EVALUATION OF SURVIVAL DATA AND TWO NEW RANK ORDER STATISTICS ARISING IN ITS CONSIDERATION[1]

Nathan Mantel[2]

SUMMARY

Survival-time patterns should be compared properly in their entirety rather than at isolated points only. Such overall comparison would require a value function for rating particular durations of survival, but no such function exists. A chi-square procedure is proposed for comparing two sets of life-table data in their entirety. The implicit value function for the procedure is reasonable in that it gives greater weight to earlier deaths. By considering the case in which the life-table intervals are arbitrarily short, it is seen to be essentially a rank order procedure. Further, the procedure is defined for general or individual right censorship, for left truncation, and for tied ranks. A second rank order procedure or statistic is developed by considering the case in which the first statistic is computed after each death in a comparative survival-time study. This second statistic is the maximum chi square computed over the course of the study. Such a statistic can be used pseudosequentially and significance or nonsignificance ascertained before the end of the study. It is suggested that the procedures may be used in the analysis of other than survival-time data. The possibility for evaluating the statistical power of the procedures for alternatives when there is a constant force of mortality ratio between the two survival-time distributions is indicated.

The results of a medical investigation frequently take the form of two sets of survival patterns for patients who have been subjected to two different therapeutic regimens. Analysis of such data, because of incomplete observation on some or perhaps most patients, is frequently by the life-table or actuarial method (1).

In comparing the two sets of data, one may ask if there is statistically significant evidence that the proportion of patients surviving some stated period, say 5 years, is greater for one of the two regimens. Or one may question if there is significant evidence that the time to which a stated proportion survive, say 50% corresponding to the median survival time, is longer for one of the regimens.

Let us consider a more advanced question that may be asked: Is there statistically significant evidence that one of the regimens is superior, as judged by the observed survival-time patterns? The answer must depend on how we define "superiority." We cannot identify differences in superiority between two survival patterns observed subject to chance variation without first being able to identify instances of differences in which the two survival patterns are known exactly and are free of variation.

As a starting point, we may reasonably say that if two exact survival patterns are such that for one the proportion surviving is sometimes greater, but never less, than for the other, it is the superior pattern. (It is assumed here that the improved survival is not accompanied by deleterious aspects such as blindness. If they were present any appropriate comparison would have to take this into

[1] Presented at Conference of Regional Statisticians in the Clinical Trials Program, National Cancer Inst, Public Health Service, May 13–14, 1965, Bethesda, Md.

[2] Biometry Branch, National Cancer Inst, Public Health Service.

account.) At whatever times the two patterns are compared, or to whatever proportion surviving, the superior pattern will never appear inferior.

It cannot however be anticipated that two survival patterns will generally stand in this superior-inferior relationship at all points. It may be that one therapeutic procedure yields a long-term survival advantage at the cost of some immediate increase in mortality—this is frequently true of surgical procedures. Alternatively, the cost of some advantageous therapy might lie in increased mortality in the more distant future. We must recognize then that two exact survival patterns may cross each other at one or more points. As a result, point comparisons between patterns of the kind generally made will depend on the point at which it is made.

Conceptually, one can make whole-pattern comparisons. What is required is a value or utility function defining the value of living to a certain age or for a certain period of time beyond therapy. Given the value function one can determine the average value for exact survival patterns. Observed survival patterns can be translated into observed average values and, with their reliabilities ascertained, the difference between two such averages can be assessed for statistical significance.

The difficulty with this approach lies in the probable lack of any general agreement on the value function to use. An economist might objectively determine the utility of surviving from any one age to any other age, perhaps doing so separately for each sex and race combination. With differing normative assumptions, the theologian and the philosopher would provide quite different value functions.

In the absence of an agreed on value function, such a concept for comparing survival patterns cannot be implemented. The need, however, for comparing survival patterns as a whole rather than at isolated points will be met by the procedures to be described. Although no specific value function for survival will be provided, whatever value function may be implicit in the procedures to be suggested will, in most actual instances, be fairly reasonable.

The Standard Life-Table Situation

The point of departure in this report will be the customary form in which survival data occur in medical investigations. The study begins with a population, presumed homogeneous, of N_{11} individuals who are followed progressively in time, with their survival status assessed periodically (but the intervals between the periodic assessments need not necessarily be equal).

