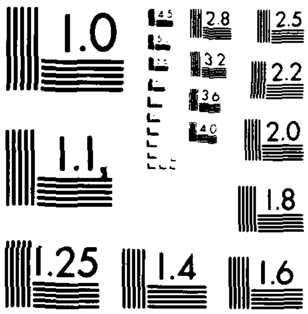
AD-A131 338 LINEAR RANK PROCEDURES FOR MATCHED OBSERVATIONS(U) 1/1
SCHOOL OF AEROSPACE MEDICINE BROOKS AFB TX
J E MICHALEK ET AL. JUN 83 USAFSAM-TR-83-16

UNCLASSIFIED F/G 12/1 NL

END
DATE
FILMED
9 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

Report USAFSAM-TR-83-16

LINEAR RANK PROCEDURES FOR MATCHED OBSERVATIONS

Joel E. Michalek, Ph.D.
Daniel Mihalko, Ph.D.

June 1983

Final Report for Period 1 January 1981 - 19 February 1982

Approved for public release; distribution unlimited.

USAF SCHOOL OF AEROSPACE MEDICINE
Aerospace Medical Division (AFSC)
Brooks Air Force Base, Texas 78235

NOTICES

This final report was submitted by personnel of the Biomathematics Modeling Branch, Data Sciences Division, USAF School of Aerospace Medicine, Aerospace Medical Division, AFSC, Brooks Air Force Base, Texas, under job order 2767-00-F1.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nations.

This report has been reviewed and is approved for publication.

JOEL E. MICHALEK, Ph.D.
Project Scientist

RICHARD A. ALBANESE, M.D.
Supervisor

ROY L. DEHART
Colonel, USAF, MC
Commander

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: USAFSAM-TR-83-16
2. GOVT ACCESSION NO.: A131338
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): LINEAR RANK PROCEDURES FOR MATCHED OBSERVATIONS
5. TYPE OF REPORT & PERIOD COVERED: Final Report, 1 Jan 1981 - 19 Feb 1982
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s):

Joel E. Michalek, Ph.D.
Daniel Mihalko, Ph.D.

8. CONTRACT OR GRANT NUMBER(s):
9. PERFORMING ORGANIZATION NAME AND ADDRESS: USAF School of Aerospace Medicine (BRM), Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas 78235
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62202F, 276700F1
11. CONTROLLING OFFICE NAME AND ADDRESS: USAF School of Aerospace Medicine (BRM), Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas 78235
12. REPORT DATE: June 1983
13. NUMBER OF PAGES: 14
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
18. SUPPLEMENTARY NOTES:

Daniel Mihalko is Assistant Professor, Department of Mathematics and Statistics, University of Nebraska, Lincoln, Neb. 68588. This report was written while he was a USAF University Resident Research Associate at the USAF School of Aerospace Medicine.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

Matched designs
Linear rank tests
Survival studies

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

General procedures are developed for matched designs in which the observations are subject to censorship on the right. The asymptotic distributions of these procedures are presented and proved. Various special cases are compared by Monte Carlo simulation. These methods are particularly applicable to epidemiological matched survival studies in which each case is matched to several controls.

CONTENTS

Tables

ACKNOWLEDGMENT

For programming the Monte Carlo study, we are grateful to Mr. Thomas J. White, Programmer (under contract with the OAO Corporation, 1222 North Main Avenue, San Antonio, Tex. 78212), in the USAFSAM Biomathematics Modeling Branch.

LINEAR RANK PROCEDURES FOR MATCHED OBSERVATIONS

1. INTRODUCTION

A statistical problem at the USAF School of Aerospace Medicine concerned an investigation of the effect of herbicide exposure on survival in Air Force personnel. A current 20-year Air Force study, of health effects in personnel after exposure to phenoxy herbicides, is a matched nonconcurrent prospective design in which each of approximately 1200 exposed subjects is assigned five control subjects matched on age, race, and job description (Ref. 8). In the mortality component of this study, the response is death, with censoring being due to loss-to-follow-up or survival to the time of analysis. An exposed subject with his matched controls, termed a "match set," constitutes a separate stratum; for survival depends on the values of the matching variables. Exposed subjects essentially determine unique strata, because age is measured in months. The number of distinct ages is nearly equal to the number of exposed subjects, so this study can be viewed as consisting of a large number of small strata, each of size 6. Here, the statistical issues of concern arise in the comparison of survival data on exposed and control subjects.

