Use of Logrank Scores in the Analysis of Litter-matched Data on Time to Tumor Appearance

Nathan Mantel and Joseph L. Ciminera

Abstract

Previously, Mantel, Bohidar, and Ciminera (Mantel-Haenszel analyses of litter-matched time-to-response data, with modifications for recovery of interlitter information. Cancer Res., 37: 3863–3868, 1977) described how the Mantel-Haenszel method could be applied to litter-matched time-to-tumor data. Each litter would be treated as a separate stratum but only for as long as there remained contrastingly treated animals in that litter. Various devices were used for recovering interlitter information from remaining animals when that condition no longer obtained.

An alternative procedure for the assessment of such data is now provided. Data from all litters are merged, and the merged data are used to assign a score to each treated or control animal, such score depending on when or whether the animal developed a tumor. The set of scores for the members of each litter now define a finite population, and those finite populations are then considered in the analysis made.

Correspondences are brought out between the Mantel-Haenszel procedure, the logrank scoring procedure of Peto and Peto, and Savage’s use of expectations of order statistics for the exponential distribution.

Where incoming data are continuously analyzed, the logrank scores have to be changed, a difficulty that does not arise with the Mantel-Bohidar-Ciminera analyses. As in the Mantel-Bohidar-Ciminera report, it is recommended that results for tumors discovered only through autopsy of dying animals be integrated with results for observable tumors in a manner suggested by Peto.

The scoring procedure used has implications for the interpretation of individual animal results. For a study of limited duration, i.e., limited to a time at which the cumulative tumor rate is low, the primary distinction in scores is as between animals developing tumor and those not developing tumor. However, for a study in which the cumulative tumor rate is high, an animal with a late-appearing tumor could be assigned a score which is not suggestive of stimulated tumor formation. This is contrary to the frequent practice of counting all tumors, however late in appearance, as indicative of the tumorigenic effect of an agent.

Introduction

Recently, Mantel, Bohidar, and Ciminera (10) showed how the Mantel-Haenszel procedure (12) as applied to time-to-response data (7, 8) could be extended to studies involving litter matching in which interest was centered on the times of tumor appearance. In this work, if a test animal died, was sacrificed, or was otherwise lost before it displayed a tumor, its tumor appearance time was considered as censored; such losses were treated as independent competing risks, although in the circumstances such treatment might be biologically questionable. Each litter provided a paired cohort of animals for which the experience of animals receiving a specified treatment could be compared with that of their littermates receiving an alternative or control treatment.

Because of limited litter size, the litter cohort sizes would also be small, and in the illustration given by Mantel, Bohidar, and Ciminera, litters of Size 3 of a given sex were allocated 1 to treatment and 2 to controls (another 2 were allocated to other levels of treatment). The Mantel-Haenszel approach could be applied to the litter-matched data only as long as the remaining animals in a litter comprised both animals receiving the specified treatment and littermates which received the alternative treatment. To avoid the loss in information which would result if data for the remnant animals in a litter (remnant animals are those for which there are no contrastingly treated littermates) were disregarded, Mantel, Bohidar, and Ciminera suggested combining these remnants into cohorts to which the Mantel-Haenszel method could be applied. Alternative ways in which remnant cohorts could be formed so as to permit recovery of interlitter information were provided.

In this communication, we will present an alternative but related approach for taking litter membership into account in analyzing time-to-response data. No problem of how to form remnant cohorts arises with this method which, presumably, is less subject to loss of information. Basic to this alternative method, test animals from all the litters are combined into a single large paired cohort comprised of all animals receiving the specified treatment versus all animals receiving the alternative treatment. The paired cohort is followed in time not for the purpose of contrasting the treatments, but rather to assign a response score for each of the animals involved. The response scores are then collected for each of the animals in a litter, with the scores of animals on the specified treatment contrasted with those of littermates on the alternative treatment. Taking litter membership into account, we can then compare the total of such scores with its expectation, on the assumption that scores for a litter are fixed, but assignment to treatment is random; the permutational variance comes into play in testing the statistical significance of the deviation of the total from its expectation.

