, and Wei (1995) generalize an approach of Robins (1989, p. 158), Robins and Tsiatis (1991), and Robins and Rotnitzky (1992, Appendix 4). In Lin et al.'s approach, an asymptotically normal and unbiased estimator of the treatment effect on time to disease was constructed under the (non-identifiable) assumption that, on a logarithmic scale, time to disease and time to informative censoring satisfied a bivariate location shift model with a completely arbitrary baseline distribution. Their key idea was to use the fact that the censoring time is known, even for subjects observed to develop disease.

The goals of this paper are to specify non-identifiable assumptions that are weaker than those made by Lin et al. and yet allow one to test for and estimate an effect of treatment on time to disease in the presence of informative censoring. In a companion paper, we provide consistent and reasonably efficient semiparametric estimators for the treatment effect under these assumptions. In this paper, we largely restrict attention to the testing problem. Since, in contrast to Lin et al. (1995), our goal is not the joint estimation of the effect of treatment on time to informative censoring (e.g., death) and disease, we choose to estimate time to informative censoring (completely) non-parametrically. Hence, our inference concerning the effect of treatment on time to disease is not invalidated by possible misspecification of a model (such as Lin et al.'s location shift model) for the effect of treatment on time to informative censoring. In this paper, we regard all components of censoring as possibly informative. In our companion paper, we consider the additional assumption that some components of the censoring mechanism (e.g., time to loss to follow-up) are *a priori* independent of both time to disease and the other informative components of censoring (e.g., time to death).

The paper is organized as follows. In Section 2, we describe two non-identifiable assumptions (Eqs. 2.8 and 2.9) sufficient to identify effects of treatment on disease in the presence of informative censoring. In Section 3, we propose a class of tests for a treatment effect under these assumptions. One non-identifiable identifying assumption is that the effect of treatment on time to censoring is rank preserving (i.e., if subject i would be censored before subject j when both are untreated, then subject i would be also censored before subject j when both are treated). The second identifying assumption is a kind of "non-interaction" assumption. The second assumption is always satisfied when time to censoring and time to failure are conditionally independent given treatment arm. It follows that the tests in our class, like standard weighted-log-rank tests, are asymptotically distribution-free α -level tests under the null hypothesis of no causal effect of treatment on time to disease when censoring and failure are conditionally independent. However, our tests remain asymptotically distribution-free α -level tests in the presence of informative censoring provided either the effect of treatment on censoring is rank-preserving or our "non-interaction" assumption holds. In contrast, a weighted log-rank test will be an α -level test in the presence of informative censoring only if (1) one of our two non-identifiable assumptions hold, and (2) the distribution of time to censoring is the same in the two treatment arms.

However, our tests, in contrast to standard log-rank tests, need not be asymptotically α -level when censoring and failure are dependent but the censoring mechanism is non-informative. We therefore suggest that tests in our class might be used to augment standard weighted log rank tests in the analysis of right-censored failure time data when the censoring time is observed even for failures.

We stress that, if censoring is informative and both of our two non-identifiable identifying assumptions are false, tests in our class will fail to provide asymptotically α -level tests of the null hypothesis of no treatment effect on time to disease. In Section 4, we apply our results to a study of the effect of treatment on the evolution of a repeated measures outcome. In this paper, we assume a dichotomous treatment. In our companion paper, we allow for non-dichotomous vector-valued treatments.

2. The Problem

2.1. The Data

Let Z denote the dichotomous (0, 1) treatment arm indicator, Y denote the logarithm of time to (informative) censoring, and X^0 denote the logarithm of time to disease. We assume we observe n independent and identically distributed realizations of

$$Z, X \equiv X^0 \wedge Y, \quad \sigma \equiv I(X^0 < Y), Y \quad (2.1)$$

That is, we always observe time to censoring Y , but we only observe time to disease X^0 if less than time to censoring. Note that implicit in our approach is the assumption that time to disease X^0 is well defined even though the subject is censored at time Y preceding X^0 . This latent failure time approach has been criticized by Kalbfleisch and Prentice (1980) when Y represents time to death, since they did not view X^0 as well defined whenever $X^0 \not\leq Y$. However, nearly all epidemiologists whose interest is in investigating the causal effects of various etiologic agents (e.g., cigarette smoking) on specific causes of death (e.g., heart disease) naturally think in terms of latent failure times (Robins, 1986). Further, in Remark A below, we show that the results of Sec. 2.2.1 below remain true, even if following Kalbfleisch and Prentice (1980), X^0 is undefined when $X^0 \not\leq Y$.