Suppose that during the first study interval r_{11} individuals are known to die while l_{11} individuals are lost track of, leaving $N_{12} = N_{11} - r_{11} - l_{11}$ individuals available for study in the succeeding period. For purposes of simplicity, and as justified in the Appendix, *Discussion 1*, consider that all losses occur at the end of the interval. From the data the apparent probability of dying in the first interval is given by r_{11}/N_{11}, while the apparent survival probability is $(N_{11} - r_{11})/N_{11}$. Assuming no essential difference between individuals lost track of and those available for second interval study, an uncorrelated but not independent (see Appendix, *Discussions 2 and 3)* estimate of the conditional probability of surviving the second period is given correspondingly by $(N_{12} - r_{12})/N_{12}$. Successively, the estimated conditional probability of surviving the i'th interval, provided there has been survival through the i — 1st interval, is given by $(N_{1i} - r_{1i})/N_{1i}$. By serial multiplication of these conditional probabilities one obtains the estimated nonconditional probability of surviving from the beginning of the study through the end of the j'th period as $\prod_{1}^{j} (N_{1i} - r_{1i})/N_{1i}$, or the nonconditional probability of dying as $1 - \prod_{1}^{j} (N_{1i} - r_{1i})/N_{1i}$. Approximate standard errors can be obtained for such estimated death or survival probabilities, permitting statistical comparison with other similarly obtained estimates.

Suppose that we have, instead, two groups of individuals with respective sets of observed numbers dying among those at risk, in each interval, r_{1i} among N_{1i} and r_{2i} among N_{2i}. Such data can be represented by several 2×2 contingency tables, one for each study interval of the form

	Died	Survived	Total
Group I	r_{1i}	$N_{1i} - r_{1i}$	N_{1i}
Group II	r_{2i}	$N_{2i} - r_{2i}$	N_{2i}
Total	M_{1i}	M_{2i}	T_i

Using the procedure described by Mantel and Haenszel (2), one can compute a one degree of freedom (df) continuity-corrected chi square (see Appendix, *Discussion 4*) which permits simultaneous comparison over all the contingency tables of the differences in death or survival probabilities for the two groups. This is given by

$$\chi^2 = \frac{(|\Sigma r_{1i} - \Sigma E r_{1i}| - \frac{1}{2})^2}{\Sigma V r_{1i}} \qquad (1)$$

where the expectations and variance, computed conditionally on the separate contingency-table marginal totals, are given by

$$E r_{1i} = N_{1i} M_{1i} / T_i \qquad (2)$$

$$V r_{1i} = N_{1i} N_{2i} M_{1i} M_{2i} / T_i^2 (T_i - 1). \qquad (3)$$

Corresponding to the comparison of observed and expected deaths for the two groups, Mantel and Haenszel (2) also give various summary formulas for the relative odds of death or of survival in an interval. One, in particular, would be

$$R = \frac{\Sigma r_{1i}(N_{2i} - r_{2i})/T_i}{\Sigma r_{2i}(N_{1i} - r_{1i})/T_i} \qquad (4)$$

which can be considered to differ significantly from unity whenever chi square is significantly large.

By the chi-square formula shown, there is a comparison, in each time interval, of the number of deaths in the first study group with its expectation on the assumption of no difference between the study groups. Any general tendency for the death rates in the first group to be larger, or smaller, than those in the second group will be self-reinforcing by use of the formula with the result that nonsignificant differences in the individual time intervals, when consistent, can give rise to a highly significant summary chi square.

It is a characteristic of the summary chi-square formula that, when applied to the life-table situation, it gives more weight to deaths which occur in an early time period than to deaths which occur in a later time period. Thus, if 2 groups are subject to the same probability of surviving through the entire study period, the formula will, nevertheless, show excess mortality for the group in which deaths occurred earlier. (As a simple illustration, consider that two groups, of 100 each, both have 50 deaths. In the first group all deaths occur in the first of 2 study periods, while for the second group all deaths occur in the second period. By the formulas given, group 1 will show a total of 50 deaths compared with an expectation of 41.7, an excess of 20%.) It is in this sense that the procedure suggested has, as indicated, an implicit survival value function which is reasonable. (However, if there is a mixture of annual cohorts, as indicated in the Appendix, *Discussion 1*, there will already be a greater weighting for early deaths.) With continued observation in time of 2 study groups an increasingly significant difference in survival patterns may emerge if earlier patterns of difference continue. In contrast, if one considers only the final survival probabilities any possibility for a significant difference disappears as the survival proportions for both groups approach zero.