As prelude to presenting our alternative approach, we will review the Mantel-Haenszel approach and its relationship, when applied to time-to-response data, to the work of others.

1 This work was partially supported by USPHS Grant CA-15686 from the National Cancer Institute.

2 To whom requests for reprints should be addressed, at Biostatistics Center, George Washington University, 7979 Old Georgetown Road, Bethesda, Md. 20014.

Received January 12, 1979; accepted August 13, 1979.
The Mantel-Haenszel Approach for Time-to-Response Data and Certain Relationships

Mantel-Haenszel (12) originally presented their method primarily for combining independent 2 × 2 tables. If the first-row entries of the ith table are A, B, and second-row entries are C, D, then conditional on A + B = N1, C + D = N2, A + C = M1, B + D = M2, A, B, + C, D = T, they gave the null hypothesis expectation of A, as Ni, M1/T, and its variance as Ni, N2, M1, M2/T2(T — 1). A single d.f. continuity-corrected chi-square test of the departure of ΣA, from its expectation, ΣE(A), was then suggested, and an overall average measure of association, (ΣA(D/T)/ΣB(C/T)), was presented.

In the report of Mantel (7), the possibility of applying this approach to survival data was suggested and then followed up in a later report of Mantel (8). If a paired cohort were followed in time, the sequence of outcomes over several time intervals would take the form of successive 2 × 2 tables. Under conditional independence of these tables, and with care taken to avoid problems of losses to observation in midterm, the Mantel-Haenszel approach would be reasonably applicable. Mantel noted that, if the width of the time intervals became arbitrarily small, only those infinitesimal intervals in which a response occurred would come into play; although in principle there should be no tied responses, any ties occurring could be routinely handled. It was immediate that in situations not involving losses to observation, the Mantel-Haenszel procedure would become a ranking procedure.

In fact, the Mantel-Haenszel method applied as a ranking method to data not involving losses to observation was essentially anticipated by Savage (Ref. 17; see also the paper of Cox (2)). Savage’s concept was to replace each order statistic among the combined N1, N2 = T observations by the expected value of the order statistics from the standard exponential distribution. Thus, the earliest response time would be assigned the score 1/T, the next 1/T + 1/(T — 1), and in general the jth, (1/(T — r + 1). The set of N1, N2 scores would define a finite population with mean of unity and determinable variance, say σ2. Asymptotically, the deviation of the total of the N1 Savage scores from its expectation of N1 can be tested as a normal deviate with finite population variance N1N2σ2/(T — 1).

It is demonstrable that the sum of the N1 Savage scores is identical to what in the Mantel-Haenszel procedure would be ΣE(A). Consider the 2 × 2 table at the jth response; it has a first column total of 1, a second column total of T — j, with row totals, say, of N1, N2. By the Mantel-Haenszel approach, the contribution of this table to ΣE(A) would be N1j/(T — j + 1), but by the Savage score procedure, each of the N1 animals at risk for this infinitesimal interval would have its Savage score augmented by 1/(T — j + 1) for a total augmentation of N1j/(T — j + 1). We obtain the same result whether we score over individual animals and cumulate or cumulate over intervals. In any case, if all animals are observed to response, ΣA, by the Mantel-Haenszel method, would be N1 so that ΣA, — ΣE(A) = N1 — Σ(Savage scores), or the negative of the deviation of the total of Savage scores from its expectation.

On the matter of variances, correspondence is not perfect. The Savage variance correctly depends only on the values of N1 and N2, but the Mantel-Haenszel variance, which is borrowed from a different milieu, depends also on the sample sequence of outcomes. However, it is demonstrable and left as an exercise that the average Mantel-Haenszel variance under the null hypothesis is identically equal to the fixed Savage variance. Use of the Mantel-Haenszel variance would seem reasonable in asymptotic situations.