2.2. Tests and Estimation of Causal Effects—Identifiability Issues.

2.2.1. Testing the Causal Null Hypothesis in a Randomized Trial.

One of our main interests will be in testing the causal null hypothesis of no treatment-arm effect on time to disease X^0 . To formulate this hypothesis, we use a counterfactual causal model. Following Rubin (1978), let $X^0(1)$ and $X^0(0)$ be random variables representing time to disease when treated ($z = 1$) and untreated ($z = 0$), respectively. $X^0 \equiv X^0(Z)$ can be observed while $X^0(1 - Z)$ is always missing. Similarly define $Y(1)$ and $Y(0)$. Again we only observe $Y \equiv Y(Z)$. The sharp null hypothesis of no effect of treatment arm on disease is

$$X^0(1) = X^0(0) \text{ with probability 1} \quad (2.2)$$

which implies

$$S_{X^0(1)}(t) = S_{X^0(0)}(t) \quad (2.3)$$

which, in turn, is equivalent to

$$S_{X^0}(t | Z = 1) = S_{X^0}(t | Z = 0), \quad (2.4)$$

since randomization of Z implies

$$Z \perp\!\!\!\perp \{X^0(1), X^0(0), Y(1), Y(0)\} \quad (2.5)$$

(Rubin, 1978) because $\{X^0(1), X^0(0), Y(1), Y(0)\}$ are, like eye color or age at randomization, fixed characteristics of a study subject and thus independent of treatment arm. Here $S_{X^0}(t | \cdot) = pr(X^0 > t | \cdot)$ and $A \perp\!\!\!\perp B | C$ means A is independent of B given C .

Results of Petersen (1976) imply that, in the absence of (2.5), $S_{X^0}(t | Z = j)$ is unrestricted by the joint law of the observables in (2.1) except for the inequality constraint

$$S_X(t | Z = j) \leq S_{X^0}(t | Z = j) < 1 - pr(X < t, \sigma = 1 | Z = j) \quad (2.6)$$

Eq. (2.6) says that, in each treatment arm Z , the non-identifiable survival function of X^0 is bounded below by the identifiable survival function of X and is bounded above by the identifiable survival function $1 - pr(X < t, \sigma = 1 | Z)$ that is obtained by assuming that any subject who is uncensored ($\sigma = 0$) would have a value of X^0 equal to infinity. Given (2.5), Rabinowitz, Williams, and Robins (1992) provide an α -level test of the sharp null hypothesis (2.2) that rejects only when there is evidence that

$$S_X(t | Z = z) > 1 - pr(X < t, \sigma = 1 | Z = 1 - z)$$

for either $z = 1$ or 0 ; that is, when the lower bound $S_X(t | Z = z)$ for treatment arm z exceeds the upper bound for treatment arm $Z = 1 - z$. They show, as one would expect, that such a test has poor power.

To construct a test with good power, such as that in Lin et al. (1995), further (non-identifiable) assumptions are required. To describe our assumptions, define $m(\cdot)$ to be a function satisfying $S_{Y(1)}(t) = S_{Y(0)}\{m(t)\}$. If, as we generally assume, $Y(1)$ and $Y(0)$ are absolutely continuous (w.r.t. Lebesgue measure), then $m(\cdot) = S_{Y(0)}^{-1}\{S_{Y(1)}(\cdot)\}$ is unique. Similarly, let $v(\cdot, \cdot)$ be a function satisfying $S_{X^0(1)}[t | m(Y(1)) = u] = S_{X^0(0)}[v(t, u) | Y(0) = u]$. Again $v(\cdot, \cdot)$ is unique in the absolutely continuous case. Note $m(\cdot)$ is identifiable, since Y is always observed and, under (2.5), the survival curve of $Y(z)$ equals that of the random variable Y given $Z = z$.

Example. Suppose, following Lin et al. (1995), we assume $(X^0 - \theta_0 Z, Y - \eta_0 Z)$ are independent of Z and identically distributed [i.e., (X^0, Y) follow a bivariate location shift model]. Then, given (2.5), $m(u) = u - \eta_0$ and $v(t, u) = t - \theta_0$.

The goal of this paper is to derive assumptions that are less restrictive than Lin et al.'s assumption that (X^0, Y) follows a bivariate location shift model but remains sufficient to test for and estimate the causal effect of treatment on X^0 in the presence of informative censoring by Y . To do so, it is essential to differentiate between (i) the key statistical fact about the joint distribution of the observables (2.1) that allowed Lin et al. (1995) to estimate

θ_0 , and (ii) the key non-identifiable assumption concerning the joint distribution of (X^0, Y) that provide θ_0 with a causal interpretation.

The key statistical fact is that, in a randomized trial satisfying (2.5), the function $v(t, u)$ is identified from the observable data (2.1) if and only if (t, u) lies in a particular subset $A(m, v)$ of R^2 . Indeed, the estimation and testing procedures of Lin et al. (1995), as well as those in both this and my companion paper (Robins, 1995), are simply methods for estimation of and testing hypotheses about the identifiable function $v(t, u)$ on the set $A(m, v)$. These methods are valid regardless of whether the function $v(t, u)$ has a causal interpretation.