Matched survival studies, such as the ongoing USAF herbicide study, motivate the need for the justification of large sample approximations to the distributions of stratified procedures for comparing censored survival data on two groups under the nonstandard asymptotics of numerous, relatively small, fixed-size strata. In such studies, the sampling unit is a match set which, in the 20-year herbicide study, is an exposed subject and his matched controls; and, in litter-matched studies on animals, is a set of litter mates. Match sets are thus bounded in size so that increasing the study size is equivalent to increasing the number of the match sets.

The purpose here is to show that the usual stratified forms of two-sample linear rank tests maintain their asymptotic properties in the nonclassical case in which sample sizes within strata are fixed and the number of strata increase without bound. Breslow has discussed the asymptotic properties of four relative risk estimators in the same situation (Ref. 2). We also present the results of a Monte Carlo study of the power properties of several linear rank statistics on data generated in this way.

We assume that $N(R+1)$ subjects have been allocated to two treatment groups, labelled 0 and 1, by matching to each of the N subjects in group 0 a set of R subjects in group 1, the matching being with respect to one or more confounding variables. The observations can be expressed by (t,Δ), where Δ is an indicator variable that equals 1 if the observation at time t is uncensored and its actual value is t; Δ equals 0 if the observation is right censored at t. The data are assumed to consist of N match sets of the form $M=\{(t_1,\Delta_1), (t_2,\Delta_2), \ldots, (t_{R+1},\Delta_{R+1})\}$, where (t_1,Δ_1) is a group 0 response, and $(t_2,\Delta_2), \ldots, (t_{R+1},\Delta_{R+1})$ are group 1 responses.

Formally, $t=\min(t^\circ,u)$, where t° is the actual response time and u is a random censoring variable. The usual assumptions about the censoring variables are maintained here (Ref. 3); namely, that all censoring variables are mutually independent and independent of all true response times, and that $P(t^\circ < u) > 0$. We also assume that we are given N independent, but not necessarily identically distributed, match sets: M_i, i=1,2 ..., N. We do not assume that $t_1, t_2, \ldots, t_{R+1}$ are mutually independent. The hypothesis, H_0, to be tested is that, for each match set M, the joint density of $(t_1^\circ, t_2^\circ, \ldots, t_{R+1}^\circ)$ is symmetric in its arguments. This hypothesis is equivalent to the equality of all joint marginal distributions for the two groups. If the observations within match sets are assumed independent, then H_0 reduces to the equality of the two population response distributions.

The analysis of match sets of the form M with R=1 with a common censoring variable for members of a given match set has been considered by Woolson and Lachenbruch who developed linear rank procedures for testing location (Ref. 18). Their method extends easily to the case R>1, but requires the added restriction of a common censoring variable within each match set, a restriction not met in many applications.

Using a generalized U-statistic, Wei analyzed match set data of the form M with R=1 under the assumption that all match sets were identically distributed; each match set member, however, was allowed to have a separate independent censoring variable (Ref. 17). His method extends to the case R>1 with the assumption of identically distributed match sets; but in some applications, particularly those in which response time depends on the matching variable, this assumption may not be warranted.

Mantel, Bohidar, and Ciminera used, without asymptotic justification, a test which we call LB in our Monte Carlo study (Ref. 10). They also propose a method for recovering data from broken match sets. For the recovery method to be valid, the assumption of identically distributed match sets is necessary.

Mantel and Ciminera, using an approach different from that considered here, have formulated an extension of the logrank test to accommodate littermatched data (Ref. 11). Their procedure is valid for separate censoring and nonidentically distributed matched sets, and is therefore a competitor to those procedures considered here. The Mantel-Ciminera test is included in our Monte Carlo study of logrank extensions for matched data.

The approach used here is the same as that for stratified samples. An established two-sample statistic is calculated on each match set. These statistics are then summed and standardized. That this standardized form is asymptotically null distributed under match set sampling as standard normal does not follow from the usual asymptotic theory for stratified samples. In Section 3, we establish the asymptotic normality of this standardized form. The advantages of this approach are that it permits separate censoring for each subject, and that it is valid for nonidentically distributed match sets.