While use of the continuity correction of 0.5 is recommended for the Mantel-Haenszel procedure in its original setting, this would seem questionable in the time-to-response setting. The total Savage score, and thus the Mantel-Haenszel ΣE(A), may have cumulation points, but these could be so finely and irregularly spaced as not to warrant correction for continuity. Where order statistics are replaced by the expected value of order statistics from the uniform distribution, as in rank-sum procedures, the cumulation points are equally spaced, making essential the use of a continuity correction. Direct use of the ranks, 1, 2, ..., T, is equivalent to the use of expected order statistics from the uniform 0 — 1, or a different uniform for each value of T. To keep the underlying distribution the same, say uniform 0 — 1, the rankings should be divided by T + 1.

In Ref. 17, van Elteren has shown optimality for just such division by T + 1 in combining rank sum statistics, without reference to the need for keeping a constant underlying distribution. Presumably, the same kind of division should apply when combining results using the Wilcoxon test generalizations of Gehan [Refs. 4 and 5; see also Mantel (9)]. A somewhat similar adjustment to van Elteren’s, division by T, was suggested by Mantel in Ref. 7 when combining ranking results. Note that, in combining such adjusted rank sum results, the need for continuity correction may disappear if the T values differ.

A simple generalization of Savage’s procedure was given by Basu in Ref. 1 to cover the case of a single point of right censorship. The procedure is essentially the same as that of Savage but with the modification that each censored observation be assigned a score unity higher than the Savage score for the latest actual time-to-response score. For this censoring situation, it will also be true that the Mantel-Haenszel ΣA, — ΣE(A) will be the negative of the deviation of the total Basu score from its null expectation. In the one case ΣA, is reduced by the number of censored observations among the N1, while in the other that same number of censored observations has had its Savage score raised by unity.

An interesting concept can be seen in connection with Basu’s modification, that of the expectation of order statistics for the exponential when subject to censoring. The expected remaining lifetime from any censorship point is unity for the exponential and, reasonably, what Basu is doing is adding unity to the cumulated score of all surviving individuals at the single fixed time of censorship. But suppose losses to observation had occurred at different times during the course of the study; it would be as reasonable to add unity to some cumulative score at that moment in time, presumably the score for the last previous animal observed to respond. Barring the possibility of tied response times, the expected values of exponential order statistics under right censoring would be for responders the sum of the reciprocals of the numbers at risk at the present and all preceding response times and for nonresponders unity plus the sum of all such prior reciprocals.

What the Mantel-Haenszel procedure is doing can now be properly recognized. It is in fact assigning just such scores of expected exponential order statistics subject to arbitrary right censoring. It accomplishes this correctly even in complex sit-
N. Mantel and J. L. Ciminera

...functions only of regressor factors or covariates. Essentially, certain assumptions, that part of the likelihood related to time on just which individuals were at risk at that moment; under...  

...he considered. Thus, what characterizes Cox’s approach is that he expressed the likelihood of a...  

...simplified and non-homogeneous setting. This is the Cox approach, and it is the method we will follow. However, the question of how to handle ties remains. Cox’s approach was not necessarily intended to follow a rigid parametric form, and where relaxation is permitted it becomes equivalent to the method suggested by Mantel.

Illustration of the Logrank Scores Method

Table 1 gives the scheme for computing our logrank scores for the data of Table 2, the nature of which will be explained further below. Column 1 lists the distinct weeks in which tumors appeared, while Column 2 shows the number of rats at risk at the beginning of each of these weeks. Where several rats appeared with tumors in the same week, namely, Weeks 73, 80, 81, etc., extra lines are left for them in the table, with the numbers at risk then shown in braces and decreasing in steps of unity for each additional tumor. Columns 3 and 4 show, respectively, the number of tumors in the week and the number of interim losses until the next tumor-appearance week. To resolve ambiguities, interim losses are defined as those occurring at the end of weeks; thus, such losses between Weeks 73 and 77 are the number of deaths without tumor occurring for Weeks 73, 74, 75, or 76, in this case 7 deaths. The at-risk number of Column 2 for a given tumor-appearance week is obtained by subtracting from the preceding Column 2 value the total of the entries in Columns 3 and 4, e.g., for Week 77, 119 = 128 - 2 - 7 (note that the 127 in parentheses in Column 2 does not enter into this calculation). In some circumstances, animals dying without apparent tumor will be found on autopsy to have tumors. Peto (14) has raised this question and Mantel, Bohidar, and Ciminera have implemented Peto’s suggestion for handling this. The issue will be raised again under ‘Discussion.’