Conditions for the identifiability of $v(t, u)$ are given in the following Lemma. Let $A(m, v) = \{(t, u); t < \min(m^{-1}(u), v^{-1}(u, u))\}$ where, for any function $h(t, u)$ that is 1 to 1 in its first argument, $h^{-1}(x, u) \equiv t$ if $h(t, u) = x$.

LEMMA 1 *Given data (2.1) and the randomization assumption (2.5), $v(t, u)$ is identifiable if and only if $(t, u) \in A(m, v)$.*

Proof: We note, by the randomization assumption (2.5), $v(t, u)$ satisfies

$$S_{X^0} [t \mid Z = 1, m(Y) = u] = S_{X^0} [v(t, u) \mid Z = 0, Y = u]. \quad (2.7)$$

However, the L.H.S. of (2.7) is only identifiable from data (2.1) for $t < m^{-1}(u)$ and the R.H.S. is only identifiable for $v(t, u) < u$. This concludes the proof. We also have the following simple lemma whose proof is left to the reader. ■

LEMMA 2 *$\{X^0(0), Y(0)\}$ and $\{v[X^0(1), m(Y(1))], m(Y(1))\}$ have the same joint distributions.*

The premise of the following corollaries to Lemma 2 provide sufficient conditions for $v(t, u)$ to have a causal interpretation as the marginal effect of Z on the distribution of X^0 .

COROLLARY 1 *If $v(t, u) = v^*(t)$ for all (t, u) and some $v^*(\cdot)$, then*

$$v^* [X^0(1)] \text{ and } X^0(0)$$

have the same distribution. Further, $v^(t) = S_{X^0(0)}^{-1} \{S_{X^0(1)}(t)\}$.*

COROLLARY 2 *If, for all (t, u) , $v(t, u) = v^*(t)$ for some $v^*(\cdot)$, then (i) the null hypothesis (2.3) holds $\Leftrightarrow v(t, u) = t \Leftrightarrow v^*(\cdot) = i(\cdot)$ where $i(\cdot)$ is the identity map, and (ii) the sharp null hypothesis (2.2) implies $v(t, u) = t$.*

Remark. Note the premises of Corollary 1 and 2 are true if $X^0(j) \perp\!\!\!\perp Y(j)$, $j = 0, 1$. Given Eq. (2.5), this is equivalent to $X^0 \perp\!\!\!\perp Y \mid Z$, i.e. conditional on treatment arm Z , the censoring variable Y and the disease variable X^0 are independent.

When the premise to the corollaries is true, $v(t, u) = v^*(t)$ has the causal interpretation as the unique function necessary to transform the distribution of X^0 when treated into

that of X^0 when untreated, i.e., $v^* \{X^0(1)\}$ and $X^0(0)$ have identical distributions. In particular, as noted above, Lin et al.'s (1995) bivariate shift model implies that the premise of the corollaries hold with $v^*(t) = t - \theta_0$ so that $X^0(1) - \theta_0$ and $X^0(0)$ have the same distribution.

Note, however, that the premise that, for all (t, u) , $v(t, u) = v^*(t)$ cannot be subjected to a consistent goodness-of-fit test, since $v(t, u)$ is not identified on $A^c(m, v)$, the complement of $A(m, v)$ in R^2 .

Example. If $v(t, u) = t - \theta_0$ on $A(m, v)$, but $v(t, u) \neq t - \theta_0$ on $A^c(m, v)$, then (i) Lin et al.'s bivariate shift model is false (i.e., misspecified); (ii) there will exist no data evidence to warn us of this misspecification; and (iii) the identifiable parameter θ_0 in the true model $v(t, u) = t - \theta_0$ for $(t, u) \in A(m, v)$ will in general have no causal interpretation; in particular, it will not be true that $X(1) - \theta_0$ and $X(0)$ have the same distribution. Table 1 exemplifies these issues. Note that, in Table 1, the sharp null hypothesis (2.2) is true, and $m(\cdot) = i(\cdot)$. However, $v(t, u) = v^*(t)$ on $A(m, v)$ with $v^*(t) = t + 1$. Note Table 1 does not contradict Corollary (2), since $v(t, u) = t - 1$ on the complement of $A(m, v)$. That is, $v(t, u)$ depends on the value of u : $v(t, u) = (t - 1)I(u \leq 2) + (t + 1)I(u > 2)$.

In summary, in Table 1, $v(t, u) = t - \theta_0$ with $\theta_0 = -1$ on $A(m, v)$, but clearly, since the sharp null hypothesis is true, θ_0 has no causal interpretation. One can regard the lack of dependence of $v(t, u)$ on u in the premise to the corollaries as a non-interaction assumption.

We have seen that the premise of corollaries 1 and 2 is a partially identifiable assumption, since $v(t, u)$ is identifiable on $A(m, v)$. To clearly separate identifiable from non-identifiable assumptions, we now consider the implications of the following non-identifiable assumption.