In the next Section, we summarize some of the most popular classes of linear rank tests for arbitrarily right censored data. The properties of these tests are then used in our asymptotic justification in Section 3. The more common special cases of these procedures are compared in a Monte Carlo study in Section 4.

2. TWO-SAMPLE PROCEDURES

In this Section, we briefly review two general classes of two-sample linear rank procedures, both of which have been extensively investigated. Suppose we have a two-sample problem with combined sample size n. Prentice introduced a class of statistics of the form (Ref 13.):

$$T' = \sum_{j=1}^{k} (c_j z_{(j)} + \sum_{\xi=1}^{m_j} C_j z_{j\xi}),$$

in which the scores, c_j and C_j, j=1,2, ..., k, are functions of a density f(.) and the observed pattern of censoring (Ref. 13: eq. 17); and, with $t_{(1)}<t_{(2)}<\ldots<t_{(k)}$ denoting the ranked times of the k (<n) uncensored observations with $t_{(0)}=0$ and $t_{(k+1)}=\infty$, m_j is the number of censored observations in $[t_{(j)}, t_{(j+1)})$; the $z_{j\xi}$ are the population identifiers of the m_j subjects censored in $[t_{(j)}, t_{(j+1)})$, j=1, 2, ...,k and ξ=1,2, ...,m_j; and $z_{(j)}$ is the population identifier of the subject with response time $t_{(j)}$.

Prentice showed that T' has certain optimality properties under the accelerated failure time model, and introduced the small sample variance estimator (Ref. 13):

$$\hat{\sigma}_1^2 = n^{-1} \sum_{i=1}^{k} (c_j^2 + m_j C_j^2).$$

Under the assumption that the censoring variables are independent and identically distributed, $\hat{\sigma}_1^2$ is an unbiased estimator of Var(T').

Prentice and Marek (Ref. 14) and Mehrotra, Michalek, and Mihalko (Ref. 12) have shown that T' can be written in the form:

$$T'' = \sum_{j=1}^{k} a_j (z_{(j)} - p_j),$$

in which $a_j = c_j - C_j$, and $p_j = n_{1j}/n_j$; n_{1j} is the number of subjects from group 1 at risk; and n_j is the total number of subjects at risk at $t_{(j)}-0$, j=1,2, ..., k. The statistic T" was introduced by Tarone and Ware (Ref. 16), and was further investigated by Schoenfeld (Ref. 15). The variance of T" was estimated by these authors as:

$$\hat{\sigma}_2^2 = \sum_{j=1}^{k} a_j^2 p_j (1-p_j).$$

Since T' and T" are equivalent, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ merely represent two different variance estimators for the same statistic. Under the hypotheses H_0, the means of T' and T" are zero, so that a third variance estimator is

$$\hat{\sigma}_3^2 = T'^2 .$$

3. ASYMPTOTIC RESULTS

Let T be a generic symbol for a two-sample linear rank statistic for H_0. Let T_i be T computed on the ith match set, M_i; let $\hat{\sigma}_i^2$ be an estimator of the variance, σ_i^2, of T_i. Then, if $E(T_i)=0$ for all i, the usual stratified procedure would be to compute:

$$Q = \left(\sum_{i=1}^{N} T_i \right) / \left(\sum_{i=1}^{N} \hat{\sigma}_i^2 \right)^{\frac{1}{2}}$$

and treat Q as approximately standard normal. We will now show the asymptotic normality of Q in the nonstandard situation, when N increases without bound while the match set sizes, R+1, remain fixed. For this purpose, we need the following three results:

Theorem 3.1. Lindeberg's Theorem (uniformly bounded case). Let T_1, T_2, ..., T_N be independent with $E(T_i)=0$ and $E(T_i^2)=\sigma_i^2$, i=1,2, ... N. If $|T_1|$, $|T_2|$, ..., $|T_N|$ are uniformly bounded for all N, and if $\sum_{i=1}^{N} \sigma_i^2$ diverges as N tends to infinity, then $\left(\sum_{i=1}^{N} T_i \right) / \left(\sum_{i=1}^{N} \sigma_i^2 \right)$ converges in distribution to a standard normal distribution.

Proof: See, for example, Ash (Ref. 1: pp. 336-337).