Column 5 shows the reciprocals of the Column 2 quantities, while Column 6 is a cumulation of the Column 5 quantities. Where a single tumor occurs in a week, the score for the tumorous rat is the Column 6 value, while the scores for any interim losses are given in Column 7 and are unity greater than the last Column 6 quantity for that week (note that the 127 in parentheses in Column 2 does not enter into this calculation). In some circumstances, animals dying without apparent tumor will be found on autopsy to have tumors. Peto (14) has raised this question and Mantel, Bohidar, and Ciminera have implemented Peto’s suggestion for handling this. The issue will be raised again under ‘Discussion.’

Column 5 shows the reciprocals of the Column 2 quantities, while Column 6 is a cumulation of the Column 5 quantities. Where a single tumor occurs in a week, the score for the tumorous rat is the Column 6 value, while the scores for any interim losses are given in Column 7 and are unity greater than the Column 6 values. For multiple tumors in a week, Column 8 shows the scores for each of the multiple tumors as the average of the Column 6 scores for each of the tumors involved. In these multiple-tumor instances, the score for interim losses is shown in Column 7 as unity greater than the last Column 6 value for the multiple-tumor week.

In Table 2, we report the data used for the computation of scores in Table 1, along with the associated scores (rounded to 3 decimal places) to which they gave rise. These data relate only to female rats and are repeated from the Mantel-Bohidar-Ciminera publication. Fifty paired cohorts each consist of a single drug-treated rat and its 2 untreated littersmates. The outcome for each rat is either its week of tumor appearance, signified by T, or its week of death without tumor, signified by D.
Logrank Scores for Analyzing Time-to-Tumor Data

Table 1
Illustration of computation of scores for pooled data combining treated and control rats from all litters

Data entered for weeks in which tumors actually occur include numbers at risk at beginning of week, number of tumors found during week, and losses to observation as from competing risks (death) in the interim to the next tumor-appearance week. (Losses in current week are treated as occurring at the end of the week.) Columns 6 and 8 show the score assignment for tumorous rats, Column 6 applying when a single rat develops tumor during a week, Column 8 applying for each of several rats developing tumor during a week. Column 7 shows the score for rats lost or dying without tumor during the week or during any weeks prior to the next tumor-appearance week.