ASSUMPTION

$$\text{If } v(t, u) = v^*(t) \text{ on } A(m, v), \text{ then } v(t, u) = v^*(t) \text{ on } A^c(m, v) \quad (2.8)$$

Assumption (2.8) says that we can rule out *a priori* joint distributions like the one represented in Table 1. We then have the following.

THEOREM 1 *Given data (2.1) and the assumption (2.8), if $v(t, u) = v^*(t)$ on $A(m, v)$, then, (a) $v^*(X(1))$ and $X(0)$ have the same distribution, and (b) the null hypothesis (2.3) is true $\Leftrightarrow v(t, u) = t$ on $A(m, v)$.*

If we impose assumption (2.8), one might use Theorem 1 to test the null hypothesis (2.3). To do so we could proceed formally as follows. First, we would perform an α_1 -level test of the identifiable hypothesis (1) that $v(t, u) = v^*(t)$ on $A(m, v)$. Conditional on this test accepting, we would perform an α_2 -level test of the identifiable hypothesis (2) that $v^*(t) = t$ on $A(m, v)$ [that is, when both hypotheses (1) and (2) are true, the second test has level α_2 conditional on the α_1 test having been accepted].

The operating characteristics of such a procedure are as follows. If assumption (2.8), hypothesis (1) and the null hypothesis (2.3) are all true, then, with probability α_1 , we will

Table 1. A Finite Population Consisting of 4 Homologous Types of Subjects

		Logarithm of Time						
	Type	No. of Subjects	1	2	3	4	5	6
Z = 1	1	500		X ⁰			Y	
	2	500	Y		X ⁰			
	3	500				X ⁰		Y
	4	500		Y			X ⁰	
Z = 0	1	500	Y	X ⁰				
	2	500			X ⁰		Y	
	3	500		Y		X ⁰		
	4	500					X ⁰	Y

Remarks:

- (1) The randomization assumption (2.5) holds.
- (2) $m(u) = u$ since $pr[Y(1) = u] = pr[Y(0) = u]$ for all u .
- (3) $v(2, 5) = 3, v(4, 6) = 5, v(3, 1) = 2, v(5, 2) = 4$, so $v(t, u) = (t - 1)I(u \leq 2) + (t + 1)I(u > 2)$
- (4) $A(m, v) = \{(2, 5), (4, 6)\}; A^c(m, v) = \{(3, 1), (5, 2)\}$
- (5) $v(t, u) = t + 1$ on $A(m, v); v(t, u) = t - 1$ on $A^c(m, v)$
- (6) The assumption (8) of rank preservation is false, since for type 3, $6 = Y(1) > Y(0) = 2$, but for type 4, $2 = Y(1), Y(0) = 6$

reject hypothesis (1) and make no statement about (2.3); with probability $(1 - \alpha_1)(1 - \alpha_2)$, we will correctly accept (2.3); and with probability $(1 - \alpha_1)\alpha_2$, we will falsely reject (2.3). [However, suppose hypothesis (1) is false but (2.3) is true. Then, given that the α_1 -level test of the identifiable hypothesis (1) is falsely accepted (due to poor power), the true conditional level of the conditional α_2 -level test of (2.3) will not equal α_2 .]

Example. In specifying a bivariate shift model, Lin et al. (1995) assumed *a priori* both the non-identifiable restriction (2.8) and the identifiable restriction that $v(t, u) = v^*(t) = t - \theta_0$ on $A(m, v)$. It would have been preferable for them, rather than just assuming this latter restriction, to have tested for dependence of $v(t, u)$ on u in $A(m, v)$, i.e., to have tested hypothesis (1), using an α_1 -level test.

Remark. In Table 2, we demonstrate that, even given assumption (2.8), the null hypothesis (2.3) does not imply $v(t, u) = t$ on $A(m, v)$ [although, by Theorem 1, the converse is true]. Hence, even given assumption (2.8), we cannot obtain an α -level test of the sharp null hypothesis (2.2) or of (2.3) simply by testing whether $v(t, u) = t$ on $A(m, v)$. Specifically, in Table 2, (i) hypotheses (2.2) and (2.3) are true; (ii) no subject is censored, so $A^c(m, v)$

Table 2. A Finite Population Consisting of 2 Homologous Types of Subjects

		Logarithm of Time						
		No. of	1	2	3	4	5	6
Type	Subjects							
Z = 1	1	500	X ⁰		Y			
	2	500		X ⁰		Y		
Z = 0	1	500	X ⁰			Y		
	2	500		X ⁰	Y			

is empty; (iii) $v(t, 3) = t + 1$ and $v(t, 4) = t - 1$, so $v(t, u)$ depends on u . Of course, since there is no censoring in Table 2, the distributional null hypothesis (2.3) can be tested using any standard 2-sample test even though X^0 and Y are dependent given Z . However, in practice, we will usually observe significant censoring.