Lemma 3.1. If $E(\hat{\sigma}_i^2)=\sigma_i^2$, i=1,2,...N, and if $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, ...,$\hat{\sigma}_N^2$ are uniformly bounded, then $\sum_{i=1}^{N} \hat{\sigma}_i^2 / \sum_{i=1}^{N} \sigma_i^2$ converges to unity almost as surely as N tends to infinity.

Proof: From the uniform boundedness an unbiasedness of $\hat{\sigma}_i^2$, it follows that the σ_i^2 are uniformly bounded and that a finite number B exists, such that $E\,|\hat{\sigma}_i^2-\sigma_i^2|^{4/3} \leq B$, for i=1,2, ..., N. This inequality implies:

$$\lim_{N\to\infty} \sum_{i=1}^{N} \frac{E(|\hat{\sigma}_i^2-\sigma_i^2|^{4/3})}{i^{4/3}} \leq B \sum_{i=1}^{\infty} i^{-4/3} < \infty$$

This inequality, together with Chung's corollary (Ref. 4: Theorem 5.4.1, p. 124), implies that $N^{-1}\,|\sum_{i=1}^{N} \hat{\sigma}_i^2 - \sum_{i=1}^{N} \sigma_i^2|$ converges to zero almost surely as N tends to infinity. The uniform boundedness of $\hat{\sigma}_i^2$ and σ_i^2 implies that both $N^{-1}\sum\hat{\sigma}_i^2$ and $N^{-1}\sum\sigma_i^2$ converge. The lemma now follows.

Our main result now follows, as a corollary.

Corollary 3.1. Under the conditions of Theorem 3.1 and Lemma 3.1, Q converges in distribution to a standard normal distribution as N tends to infinity.

Proof: Apply Slutsky's Theorem to the results of Theorem 3.1 and Lemma 3.1.

We now show that T' (and thus its equivalent form T"), σ_i^2, $\hat{\sigma}_{1i}^2$, $\hat{\sigma}_{2i}^2$, and $\hat{\sigma}_{3i}^2$ satisfy the conditions of Theorem 3.1 and Lemma 3.1.

For any particular scoring system, T' is finite, thus implying that $|T_i'|$ and $\hat{\sigma}_{1i}^2$, $\hat{\sigma}_{2i}^2$, and $\hat{\sigma}_{3i}^2$ are all uniformly bounded. In addition, the assumption that $P(t^\circ < u) > 0$ implies that the σ_i^2 are all uniformly bounded from below, away from zero, making $\sum\sigma_i^2$ diverge. The statistic T' has zero expectation under H_0. Finally, all three variance estimators, $\hat{\sigma}_{1i}^2$, $\hat{\sigma}_{2i}^2$, and $\hat{\sigma}_{3i}^2$, are unbiased. Thus T', σ_i^2, and the three variance estimates satisfy the necessary conditions of the theory.

With the equivalence of T' and T", we have three classes of statistics of the form Q, each asymptotically normal under H_0. In the next Section, we compare the powers of representatives of each of these three classes via a Monte Carlo study.

4. A MONTE CARLO STUDY

4.1. Data Generation Methods

All computations used to produce the tables of this Section were performed on a DEC VAX 11/780 computer. All pseudorandom numbers were generated using the Linear Congruential Sequence Method (Ref. 7: p. 9).

Our approach was to simulate match sets with the assumption that the single subject in each match set from population 0 had known, and therefore given and fixed, matching variables. In practice, these variables are usually chosen by some random process beyond the control of the experimenter. In these simulations, the matching variables were represented by a single random variable. We first chose N random values for this random variable; each of these N values was then used as a parameter for the true response time distributions of the observations composing each match set. In so doing, we simulated match set data wherein all members of a given match set had the same value of the matching variable, and all response time distributions depended on the value of the matching variable in the same way.

After true response times were generated for a given match set, censoring variables were generated as the minimum of these times, as specified in Section 1. Censoring variable distributions were so specified as to generate data with 0%, 25%, 50%, and 75% censored observations. One thousand samples of 50 match sets were generated: for each fixed set of N=50 matching variable observations; for each number R, R=1, 2, 3, 4, 5, of match set observations from population 1; and for each censoring percentage. Tests were performed at the one percent level of significance on each of these one thousand samples, and the number of times that each test rejected H_0 was recorded.