<table>
<thead>
<tr>
<th>Wk of tumor appearance (1)</th>
<th>At risk, beginning of wk (2)</th>
<th>Tumors during wk (3)</th>
<th>Reciprocals of risk nos. (4)</th>
<th>Cumulation of reciprocals (score for single tumor in wk) (5)</th>
<th>Last cumulation for wk plus unity (score for interim losses) (6)</th>
<th>Av. cumulation for multiple tumors in a wk (score for each tumor where there are tied wk of appearance) (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>150</td>
<td>1</td>
<td>0</td>
<td>0.006667</td>
<td>0.006667</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>149</td>
<td>1</td>
<td>0</td>
<td>0.006711</td>
<td>0.013378</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>148</td>
<td>1</td>
<td>0</td>
<td>0.006757</td>
<td>0.020135</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>147</td>
<td>1</td>
<td>1</td>
<td>0.006803</td>
<td>0.026938</td>
<td>1.026938</td>
</tr>
<tr>
<td>49</td>
<td>145</td>
<td>1</td>
<td>1</td>
<td>0.006887</td>
<td>0.033835</td>
<td>1.033835</td>
</tr>
<tr>
<td>50</td>
<td>143</td>
<td>1</td>
<td>0</td>
<td>0.006993</td>
<td>0.040828</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>142</td>
<td>1</td>
<td>1</td>
<td>0.007042</td>
<td>0.047870</td>
<td>1.047870</td>
</tr>
<tr>
<td>55</td>
<td>141</td>
<td>1</td>
<td>2</td>
<td>0.007092</td>
<td>0.054962</td>
<td>1.054962</td>
</tr>
<tr>
<td>64</td>
<td>138</td>
<td>1</td>
<td>0</td>
<td>0.007246</td>
<td>0.062208</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>137</td>
<td>1</td>
<td>0</td>
<td>0.007299</td>
<td>0.069507</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>136</td>
<td>1</td>
<td>0</td>
<td>0.007353</td>
<td>0.076660</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>135</td>
<td>1</td>
<td>2</td>
<td>0.007407</td>
<td>0.084267</td>
<td>1.084267</td>
</tr>
<tr>
<td>70</td>
<td>132</td>
<td>1</td>
<td>1</td>
<td>0.007576</td>
<td>0.091843</td>
<td>1.091843</td>
</tr>
<tr>
<td>72</td>
<td>130</td>
<td>1</td>
<td>1</td>
<td>0.007692</td>
<td>0.099535</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>128</td>
<td>2</td>
<td>7</td>
<td>0.007812</td>
<td>0.107347</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>119</td>
<td>1</td>
<td>4</td>
<td>0.008403</td>
<td>0.123624</td>
<td>1.123624</td>
</tr>
<tr>
<td>78</td>
<td>114</td>
<td>1</td>
<td>1</td>
<td>0.008772</td>
<td>0.132396</td>
<td>1.132396</td>
</tr>
<tr>
<td>79</td>
<td>112</td>
<td>1</td>
<td>3</td>
<td>0.008929</td>
<td>0.141325</td>
<td>1.141325</td>
</tr>
<tr>
<td>80</td>
<td>108</td>
<td>2</td>
<td>1</td>
<td>0.009259</td>
<td>0.150584</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>105</td>
<td>2</td>
<td>4</td>
<td>0.009524</td>
<td>0.169454</td>
<td>1.169454</td>
</tr>
<tr>
<td>84</td>
<td>99</td>
<td>2</td>
<td>1</td>
<td>0.010101</td>
<td>0.189170</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>96</td>
<td>1</td>
<td>2</td>
<td>0.010417</td>
<td>0.209791</td>
<td>1.209791</td>
</tr>
<tr>
<td>88</td>
<td>93</td>
<td>1</td>
<td>1</td>
<td>0.010753</td>
<td>0.220544</td>
<td>1.220544</td>
</tr>
<tr>
<td>89</td>
<td>91</td>
<td>2</td>
<td>7</td>
<td>0.010989</td>
<td>0.231533</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td>0.011111</td>
<td>0.242644</td>
<td>1.242644</td>
</tr>
<tr>
<td>92</td>
<td>82</td>
<td>1</td>
<td>2</td>
<td>0.012195</td>
<td>0.254839</td>
<td>1.254839</td>
</tr>
<tr>
<td>94</td>
<td>79</td>
<td>1</td>
<td>4</td>
<td>0.012658</td>
<td>0.267497</td>
<td>1.267497</td>
</tr>
<tr>
<td>96</td>
<td>74</td>
<td>2</td>
<td>3</td>
<td>0.013514</td>
<td>0.281011</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>69</td>
<td>1</td>
<td>1</td>
<td>0.013699</td>
<td>0.294710</td>
<td>1.294710</td>
</tr>
<tr>
<td>102</td>
<td>67</td>
<td>2</td>
<td>1</td>
<td>0.014295</td>
<td>0.309203</td>
<td>1.309203</td>
</tr>
<tr>
<td>103</td>
<td>64</td>
<td>3</td>
<td>1</td>
<td>0.015625</td>
<td>0.339280</td>
<td>1.339280</td>
</tr>
<tr>
<td>104</td>
<td>60</td>
<td>1</td>
<td>59</td>
<td>0.016129</td>
<td>0.386907</td>
<td>1.386907</td>
</tr>
</tbody>
</table>

While litter membership has been ignored in obtaining the scores shown in Table 2, it is taken into account in analyzing those scores. Thus, the successive columns of Table 3 show for each litter: the total of all scores for that litter, \( \sum Y \); the sum of squared scores, \( \sum Y^2 \); the sum of scores for treated rats only in that litter, \( \sum Y^T \); the expected sum of scores for treated rats in that litter, \( E \sum Y^T / N \); the variance of the sum of scores for treated rats in that litter, \( \text{VAR}(\sum Y^T) \). These quantities are totalled across all litters at the bottom of the table.