Extension of this analytic procedure to heterogeneous study groups is immediate. This is accomplished by setting up separate life tables for various age, race, sex, or other subgroupings as necessary, perhaps even by annual cohorts, and then carrying out the summations indicated in the preceding formulas over both the various subgroupings and the various study intervals. Extension may be made also to the case in which there are more than 2 study groups compared. Mantel and Haenszel (2) discuss the problem of computing chi square with more than 1 df for the case of more than two levels for one axis of classification and give specific formulas for the 2-df case. Mantel (3) provides 1-df procedures for arbitrarily many, but orderable, levels.

The Continuous Observation Situation and its Corresponding Rank Order Statistic

Suppose that in the life-table situation we allow the periods of observation to become arbitrarily short, and arbitrarily numerous. In equation 1 (col. 1) there will be a contribution to the numerator and to the denominator of chi square only for periods in which deaths occur, so that only such periods need be considered. Intervals in which there are no deaths or only losses to follow-up need not be considered, and with sufficiently short-time periods there will be one period per death which will require consideration.

In equations 1–3 (col. 1), corresponding to each death, r_{1i} will be 1 or 0 according to

whether the death is from group 1 or group 2; $Er_{1i} = N_{1i}/T_i$ and $Vr_{1i} = N_{1i}N_{2i}/T_i^2 = Er_{1i}(1 - Er_{1i})$, where N_{1i} and N_{2i} are the number of survivors in the two groups respectively immediately prior to the death in question and $T_i = N_{1i} + N_{2i}$.

In the instance of application of the chi-square formula to continuous observation, it follows that chi square depends only on the ordering in time at which deaths and losses to follow-up occur in the 2 groups and is not affected by the actual times at which such deaths occur. As such it can be recognized to be a rank order statistic.

As a rank order statistic, it can be applied also in the analysis of data in general, and not solely for the analysis of survival data. While it is defined when all individual values are known, the computation of the rank order statistic is also appropriate even though any or all three of the following may occur:

1. There is a general point of right-censored observation that corresponds to the point of termination of observation in a life-table study.

2. There are a variety of points of right-censored observation that correspond to the individual times of loss to follow-up in a life-table study.

3. There are arbitrarily many tied ranks, which is essentially the case when, in a life-table study, one uses wide enough intervals to permit multiple deaths in the same interval.

With these properties, the statistic proposed can be used as an alternative rank order procedure in certain problem situations. For example, it could be used alternatively to the procedure proposed by Halperin (4) for taking into account both the difference in average rank below the point of general censorship, and the difference in the proportion of observations above the censorship point. More recently Weiler (5) has considered how to combine both quantal and quantitative responses, though not by ranking procedures. In the examples Weiler gives, however, the quantal response can be considered to be an extreme quantitative response, either poorer than the poorest quantitative response, or better than the best. This would suggest that the situation can be handled as a censored data problem, though Weiler uses an alternative procedure.

The various extensions cited in the preceding section for the standard life table can also be made for the continuous observation situation. Some question may be raised as to the continuous observation since the individual expectations will ordinarily be small. Mantel and Haenszel (2), however, assert that the process of summation involved in the procedures they give can justify the use of chi square, albeit the individual contributions to the total are small. It may be, however, that an alternative likelihood procedure can be obtained.

In some instances it may be desirable to compute the chi-square statistic beginning at some point later than the start of the study. This can be used to minimize the effect of postoperative deaths when surgical procedures are used, or for other reasons.

For that matter, the methodology can be used to incorporate arbitrarily and individually left-truncated data, as well as the right-censored data already indicated. Suppose the records for the first 2 years of a group are lost. Survival in the third year and subsequent years can nevertheless be evaluated and the data for such later periods combined with fuller data from other groups. Thus the data for any individual or individuals can be introduced at any point into a life-table type of analysis without concern as to the size of the initial group of whch these individuals are a remnant.

A Second Rank Order Statistic

Suppose a comparative trial of two alternative therapies is conducted and, for purposes of describing the alternative rank order statistic, under conditions in which there are no losses to follow-up. With the procedures described, a statistical comparison can be made at some predesignated point in time or when a predesignated number of deaths occur, which may equal $N_1 + N_2$, the total number of individuals in the study.

During the course of the study the constant accrual of the data would suggest the usefulness of making statistical comparisons before the actual completion of the study. As each item of data is obtained a statistical test is done. But for mathematical validity, the procedures already described are not satisfactory. These procedures will have the nominal error probabilities only if employed once rather than repeatedly.