In Subsection 4.2, we present a Monte Carlo comparison of the Mantel-Ciminera procedure and three logrank procedures of the form Q. We compared each set of four tests on data for which the statistic T' is efficient. That is, the logrank procedures were compared on simulated arbitrarily right censored exponential data; the Wilcoxon procedures were compared on simulated arbitrarily right censored log logistic data.

4.2. Logrank Procedures

Uncensored response times were exponentially distributed. We generated values, $\lambda_1, \lambda_2, \ldots, \lambda_{50}$, of a matching variable from a uniform distribution on (0,1). The true response from population 0 for M_i was then generated from an exponential distribution with mean $1/\lambda_i$. The R true responses from population 1 for M_i were generated from an exponential distribution with mean D/λ_i. The censoring variable for the population 0 response in M_i was generated from a uniform distribution on $(0,c/\lambda_i)$. The censoring variables for the R population 1 responses were generated from a uniform distribution on $(0,cD/\lambda_i)$. With this censoring mechanism, the probability, p, for an observation being censored, was $p=(1-e^{-c})/c$. We chose c so that p would be 0, .25, .50, and .75. The parameter D was specified as 1, $\frac{3}{2}$, 2, and 4. The null hypothesis is equivalent to D=1.

Since the Mantel-Ciminera procedure is a two-sided test, all tests in this subsection have been run as two-sided tests. The four tests compared on these data were--

LU: a test of the form Q with each T_i of the form T' with weights

$$c_j=\sum_{\upsilon=1}^{j} n_\upsilon^{-1}-1,\ C_j=\sum_{\upsilon=1}^{j} n_\upsilon^{-1},\ j=1,2,\ \dots,k,$$

and with $\hat{\sigma}^2=\hat{\sigma}_3^2$. The weights c_j and C_j, j=1,2, ...,k, are derived in Kalbfleisch and Prentice (Ref. 6), from Prentice (Ref. 13: eq. 17).

LB: a test of the form Q with each T_i of the form T" with $a_j=-(c_j-C_j)$, where c_j and C_j are as in LU, j=1,2, ...k, and with $\hat{\sigma}^2=\hat{\sigma}_2^2$.

LP: a test of the form Q with T_i of the form T' with weights c_j and C_j, j=1,2, ...k, as in LU and with $\hat{\sigma}^2=\hat{\sigma}_2^2$.

LM: the logrank procedure of Mantel and Ciminera (Ref. 11).

The results, in Table 1, show that LU, LB, and LP perform at about the same power, with LB having a slight advantage. The powers of these three tests are equal at R=1, because these three statistics are algebraically identical for matched pair data. The Mantel-Ciminera procedure, LM, tends to have the lowest power for all considered alternatives when p>.50. The power of all procedures decreases, as expected, with increasing censoring.

4.3 Wilcoxon Procedures

The logarithm of response time is logistically distributed, thus resulting in response time having the density:

$$g(t)=\lambda D^z(t+\lambda D^z)^{-2},\ t>0,\ \lambda>0,\ D>0$$

with z being the population indicator. We generated matching variable values, $\lambda_1, \lambda_2, \dots, \lambda_{50}$, from a uniform distribution on (1,11). For the true response from population 0 in M_i, we generated a variable from a distribution with the density g(t) with z=0 and D=1. The R true responses from population 1 were generated using g(t) with z=1 and the parameter D, having values 1, 1.5, 2, and 2.2. The null hypothesis is equivalent to D=1. The

TABLE 1. LOGRANK PROCEDURES

Entries are the number of times each of the four logrank tests, LU, LB, LP and LM, rejected the null hypothesis at the 1% level of significance in 1000 simulations of a sample of size 50 match sets, each of size R+1, with one observation from population 0 and R observations from population 1 per match set. The uncensored population 0 observation in the <u>ith</u> match set is exponential with mean $1\lambda_i$, the R uncensored population 1 observations in the <u>ith</u> matched set are exponential with mean D/λ_i, with D=1, 3/2, 2, and 4. The values λ_i are chosen uniformly from (0,1). The censoring variables for the <u>ith</u> match set are uniform on $(0, cD/\lambda_i)$, where c_i is chosen so that the proportion of censored observations is p;p=0, .25, .50, and .75.