Rank Preservation

An alternative non-identifiable assumption under which the sharp null hypothesis (2.2) does imply $v(t, u) = t$ on $A(m, v)$ is the assumption that censoring Y is rank preserving, i.e., by definition,

$$m\{Y(1)\} = Y(0) \text{ w.p. } 1 \tag{2.9}$$

We call assumption (2.9) rank preserving, since it is equivalent to the assumption that given 2 realizations i and k of $Y(0)$ and $Y(1)$, if $y_i(0)$ exceeds $y_k(0)$, then $y_i(1)$ exceeds $y_k(1)$, i.e., ranks under the absence of treatment are preserved under treatment. The assumption of rank preservation is non-identifiable. We have

THEOREM 2 Given data (2.1) and the rank preservation assumption (2.9), the sharp null hypothesis (2.2) implies $v(t, u) = t$ on $A(m, v)$.

Proof: Obvious. ■

Remark. Given data (2.1) and the rank preservation assumption (2.9), if the sharp null hypothesis (2.2) is false but the distributional null hypothesis (2.3) is true, we cannot conclude $v(t, u) = t$ on $A(m, v)$.

Example. The rank preservation assumption (2.9) does not hold for Table 1 or Table 2. Note, in Table 2, the sharp null hypothesis is true but $v(t, u) \neq t$ on $A(m, v)$ since rank preservation is false. However, in Table 2, if we relabel, for $Z = 1$, type 1 subjects as type

2 and vice-versa, then we do have rank preservation, the null hypothesis (2.3) is true, but the sharp null hypothesis (2.2) is false, so Theorem 2 is not contradicted.

Except for special circumstances such as those described in the following example, the assumption of rank preservation is usually biological untenable.

Example. Suppose, as in Robins and Rotnitzky (1992, Appendix 4) and Robins (1989), censoring of X^0 is only due to time to end of follow-up Y , defined as the final common date of study termination minus the date of study entry (i.e., there is no censoring by death or loss to follow-up). Then, $Y = Y(1) = Y(0)$ w.p.1 since Y for a subject is determined at calendar date of entry prior to random assignment to treatment Z . Thus assumption (2.9) holds and $m(u) = u$.

Remark A. Suppose the only type of informative censoring was death, and, following Kalbfleisch and Prentice (1980), we make the following assumption.

ASSUMPTION (*Kalbfleisch and Prentice*): $X^0(z)$ is defined if and only if $X^0(z) \leq Y(z)$.

In this setting, Robins (1986) proposed the following definition.

DEFINITION *The sharp null hypothesis of no causal effect of treatment Z on time to disease X^0 holds if and only if*

(i) $X^0(1) = X^0(0)$ whenever $X^0(z) \leq Y(z)$ for both $z = 0$ and $z = 1$ and

(ii) If $X^0(z) \leq Y(z)$ but $X^0(1-z) \not\leq Y(1-z)$, then $X^0(z) > Y(1-z)$, $z = 0, 1$.

To see the reason for (ii), suppose treatment had no effect on X^0 . Then if, as for type 4 subjects in Table 1, $X^0(0) = 5$ and $Y(0) = 6$ and $Y(1) = 2$, then $X^0(1)$ should not be less than 2. The data in Table 1 satisfies the Robins (1986) definition of the sharp causal null hypothesis. However, $v(t, u) = t + 1 \neq t$ where identified (i.e., on $A(m, v)$). Thus we see that, even under the Kalbfleisch–Prentice assumption an α -test of the identifiable hypothesis $v(t, u) = t$ on $A(m, v)$ in Table 1 is not an α -level test of the sharp causal null hypothesis.

2.2.2. Estimation.

As described above, we can use the function $S_{X^0(0)}^{-1}\{S_{X^0(1)}(t)\} \equiv v^\dagger(t)$ to quantify the effect of treatment z on the marginal distribution of time to disease X^0 since $v^\dagger(X(1))$ and $X(0)$ have the same distribution. When $v(t, u) = v^*(t)$ for all (t, u) , $v^*(t) = v^\dagger(t)$ by Corollary 1.

Example. Suppose (2.5) holds and $v(t, u) = v^*(t) = t + \theta_0^{(0)}$, then $v^\dagger(t) = t + \theta_0^{(0)}$, i.e., as in Lin et al. (1995), X^0 follows a shift model with parameter $-\theta_0^{(0)}$, equivalently, $(X^0 + \theta_0^{(0)}Z) \perp\!\!\!\perp Z$.

However, if $v(t, u)$ depends on u , then even were $m(u)$ and $v(t, u)$ known for all (t, u) and even if we knew that we had strong rank preservation [in the sense that $\{v[X^0(1), m(Y(1))], m(Y(1))\}$ of Lemma 2 equalled $(X^0(0), Y(0))$ with probability 1], $v^\dagger(t)$ would not be identified when there is censoring of X^0 by Y among subjects with $Z = 0$.