A pseudosequential procedure can permit this repeated testing while maintaining error probabilities. Such a procedure permits deter-

ming, before all the data are in, whether or not the final statistic based on all the data will prove significant. The pseudosequential procedure differs from the true sequential in that the sample size does not vary but is fixed in advance. But there is sequential testing in time. In a survival-time study this can be a great advantage. Since the data for the last few deaths may trickle in rather slowly, it is likely that a pseudosequential decision can be made long before the study would be otherwise completed. In the Appendix, *Discussion 5* the Kolmogorov-Smirnov 2-sample statistic (6, 7) is used to exemplify this principle.

In a comparative trial the pseudosequential statistic can be devised in the following manner: Consider that after each death the chi-square statistic is computed, as indicated, and as if the study had just terminated. At some point in the study this chi square will attain a maximum, and it is this maximum chi square which is the suggested test statistic. The observed maximum chi square over the course of the study must be judged on the basis of the distribution of such maxima. However, if a critically large maximum chi square is attained before the termination of the study, statistical significance can be claimed immediately. And at any point where significance has not yet been obtained, it is possible to determine whether there is a chance that a significant maximum chi square may yet arise. This is done by computing the terminal chi square which would arise if the remaining deaths were, first, all from one study group, and then all from the other group.

The distribution of the maximum chi square can, in principle, be obtained combinatorially. Under the null hypothesis of no difference between the groups there is a total of $(N_1 + N_2)!/N_1!N_2!$ equally probable orderings in which the deaths may occur. For each such ordering a maximum chi square must be computed, which permits determining its distribution over all the possible orderings. The use of this statistic may have to await the computation of such distributions. Monte Carlo procedures, however, may permit determining relatively accurate empirical distributions of maximum chi square without the need for complete enumeration.

Although the two rank order statistics considered, the terminal chi square and maximum chi square, have been envisaged in connection with a survival-time study, the possibility for application to other kinds of observations is not precluded. Given any two sets of observations, they can be ordered and both terminal and maximum chi square computed for testing. There is perhaps some special appropriateness of the statistics for life data in that both terminal and maximum chi-square values are computed by ascending from smaller to larger values. In some other context a reverse ordering, giving rise to different chi-square values, would be appropriate. In general, other rank order statistics, not specifically appropriate to the life data, yield the same level of significance whether increasing or decreasing ranks are assigned.

Power Considerations

The procedures described are essentially sensitive to how instantaneous death rates or probabilities of death in an interval compare for the two survival-time distributions. The force-of-mortality function for a survival-time distribution is given by $Z(t) = f(t)dt/(1 - F(t))$, where $f(t)$ is the distribution density at t, $F(t)$ is the proportion of the population dying before t, and $1 - F(t)$ is the proportion surviving beyond t.

It is unlikely that in any real instance in which the two force-of-mortality functions, $Z_1(t)$ and $Z_2(t)$, differ that any simple relationship exists between them. In principle, however, it is possible to determine the power of the procedures given for alternatives which are expressible in the form $Z_1(t) = kZ_2(t)$, $k \neq 1$. This can be accomplished without knowledge of the actual forms of the distribution functions, since the probability of any particular permutation of deaths will depend only on k, and the value of the test statistic depends only on the permutation observed.

To illustrate, consider that in a study with 4 subjects in group 1 and 6 subjects in group 2, the sequence of deaths is

1,1,2,2,1,2,1,2,2,2.

The probability of this permutation is given by

$$\frac{4k}{4k+6} \times \frac{3k}{3k+6} \times \frac{6}{2k+6} \times \frac{5}{2k+5} \times \frac{2k}{2k+4} \times \frac{4}{k+4} \times \frac{k}{k+3} \times \frac{3}{3} \times \frac{2}{2} \times \frac{1}{1}$$

in which the denominators in each factor reduce by k with each prior group-1 death and by unity with each prior group-2 death. The product of the factor numerators simplifies to $4!6!k^4 = N_1!N_2!k^{N_1}$. For $k = 1$, the probability for the permutation reduces to $4!6!/10! = N_1!N_2!/(N_1 + N_2)!$, the reciprocal of the number of equally probable permutations.

Since it is feasible to compute the probability of all possible permutations under any alternative k value, it is possible to determine the alternative distribution of the chi-square statistics. In turn one can evaluate the probability that with a particular k value a statistically significant chi-square or maximum chi-square statistic will arise.