In these instances, each \( N_1 = 1 \), each \( N_2 = 2 \), each \( N = N_1 + N_2 = 3 \). In other instances, \( N_1, N_2, \) and \( N \) values might differ. Also, had there been any interim losses (deaths without tumor) prior to the first tumor-appearance week, it would probably be more preferable to exclude the animals involved from the analysis made than to include them with a score of unity. In that case, the \( N_1, N_2, \) and \( N \) values would also have changed. In the present example, this did not occur.

Two alternative analyses are shown in Table 3 for the illustrative data, litter membership being taken into account in the first but ignored in the second. For both analyses, one obtains a difference between the total of observed and expected scores for treated animals, with a 1 d.f. \( \chi^2 \) value resulting when the square of that difference is divided by its variance. In the present instance, because of uniformity of the \( N_1 \) values and of the \( N_2 \) values across litters (more properly, equality of the \( N_1/N_2 \) ratios), both analyses yield the same total of expected scores, except for rounding effects, 50.007 and 50.009 (totals of observed scores are necessarily equal). Yet other rounding effects have caused these to deviate from the total number of treated animals, 50. The total of all scores, 150.026 (Column 2), differs from the total number of animals primarily because the scores for the 60 rats remaining to Week 104 have each been rounded up by 0.000426 for a total rounding effect of 0.02556.
Nevertheless, the analysis allowing for litter effects shows a total deviation of \(-8.598\) with variance \(7.799\) to yield a \(\chi^2\) of 9.48. This contrasts with the 3 analyses involving recovery of interlitter information report by Mantel, Bohidar, and Ciminera; \(\Sigma A - \Sigma E(A)\) values ranged from \(+8.6476\) to \(+8.9193\) and \(\Sigma VAR(A)\) from 7.4281 to 7.5085. Continuity-corrected \(\chi^2\) values ranged from 8.84 to 9.51, but if the continuity correction is dropped, as brought out above, the range of \(\chi^2\) values would be from 9.96 to 10.67, all above the present \(\chi^2\) of 9.48.

An interesting comparison results when we consider the analysis ignoring litter membership. Here the score deviation is \(-8.600\), the variance is 8.863, and the \(\chi^2\) value is 8.34. The Mantel-Bohidar-Ciminera analysis ignoring litter membership showed a \(\Sigma A - \Sigma E(A)\) value of \(+8.5484\), a \(\Sigma VAR(A)\) value of 8.4905, and a corrected \(\chi^2\) of 7.63; without continuity correction the \(\chi^2\) would be 8.61, or greater than the value of 8.34.

We can recognize that the 2 deviations, \(-8.600\) and \(+8.5484\), are closely similar except for sign. Apart from rounding effects, the difference can be attributed to the variant handlings of tied observations by the Mantel-Haenszel approach and by the logrank score approach. The contrast in variances, 8.863 versus 8.4905, is not excessive. As noted above, in the absence of ties and interim losses to observation, the variance by the Mantel-Haenszel procedure would, on the average, equal the fixed logrank score variance. If lower Mantel-Haenszel variances occur when the data seem to contradict the null hypothesis, that would account for the moderately lower Mantel-Haenszel variance in this case and the moderately higher uncorrected \(\chi^2\). It may be recognized now that, for the analyses taking litter membership into account, smaller variances also attended the Mantel-Haenszel-type analyses; those smaller variances could have offset any loss of information relative to the logrank score analysis now proposed.