Example. Suppose

$$m(u) = u, v(t, u) = t + \theta_0^{(0)} + \theta_0^{(1)}u \quad (2.10)$$

for known *non-zero* $\theta_0^{(0)}$ and $\theta_0^{(1)}$. Then, by its definition, $v^\dagger(t)$ satisfies

$$pr[\theta_0^{(0)} + \theta_0^{(1)}Y(0) + X^0(0) > t] = pr[X^0(0) > v^\dagger(t)].$$

$v^\dagger(t)$ is not identified since, given the data (2.1) and randomization assumption (2.5), the joint distribution of $(X^0(0), Y(0))$ is not identified.

We now discuss the implications of the above results for causal inference. Suppose we correctly specify the identifiable model

$$v(t, u) = t + \theta_0^{(0)} + \theta_0^{(1)}u \text{ on } A(m, v) \quad (2.11)$$

where $\theta_0 = (\theta_0^{(0)}, \theta_0^{(1)})'$ is a parameter vector to be estimated. In Robins (1995), I will describe how to obtain joint $(1 - \alpha)$ percent large sample confidence intervals for the identifiable parameters $(\theta_0^{(0)}, \theta_0^{(1)})$ given data (2.1) and the randomization assumption (2.5).

Here, we discuss the causal interpretation of the parameters of model (2.11) under various assumptions. Suppose first that we impose the non-identifiable assumption that model (2.11) holds on $A^c(m, v)$ as well and thus holds for all (t, u) . Note that, given model (2.11), the assumption (2.8) is a special case of this assumption. Then, if $\theta_0^{(0)} = \theta_0^{(1)} = 0$, the null hypothesis (2.3) is true. If $\theta_0^{(0)} \neq 0, \theta_0^{(1)} = 0$, then $v^\dagger(t) = t + \theta_0^{(0)}$ and the causal effect of Z on the law of X^0 is a shift of size $-\theta_0^{(0)}$. If $\theta_0^{(0)} \neq 0$ and $\theta_0^{(1)} \neq 0$, the causal parameter $v^\dagger(t)$ is nowhere identified and $\theta_0^{(0)}$ and $\theta_0^{(1)}$ have no meaningful causal interpretation. However, if we impose the rank preservation assumption (2.9) then, even if we do not assume model (2.11) holds on $A^c(m, v)$, $\theta_0^{(0)} + \theta_0^{(1)}u$ equals the causal effect of treatment z on the subset of subjects with $Y(0) = u$ for any $(t, u) \in A(m, v)$. That is, if we define $v^\dagger(t, u) \equiv S_{X^0(0)|Y(0)=u}^{-1} \{S_{X^0(1)|Y(0)=u}(t)\}$ so $v^\dagger(X(1), u)$ and $X(0)$ have the same distribution given $Y(0) = u$, then

$$t + \theta_0^{(0)} + \theta_0^{(1)}u = v^\dagger(t, u), (t, u) \in A(m, v).$$

If we assume that (2.11) holds in $A^c(m, v)$ then $\theta_0^{(0)} + \theta_0^{(1)}u$ equals $v^\dagger(t, u)$ for all t, u .

Remark. Suppose we impose (2.11), but we do not assume the model in (2.11) holds on $A^c(m, v)$. Then even if we assume (2.9) and $\theta_0^{(1)} = 0$, we cannot conclude $v^\dagger(t) = \theta_0^{(0)} + t$ and $v^\dagger(t)$ is not identified. In addition, even if $\theta_0^{(0)} = \theta_0^{(1)} = 0$, we cannot conclude that the null hypothesis (2.3) is true even if (2.9) is assumed true. Further, if we do not assume (2.9), then even if $\theta_0^{(0)} \neq 0, \theta_0^{(1)} \neq 0$, we cannot conclude the null hypothesis (2.2) is false.

3. A Class of Tests

Recall that a standard weighted log rank test based on weight function $w(t)$ is given by $\psi = \text{Num}/\text{Denom}$ where

$$\text{Num} = \sum_{(i:\sigma_i=1)} w(X_i) (Z_i - E_i), \quad E_i = \sum_j I(X_j > X_i) Z_j / \sum_j I(X_j > X_i),$$

$$\text{Denom} = \left\{ \sum_{(i:\sigma_i=1)} \{w(X_i)\}^2 E_i (1 - E_i) \right\}^{\frac{1}{2}}.$$

It is well known that the test that rejects when $|\psi| > z_{\alpha/2}$, with $z_{\alpha/2}$ the upper $1 - \alpha/2$ percentile of a standard normal distribution is an asymptotically α -level test of the sharp null hypothesis (2.2) when the randomization assumption (2.5) and the non-identifiable assumption of non-informative censoring, i.e.,

$$\lambda_{X^0}(t | Y > t, Z) = \lambda_{X^0}(t | Z) \quad (3.1)$$

are true, where $\lambda_{X^0}(t | \bullet)$ is the hazard of X^0 given \bullet .