The concept of a constant ratio k between two force-of-mortality functions can be shown graphically. Consider a graph of the relationship between log $(1 - F_2(t))$ and t. Suppose we make a continuous distortion of the time scale, $h(t)$, with the result that log $(1 - F_2(t))$ plots linearly on the distorted scale. It will then be true that with the distorted scale log $(1 - F_1(t))$ will also plot linearly and with a negative slope k times as great.

This approach for determining the power of the statistical procedures proposed can be used more generally to calculate the power of other rank order methods for alternatives of the kind indicated. In this connection, Lehmann (8) has proposed somewhat similar alternatives and has recommended their usefulness for calculating the power of various nonparametric procedures.

APPENDIX: *Discussion 1*

Losses are ordinarily of two kinds. There is the true loss to follow-up in which the individual is lost track of and this occurs only to a limited extent. The second and more important kind is the administrative loss which arises artificially from the study procedure employed. To evaluate survival of a treatment group it is customary to combine data for several annual cohorts. Suppose there are 10 such cohorts, one each for 10 successive years. In the analysis, all 10 groups are combined for evaluating survival in the 1st year after treatment; only data for the first 9 groups are available for evaluating 2d-year survival; only data for the first 8 groups are available for evaluating 3d-year survival, and so forth. Take an individual who entered the study 3 years and 4 months before the date of the analysis: if he were alive at the time of the analysis he would be treated as an administrative loss during the 4th year and, with most methods of analysis, would be considered as having contributed a half year of exposure in the 4th year.

The following suggestion is made to minimize the complications of such administrative losses: Only full years (or full-time periods) of possible exposure for an individual should be counted. Thus, for the 3-year, 4-month person, his survival in the first 3 years would be counted, but his experience in the 4th year, or its first 4 months, would be disregarded, whether he survived or died. Only a small amount of information would be lost by this approach, the last fractional year for each person. And the information can be recovered in two ways. On reevaluation 8 or more months later, the full 4th year's data would be used. Or, if the interval of follow-up was reduced to a month, only the last fractional month's information would be lost. With the device suggested, all administrative losses would occur at the end of the study intervals.

APPENDIX: *Discussion 2*

The dependence of the successive estimates of the conditional survival or death probabilities results only from the number of cases involved. Whether there is low or high mortality in the 1st year, on the average the proportion of 1st-year survivors dying in the 2d year will remain the same. But with low mortality in the 1st year, there will be more cases on which to base 2d-year mortality, permitting a more reliable estimate to be made. With high 1st-year mortality, the reliability for the 2d year will be low. (It is true that by more diligent care one might keep people alive into the 2d year who would otherwise have died in the 1st year, thus seeming to produce a negative correlation. The negative correlation would apply only if we considered diligence of care a random variable. If we take the estimates as corresponding to the care given, the negative correlation disappears.)

APPENDIX: *Discussion 3*

The assumption that individuals lost during the course of a study are no different, except

for chance, from those remaining for further follow-up underlies use of the life-table procedure. In each instance it is important to consider whether the assumption holds. Thus in the combined cohort approach described in *Discussion 1*, the administrative losses come from a different annual cohort than those followed in the succeeding year. But in combining cohorts in the first place, we are making the assumption that cohorts show essentially the same survival, and such an assumption can be open to question. If there are true losses to follow-up, it is necessary to consider whether such losses show some kind of selection effect.

Use of the life-table procedure to evaluate the carcinogenic effects of drugs may illustrate more pointedly the nature of the assumption made. It is generally the lethal effect of the drug tested which obscures its carcinogenic effect. Thus, if a test animal should die without a tumor, it cannot be known when, or if, the animal would subsequently have developed a tumor. Treating deaths without tumor as losses to observation, the life-table procedure is used to estimate the probabilities of tumor incidence over the period of observation. The question then arises whether animals dying of lethal toxic effects are similar to those surviving with respect to the incidence of drug-induced tumors. On consideration, it is a very real possibility that they are not. Animals that died may have received effectively higher dosages, whether through actual minor variation in dose, less rapid elimination of the drug, or to a greater dose administered relative to some metabolic function of the animal. Alternatively, these animals may have received effectively lower dosages, the reduction resulting from more of the administered dose concentrating at sites of toxic activity, less at sites of carcinogenic activity. For reasons such as these, the results of life-table computations must be interpreted with caution and taken only as first attempts at correcting for the obscuring effects of toxicity, rather than as completely valid procedures for such correction. The procedure will, at least, preclude coming to the conclusion that a drug has little carcinogenic potential when it has been highly and rapidly lethal.