Discussion

The primary justification for the method proposed here is that it leaves no loose ends to the question of whether there has been full and proper recovery of information. It was clear that, of the 3 methods given by Mantel, Bohidar, and Ciminera for recovery of interlitter information, the first 2 were subject to some loss of such information, while for the third there was an admitted lack of complete propriety. In fact, where litters are of arbitrary and varying size with arbitrary allocation to treatment or controls, none of the methods given by Mantel, Bohidar, and Ciminera would be fully appropriate. By the approach here given, a strongly related method is applied collectively to all the litters rather than to the separate litters, but only for the purpose of affixing scores to each individual response. Once the scores are fixed, the individual litter membership of each animal is taken into account. This method of assigning scores using combined litters is analogous to combining strata before assigning ranks rather than assigning ranks within strata in comparing 2 or more treatment groups. Mantel (7) suggested just such a pooled ranks procedure as an alternative to pooling within strata, while Mantel and Schneiderman (13) used that device in order to avoid the need for repeated reranking which otherwise would have arisen. Kruskal and Wallis (6) do use pooled ranks in multiple-sample situations, but conceptually all the samples are from a single stratum and the question is whether they are also from the same population; there is no problem of removing stratum, i.e., litter, effects.

Suppose that instead of the many small strata as occurred in the example (each stratum consisting of 3 rats from a litter) we had several large strata; might not the method of assigning scores using pooled strata be used? The answer would be no, because the residual information in the remnants of a large stratum would be small compared to the information extracted from the stratum before it was reduced to a remnant condition.

One advantage of the Mantel-Bohidar-Ciminera analyses was that they permitted simple updating as a study progressed or, if the study were already completed, determining what the results of analysis would have been at each time point. Only the outcomes in a given week would be required for determining the change from the analysis for the previous week. This
Calculation of \( x^2 \), allowing for litter effects:

\[
\chi^2 = \left( \sum \frac{Y}{N} - \Sigma \frac{E(Y)}{N} \right)^2 / \text{VAR} \sum \frac{Y}{N}
\]

in which \( E(Y) = \Sigma \frac{N_i}{N} \), where \( N_i \) is the number of rats in each litter of the treated group.

Calculation of \( x^2 \), ignoring litter effects:

\[
\chi^2 = \left[ \sum \frac{Y}{N} - \Sigma E(Y) \right]^2 / \text{VAR} \sum \frac{Y}{N}
\]

\( \text{VAR} \) is calculated from Table 3 as shown in Table 3—Continued.
rates among treated and control animals, and without concern as to when the tumors arose. In the present example, there were no rats dying without tumor before tumors started appearing; therefore, there was no problem of not counting as at risk those rats dying early in the study without tumor. Such modification is sometimes almost essential if one wishes to look only at crude tumor occurrence rates. However, consider the highest logrank score of 0.404 for tumors in rats, $1 - \exp(0.404) = 0.332$, suggesting that as of the end of the study, Week 104, there was only a one-third chance of having developed tumor. Suppose the study continued somewhat longer in time, with an appreciably higher cumulative probability of tumor occurrence. Logrank scores for tumors in rats could start approaching and possibly exceeding those for some nontumorous rats. The similarity to a 2-point distribution no longer exists, and it is no longer so reasonable to reduce the data to a single $2 \times 2$ table to be analyzed as such.

Yet another object lesson follows, seemingly contrary, if we consider that a tumorigenesis trial has been followed so long that the cumulative probability of developing tumor is getting somewhat closer to unity. The logrank score for a late-appearing tumor may now become rather large, even considerably in excess of unity and in excess of the logrank scores for rats dying without tumor or otherwise lost to observation without tumor somewhat earlier in the study. For interpretation, what is important in such instances is not that the rat has developed a tumor but that it went for such a long time without developing one. Late development of tumors may essentially constitute negative evidence for a tumorigenic effect. On other bases, it may be argued that a tumor appearing towards the end of a lifetime is nowhere near as important as early-appearing tumors and so should be discounted; in our experience, responsible authorities do not accept this argument and claim instead that a tumor is a tumor whenever it appears. However, consideration of the logrank scores for such late tumors brings out the opposite. A logrank score analysis will not count late-appearing tumors as positive evidence. If one wishes instead to make a simple-minded analysis relating only to the crude tumor rates, then the very late tumors in the study should be excluded in calculating and comparing tumor rates.