For pedagogic purposes, suppose for the moment that the distribution of Y given Z , and thus the function $m(u)$ was known. Let $X^* = \min(X^0, Y^*)$, where $Y^* = \min(Y, m(Y))$ if $Z = 1$ and $Y^* = \min(Y, m^{-1}(Y))$ if $Z = 0$, $\sigma^* = I(X^* = X^0)$, and $\psi^* = \text{Num}^*/\text{Denom}^*$ where Num^* and Denom^* are defined like Num and Denom but with σ^* and X^* replacing σ and X . We then have the following results.

THEOREM 3 *If (i) $v(t, u) = v^*(t)$ for all (t, u) and some $v^*(t)$, or (ii) the assumption (2.9) of rank preservation holds, then the test that rejects when $|\psi^*| > z_{\alpha/2}$ is an asymptotically α -level test of the sharp null hypothesis (2.2).*

Proof: We have seen previously that the premises of the theorem plus the sharp null hypothesis imply that $v(t, u) = t$ for all (t, u) which, by Lemma 2, implies $\{X^0, Y^*\}$ is independent of Z . Hence,

$$\lambda_{X^*, \sigma^*=1}(u | Z) = \lambda_{X^*, \sigma^*=1}(u) \quad (3.2)$$

where $\lambda_{X^*, \sigma^*=1}(u | Z)$ is the cause-specific hazard of X^* corresponding to $\sigma^* = 1$ given Z . However, the standard theory of weighted log-rank tests implies $|\psi^*| > z_\alpha$ is an asymptotically α -level test of (3.2). ■

COROLLARY 3 *If $X^0 \perp\!\!\!\perp Y \mid Z$, then $|\psi^*| > z_{\alpha/2}$ and $|\psi| > z_{\alpha/2}$ are both α -level tests of (2.2).*

Proof: As noted previously, $X^0 \perp\!\!\!\perp Y \mid Z$ implies premise (i) of Theorem 3 and also implies the assumption (3.1) of non-informative censoring. ■

COROLLARY 4 *If $Y \perp\!\!\!\perp Z$ so that $m(u) = u$ and either premise (i) or (ii) of Theorem 3 are true, then $|\psi| > z_{\alpha/2}$, as well as $|\psi^*| > z_{\alpha/2}$, is an asymptotically α -level test of (2.2).*

Proof: If $m(u) = u$ then $m(Y) = Y$, $X^* = X$, $\sigma^* = \sigma$, and thus Eq. (3.2) is equivalent to the statement that the cause-specific hazard of X corresponding to $\sigma = 1$ does not depend on Z , which is the hypothesis tested by $|\psi| > z_{\alpha/2}$. ■

The power of the tests $|\psi| > z_{\alpha/2}$ and $|\psi^*| > z_{\alpha/2}$ under the assumptions of Corollaries 3 and 4 will be briefly compared in the companion paper (Robins, 1995).

Remark. Note that if the assumption (3.1) of non-informative censoring is true but premises (i) and (ii) of Theorem 3.1 are false (and thus $X^0 \perp\!\!\!\perp Y \mid Z$ is false), then $|\psi| > z_{\alpha/2}$ is, but $|\psi^*| > z_{\alpha/2}$ is not, an asymptotically α -level test of (2.2) [since (3.1) in contrast to the assumption that $Y \perp\!\!\!\perp X^0 \mid Z$, does not imply that (3.2) is true]. Conversely, if (3.1) is false, but either premise (i) or premise (ii) of Theorem 3 is true, $|\psi^*| > z_{\alpha/2}$ is, but $|\psi| > z_{\alpha/2}$ is not, an asymptotically α -level test of (2.2).

A setting in which the assumption (3.1) of non-informative censoring could be true but the assumption $Y \perp\!\!\!\perp X^0 \mid Z$ would be false would be a setting in which the hazard of censoring increases when subjects have been observed to develop disease.

In practice, since $m(u)$ is an unknown function, we would use a test statistic numerator \widehat{Num}^* in which $m(u)$ in Num^* is replaced by a non-parametric estimate. Further, we must replace $Denom^*$ by a new variance estimate \widehat{Denom}^* that properly accounts for the added variability in \widehat{Num}^* due to estimation of $m(u)$. The resulting test statistic $|\widehat{\psi}^*| > z_{\alpha/2}$, $\widehat{\psi}^* \equiv \widehat{Num}^*/\widehat{Denom}^*$ will be an asymptotically α -level test of the sharp null hypothesis under the premises of Theorem 3.

In our companion paper, we present an alternative class of test statistics that also produce asymptotically α -level tests of the sharp null hypothesis (2.2) under the premises of Theorem 3 but are even powerful than the class of test statistics $\widehat{\psi}^*$. These tests are based on the fact that the premises of Lemma 3 imply $(X^*, \sigma^*) \perp\!\!\!\perp Z \mid M$ where $M \equiv m(Y)$ if $Z = 1$ and $M \equiv Y$ otherwise. Further, we present a test of premise (i) of Theorem 3 by constructing an α_1 -level test of the hypothesis that $v(t, u) = v^*(t)$ on $A(m, v)$ for some $v^*(t)$.