	D=1				D=3/2				D=2				D=4			
	LU	LB	LP	LM	LU	LB	LP	LM	LU	LB	LP	LM	LU	LB	LP	LM
p=0																
R=1	.005	.005	.005	.010	.099	.099	.099	.184	.381	.381	.381	.615	.979	.979	.979	1.00
R=2	.013	.013	.013	.016	.228	.217	.165	.243	.693	.691	.613	.774	1.00	1.00	.999	1.00
R=3	.007	.007	.008	.007	.284	.269	.196	.219	.795	.774	.728	.801	1.00	1.00	1.00	1.00
R=4	.001	.004	.005	.009	.373	.326	.241	.302	.908	.904	.837	.871	1.00	1.00	1.00	1.00
R=5	.011	.008	.007	.010	.388	.347	.260	.270	.907	.901	.826	.881	1.00	1.00	1.00	1.00
p=.25																
R=1	.008	.008	.008	.011	.092	.092	.092	.095	.306	.306	.306	.324	.931	.931	.931	.953
R=2	.008	.006	.007	.009	.137	.142	.114	.111	.521	.530	.478	.443	.990	.991	.987	.979
R=3	.004	.006	.003	.011	.229	.238	.181	.151	.652	.667	.596	.470	1.00	1.00	.999	.987
R=4	.006	.008	.008	.006	.225	.225	.187	.138	.673	.717	.625	.533	1.00	1.00	.999	.994
R=5	.007	.007	.007	.009	.281	.290	.203	.131	.760	.767	.663	.515	1.00	1.00	1.00	.995
p=.5																
R=1	.014	.014	.014	.006	.069	.069	.069	.060	.216	.216	.216	.178	.741	.741	.741	.666
R=2	.007	.009	.010	.008	.097	.110	.096	.064	.315	.360	.324	.216	.900	.930	.910	.742
R=3	.006	.007	.006	.001	.106	.123	.094	.060	.354	.426	.361	.205	.954	.971	.951	.793
R=4	.000	.002	.003	.011	.135	.191	.138	.126	.417	.502	.418	.323	.942	.987	.954	.740
R=5	.006	.006	.008	.012	.166	.204	.147	.082	.504	.593	.458	.264	.958	.976	.954	.792
p=.75																
R=1	.013	.013	.013	.011	.030	.030	.030	.025	.083	.083	.083	.059	.353	.353	.353	.239
R=2	.013	.010	.009	.015	.035	.060	.056	.034	.093	.175	.147	.069	.468	.621	.578	.313
R=3	.004	.006	.004	.009	.033	.083	.065	.030	.117	.233	.197	.079	.518	.751	.690	.361
R=4	.006	.009	.009	.014	.052	.094	.077	.056	.149	.311	.236	.120	.525	.778	.682	.370
R=5	.006	.010	.008	.005	.034	.102	.071	.029	.131	.311	.212	.094	.573	.841	.719	.377

censoring variable for the population 0 response in M_i was generated from a uniform distribution on $(o, c\lambda_i)$. The censoring variables for the population 1 responses were generated from a uniform distribution on $(0, cD\lambda_i)$. With this censoring mechanism, the probability of obtaining a censored response was $p = c^{-1}\log(1+c)$.

The four tests compared on these data were--

WU: a test of the form Q with each T_i of the form T" with weights

$$c_j = 1 - 2\prod_{\upsilon=1}^{j} \frac{n_\upsilon}{n_\upsilon + 1}, \quad C_j = 1 - \prod_{\upsilon=1}^{j} \frac{n_\upsilon}{n_\upsilon + 1}, \quad j = 1, 2, \ldots, k,$$

and with $\hat{\sigma}^2 = \hat{\sigma}_3^2$. The weights c_j and C_j, $j=1,2, \ldots, k$, are derived in Kalbfleisch and Prentice (Ref. 6: p. 147), from Prentice (Ref. 13: eq. 17).

WB: a test of the form Q with each T_i of the form T" with $a_j = c_j - C_j$, $j=1,2, \ldots, k$, where c_j and C_j are as in WU, with $\hat{\sigma}^2 = \hat{\sigma}_2^2$.

WP: a test of the form Q with each T_i of the form T' with the same weights as in WU and with $\hat{\sigma}^2 = \hat{\sigma}_1^2$.