APPENDIX: *Discussion 4*

Theoretical justification of this combination chi-square approach can be seen from the consideration of certain orthogonal variables. The results of a comparative study, without losses to follow-up, can be viewed as giving rise to a $2 \times (k + 1)$ contingency table. The entries in the i'th column, $i = 1, \text{---}, k$, are r_{1i} and r_{2i}, the respective numbers of deaths in the i'th interval. The $k + 1$'st column consists of s_1 and s_2, the respective k interval survivors. All marginal totals of this $2 \times (k + 1)$ contingency table will now be considered fixed.

The following orthogonal variables, with determinable variances, can now be constructed.

$$Y_1 = r_{11} - E(r_{11})$$
$$Y_2 = r_{12} - E(r_{12}) \mid r_{11}$$
$$Y_3 = r_{13} - E(r_{13}) \mid r_{11}, r_{12}$$
$$\vdots$$
$$Y_k = r_{1k} - E(r_{1k}) \mid r_{11}, r_{12}, \text{---}, r_{1,k-1}$$

This formulation permits computing a k df chi square, $\Sigma Y_i^2/\text{Var } Y_i$. For power purposes, however, we compute alternatively a 1 df chi square as $(\Sigma Y_i)^2/\Sigma \text{ Var } Y_i$.

But it can be readily seen, or shown, that

$$Y_i = r_{1i} - E(r_{1i}) \mid r_{11}, r_{12}, \text{---}, r_{1,i-1}$$
$$= r_{1i} - E(r_{1i}) \mid N_{1i}, N_{2i}, M_{1i}, M_{2i}$$

Thus, except for the continuity correction, equation (1) yields the same chi square as would consideration of the orthogonal Y variables. Losses to follow-up are handled by including conditionality on all such previous losses, which serve only to increase the number of columns in our contingency table.

APPENDIX: *Discussion 5*

The Kolmogorov-Smirnov 2-sample statistic (6, 7) is used to test the difference between 2 empirical cumulative distributions. The test statistic, which can be evaluated nonparametrically, is the maximum spread in probability between the 2 empirical cumulative distributions when comparison is made at all possible points. Suppose that in a comparative trial as considered in the text, it is planned to use the Kolmogorov-Smirnov test statistic. It will then be true that if a large enough difference in the empirical survival-time distribution is observed during the course of the study, the Kolmogorov-Smirnov statistic at the end of the study will necessarily be significant. Conversely, if the difference has not yet attained significance

and remains small as the study approaches completion, it is possible to determine at some point that the final Kolmogorov-Smirnov statistic cannot be significant. Thus the statistic can be used pseudosequentially.

For the purposes indicated in this report, however, the Kolmogorov-Smirnov statistic is unsatisfactory. It is sensitive to the difference in 2 cumulative distributions at single points in time, rather than to the difference in survival patterns. It is also subject to certain weaknesses in power. In a large study, clear and substantial differences at the tail of 2 cumulative distributions may fail to be significant because equally large differences can arise by chance in the central area. Similar weaknesses arise in the Kolmogorov-Smirnov 1-sample statistic, which is used to compare a single empirical distribution with a hypothetical distribution. Anderson and Darling (9) have considered how such statistics can be modified and strengthened.

REFERENCES

1. SPIEGELMAN, M. The versatility of the life table. Amer J Public Health 47:297–304, 1957.
2. MANTEL, N., and HAENSZEL, W. Statistical aspects of the analysis of data from retrospective studies of disease. J Nat Cancer Inst 22:719–748, 1959.
3. MANTEL, N. Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. J Amer Statist Ass 58:690–700, 1963.
4. HALPERIN, M. Extension of the Wilcoxon-Mann-Whitney test to samples censored at the same fixed point. J Amer Statist Ass 55:125–138, 1960.
5. WEILER, H. A significance test for simultaneous quantal and quantitative responses. Technometrics 6:273–285, 1964.
6. KOLMOGOROV, A. N. Confidence limits for an unknown distribution function. Ann Math Statist 12:461–463, 1941.
7. SMIRNOV, N. On the estimation of the discrepancy between empirical curves of the distribution for two independent samples. Bull Math Univ Moscou 1939, 2(fasc. 2).
8. LEHMANN, E. L. The power of rank tests. Ann Math Statist 24:23–43, 1953.
9. ANDERSON, T. W., and DARLING, D. A. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Ann Math Statist 23:193–212, 1952.