4. Repeated Measures Outcomes

Suppose now our interest is in testing for and estimating the effect of treatment on CD4 lymphocyte count X_t recorded at T predetermined times from randomization. Without loss of generality, denote the T times by $t = 1, \dots, T$. We observe $Z_i, \bar{X}_i(Y_i), Y_i, i = 1, \dots, n$

where $\bar{X}(Y) \equiv (X_1, \dots, X_{int(Y)})'$ and $int(Y)$ is the largest integer less than or equal to both Y and T . With Y , Z , and M as defined previously, define $v(t, u) = E[X_t | M = u, Z = 1] - E[X_t | M = u, Z = 0]$. Identifiability results wholly analogous to those obtained in Sec. 2 can be derived. Specifically, let $X_t(z)$ be the possibly counterfactual value of CD4 count if treatment z were given. If Z were assigned by physical randomization [i.e. Eq. (2.5) holds] and the rank preservation assumption (2.9) holds, then

$$v(t, u) = E[X_t(1) - X_t(0) | Y(0) = u]$$

and thus has the causal interpretation as the mean effect of treatment 1 compared to treatment 0 amongst subjects with $Y(0) = u$. If Z is randomized, and $v(t, u) = v^*(t)$ for all (t, u) , then $v^*(t) = E[X_t(1) - X_t(0)]$ is the marginal mean effect of treatment Z . $v(t, u)$ is only identified on $A(m, v) = \{(t, u); t < \min[m^{-1}(u), v^{-1}(u, u)]\}$. Further, $v(t, u) = v^*(t)$ on $A(m, v)$ does not imply that $v(t, u) = v^*(t)$ for all (t, u) and thus does not imply either that $E[X_t(1) - X_t(0)] = v^*(t)$ or that the identifiable function $v^*(t)$ on $A(m, v)$ has any causal interpretation. In our companion paper, estimation and testing of the function $v(t, u)$ is based on the identity $E[I(Y > t)\varepsilon_t | Z, M, I(Y > t)] = E[I(Y > t)\varepsilon_t | M, I(Y > t)]$ where $\varepsilon_t = X_t - v(t, M)$ if $Z = 1$ and $\varepsilon_t = X_t$ otherwise.

References

1. J. D. Kalbfleisch, and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley, New York, 1980.
2. D. Y. Lin, J. M. Robins, and L. J. Wei, "Comparing Two Failure Time Distributions in the Presence of Informative Censoring," *Biometrika*, 1995, (to appear).
3. A. V. Peterson, "Bounds for a Joint Distribution Function with Fixed Sub-distribution Functions: Application to Competing Risks." *Proc. Natl. Acad. Sci.*, vol. 73, pp. 11-13, 1976.
4. D. Rabinowitz, P. Williams, and J. Robins, "Testing for Treatment Effects with Dependent Competing Risks." *Technical Report*, Department of Epidemiology, Harvard School of Public Health, Boston, MA, 1992.
5. J. M. Robins, "A New Approach to Causal Inference in Mortality Studies with Sustained Exposure Periods—Application to Control of the Healthy Worker Survivor Effect", *Mathematical Modelling*, Vol 7, pp. 1393-1512, 1986.
6. J. M. Robins, "Addendum to 'A New Approach to Causal Inference in Mortality Studies with Sustained Exposure Periods—Application to Control of the Healthy Worker Survivor Effect'," *Computers and Mathematics with Applications*, Vol. 14, pp. 923-945, 1987a.
7. J. M. Robins, "A Graphical Approach to the Identification and Estimation of Causal Parameters in Mortality Studies with Sustained Exposure Periods," *Journal of Chronic Disease*, Vol. 40, Supplement 2, pp.-139s-161s, 1987b.
8. J. M. Robins, "The Analysis of Randomized and Non-randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies", *Health Service Research Methodology: A Focus on AIDS*, L. Sechrest, H. Freeman, A. Mulley, NCHSR, U.S. Public Health Service, pp. 113-159, 1989.
9. J. M. Robins, "Analytic Methods for HIV Treatment and Cofactor Effects," *Methodological Issues of AIDS Behavioral Research*, D. G. Ostrow, R. Kessler, Plenum Press, New York, pp. 213-287, 1993.
10. J. M. Robins, "An Analytic Method for Randomized Trials with Informative Censoring, Part II", *Lifetime Data Analysis*, Vol. 1, 1995.

11. J. M. Robins, and A. Rotnitzky. "Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers", *AIDS Epidemiology—Methodological Issues*, N. Jewell, K. Dietz, V. Farewell, Birkhäuser, Boston, MA, pp. 297–331, 1992.
12. D. B. Rubin, (1978). "Bayesian Inference for Causal Effects: The Role of Randomization", *The Annals of Statistics*, Vol. 6, pp. 34–58, 1978.