WG: a test of the form Q with each T_i being a Gehan (Ref. 5) extension of the Wilcoxon two-sample procedure and with $\hat{\sigma}^2 = \hat{\sigma}_3^2$.

Note that WU, WB, and WP are the Wilcoxon analogues of LU, LB, and LP. The test WG is a viable competitor for WU, WB, and WP, since WG is a Wilcoxon extension, not of the form T'--a fact which clearly satisfies the conditions of Theorem 3.1 and Lemma 3.1.

The results of this comparison, in Table 2, show much the same behavior as in Table 1. All four tests operate at about the same power, with WB tending to do slightly better. The tests WU and WG are very close to each other and, in fact, are algebraically identical for all values of R in the absense of censoring.

5. CONCLUSIONS

In both the logrank and Wilcoxon situations, all tests performed close to the asymptotic level of one percent. In both tables, the statistics of the form T" with the binomial variance estimates $\hat{\sigma}_2^2$, LB, and WB tend to be more powerful than the other tests. The only exception to this trend is in Table 1, without censorship where none of the four logrank tests is most powerful for all specified alternatives and for all values of R. All Wilcoxon tests and all logrank tests, except LM, appear to operate at the same power

TABLE 2. WILCOXON PROCEDURES

Entries are the number of times each of the four Wilcoxon tests, WU, WB, WP, and WG, rejected the null hypothesis at the 1% level of significance in 1000 simulations of a sample of size 50 match sets, each of size R+1, with one observation from population 0 and R observations from population 1 per matchset. The population 0 uncensored observation in the ith match set is log logistically distributed with parameters λ_i and Z=0 and D=1. The R population 1 uncensored observations in the ith match set are log logistic with parameters λ_i, Z=1 and D=1, 1.5, 2, and 2.2. The values λ_i are chosen uniformly from (1,11) on each simulation. The censoring variables for the ith match set are uniform on $(0,cD\lambda_i)$, where c is chosen to make the proportion of observations censored equal to p; p=0, .25, .50, and .75.

	D=1				D=1.5				D=2				D=2.2			
	WU	WB	WP	WG	WU	WB	WP	WG	WU	WB	WP	WG	WU	WB	WP	WG
p=0																
R=1	.003	.003	.003	.003	.028	.028	.028	.028	.117	.117	.117	.117	.164	.164	.164	.164
R=2	.011	.012	.016	.011	.070	.082	.082	.070	.241	.268	.268	.241	.334	.382	.382	.334
R=3	.007	.008	.009	.007	.076	.109	.079	.076	.302	.377	.330	.302	.432	.501	.445	.432
R=4	.012	.023	.016	.012	.107	.133	.117	.107	.368	.442	.405	.368	.468	.528	.492	.468
R=5	.015	.015	.014	.015	.111	.146	.110	.111	.368	.447	.360	.368	.506	.598	.525	.506
p=.25																
R=1	.006	.006	.006	.006	.044	.044	.044	.044	.150	.150	.150	.150	.202	.202	.202	.202
R=2	.014	.013	.013	.013	.063	.080	.067	.064	.255	.297	.282	.252	.322	.376	.353	.323
R=3	.008	.009	.009	.011	.095	.117	.106	.095	.291	.369	.324	.297	.361	.440	.389	.358
R=4	.014	.018	.016	.013	.110	.153	.135	.106	.310	.384	.345	.305	.412	.508	.456	.418
R=5	.101	.013	.008	.011	.088	.131	.104	.089	.346	.445	.393	.343	.472	.568	.501	.464
p=.5																
R=1	.008	.008	.008	.008	.045	.045	.045	.045	.118	.118	.118	.118	.165	.165	.165	.165
R=2	.011	.009	.009	.011	.039	.064	.055	.038	.173	.234	.216	.167	.247	.320	.298	.245
R=3	.011	.013	.013	.010	.040	.077	.065	.038	.227	.323	.294	.226	.251	.384	.331	.244
R=4	.011	.011	.012	.010	.054	.094	.076	.050	.279	.396	.343	.276	.264	.380	.334	.259
R=5	.010	.009	.008	.008	.071	.122	.101	.069	.242	.381	.322	.233	.350	.500	.429	.346
p=.75																
R=1	.005	.005	.005	.005	.018	.018	.018	.018	.063	.063	.063	.063	.098	.098	.098	.098
R=2	.009	.011	.011	.012	.023	.057	.054	.022	.068	.135	.124	.068	.109	.171	.161	.103
R=3	.007	.011	.008	.012	.025	.067	.063	.017	.078	.192	.175	.072	.110	.253	.230	.101
R=4	.016	.007	.009	.016	.018	.073	.061	.018	.086	.233	.198	.078	.135	.329	.289	.128
R=5	.015	.006	.007	.016	.019	.082	.065	.028	.083	.233	.201	.073	.108	.312	.271	.091

level until the censoring percentage is large; as the percentage of censoring increases, LB and WB tend to dominate the others. In Table 1, the first dramatic power dominance occurs for LB at p=.75. Data in Table 2 indicate that WB is also beginning a significant power dominance as p increases. The procedure WP is the second most powerful in Table 2, while LU is second most powerful in Table 1 in the presence of censoring.

We emphasize again that LB, WB, LP, and WP are standard procedures for two-sample stratified analyses applied to match set data, with each match set constituting a separate stratum. We have shown that the asymptotic normality of these procedures is justified even when, as occurs with match set sampling, the number of strata increase while the size of each stratum remains fixed.

The tests WU and LU are more generally applicable than the other procedures considered here; for their variance estimates are unbiased independently of the symmetry of the hypothesis H_0, whereas the unbiasedness of the variance estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ does depend on that symmetry. So, in cases in which $E(T')=0$ but the joint distribution of the true response times is not symmetric, the only applicable procedure (among those considered here) for testing--for example, the hypothesis of equality of one-dimensional marginal distriibutions--is to use $\hat{\sigma}_3^2$ as the variance estimate in the form Q.

In this report we have considered a one-to-R matched design in which all match sets were assumed to consist of R+1 observations (t,Δ). These procedures are easily extended to many-to-many matched designs in which the sizes of the match sets may vary.

REFERENCES

1. Ash, R. B. Real analysis and probability. New York: Academic Press, 1972.

2. Breslow, N. E. Odds ratio estimators when the data are sparse. Biometrika 68:73-84 (1982).

3. Breslow, N., and J. Crowley. A large sample study of the life table and product limit estimates under random censorship. Statistics 2:437-453 (1974).

4. Chung, K. L. A course in probability theory. New York: Academic Press, (1974).

5. Gehan, E. A. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52:203-223 (1965).

6. Kalbfleisch, J. D., and R. L. Prentice. The statistical analysis of failure time data. New York: John Wiley, 1980.

7. Knuth, D. E. The art of computer programming, vol. 2. Reading, Pa.: Addison Wesley, 1969.

8. Lathrop, G. D., P. M. Moynahan, R. A. Albanese, and W. H. Wolfe. Epidemiologic investigation of health effects in Air Force personnel following exposure to herbicides: Baseline questionnaires. SAM-TR-82-42, Nov 1982.

9. Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 50:163-170 (1966).

10. Mantel, N., N. R. Bohidar, and J. L. Ciminera. Mantel-Haenszel analysis of litter-matched time-to-tumor response data, with modifications for recovery of interlitter information. Cancer Res 37:3863-3868 (1977).

11. Mantel, N., and J. L. Ciminera. Use of logrank scores in the analysis of litter-matched data on time to tumor appearance. Cancer Res 39:4308-4315 (1979).

12. Mehrotra, K., J. Michalek, and D. Mihalko. A relationship between two forms of linear rank procedures for censored data. Biometrika 69:674-676 (1982).

13. Prentice, R. L. Linear rank tests with right censored data. Biometrika 65:167-179 (1978).

14. Prentice, R. L., and P. Marek. A qualitative discrepancy between censored data rank tests. Biometrics 35:861-867 (1979).

15. Schoenfeld, D. The asymptotic properties of nonparametric tests for comparing survival distributions. Biometrika 68:316-319 (1981).

16. Tarone, R., and J. Ware. On distribution-free tests for equality of survival distributions. Biometrika 64:156-160 (1977).

17. Wei, L. J. A generalized Gehan and Gilbert test for paired observations that are subject to arbitrary right censorship. J Am Statistic Assoc 75:634-637 (1980).

18. Woolson, R. F., and P. A. Lachenbruch. Rank tests for censored matched pairs. Biometrika 67:597-606 (1980).