Although the present illustration has been one of a comparison between treated and control animals, the extension to the case of several levels of treatment, with some kind of dosage or dosage score assigned to controls and to each treatment level, is immediate and would follow the lines indicated by Mantel (7). Where treatments are distinctive and not orderable or scorable, a multi-d.f. $\chi^2$ can be calculated. Such multi-d.f. situations are discussed by Mantel and Haenszel who provide a calculating device for 2-d.f. situations, while Mantel and Byar (11) provide a general procedure for multi-d.f. $\chi^2$ values.

The logrank scores herein described could have application in studies in which all members of a litter receive identical treatment. Such identity of treatment may apply in multigenerational studies, necessarily occurs where treatment is in utero, and frequently applies where treatment is neonatal. The results for all animals in all litters are combined in order to get a logrank score for each animal, the total of which for a litter provides a score for that litter. Where litter sizes (of a sex) are equal, as might be accomplished by culling, between-litter differences provide the error term for evaluating treatment differences. For moderate differences in litter size, it could be reasonable to take as the litter score the average of the scores for the members of that litter, although other devices might also apply, e.g., stratification by litter size or use of litter size as a covariate.

**Appendix**

A reviewer has suggested a modified logrank score procedure for application to litter-matched data. The results of application of that scoring procedure are readily obtained from those given herein.

The present procedure can be thought of as one in which the total logrank score for each group, treated or controls, is exactly that which would obtain in the homogeneous case, all the rats being treated as one enormous litter. However, litter membership is recognized in the calculation of the variance or variances of those total logrank scores. This approach maintains the average logrank score across all rats at unity, but the averages for the separate litters can deviate from unity. This contrasts with a pure within-litter analysis, cf. Mantel, Bohidar, and Ciminera (10), in which, effectively, the separate litter averages are maintained at unity. Under the reviewer’s suggestion, each logrank score, as calculated, would be divided by the average score for its litter.

That change can be effected with no need to recalculate individual scores. Let the average score for a litter be $\bar{Y} = \Sigma Y/N = \text{Column 2}/N$; replace each Column 4 entry by $(4)_l = (4)/\bar{Y}$ and each Column 6 entry by $(6)_l = (6)/\bar{Y}^2$. (Column numbers refer to Table 3.) The Column 5 entry would be replaced by the number of treated rats in each litter or can be deleted, because the total expected score for treated rats would be simply the total number of treated rats across all litters. With these modifications, the reviewer’s $\chi^2$ would be $[\Sigma(4)_l - \Sigma(5)_l]2/\Sigma(6)_l$.

When applied to the data of the instant example, the reviewer reports obtaining a reduced $\chi^2$ of 5.69 as against 9.48 here reported, or 5.99 by our earlier [Mantel, Bohidar, and Ciminera (10)] intralitter analysis without recovery of interlitter information. Associated with this reduced $\chi^2$ is a moderate increase in the total deviation from expectation, but a sharp increase in variance, by a factor of 2.27 relative to the present analysis or 2.64 relative to the analysis in Ref. 10.

In his hypothetical example proposing the modified approach, the reviewer has postulated a situation with sharp litter differences, but no losses to observation are allowed in the illustration. Thus, all responses occur first in one litter, then all in another litter, than all in yet another litter, etc. With such sharp litter differences and no censoring, the reviewer’s suggestion could indeed be advantageous, but it is questionable whether actual litter differences could show patterns like those of the reviewer’s example. For our own example, litter differences would seem to be moderate. In our data, differences between litter averages of logrank scores would seem to be dominated by chance variations in the number of censored observations, i.e., death or sacrifice without tumor, rather than by differences in typical time to tumor appearance. In the particular example, the reviewer’s approach leads to greater chance variation of the total score, reduced statistical significance. The reader can use either or both approaches as he deems suitable. In any case, the reviewer has declined to be identified. Users of his proposal may wish to identify it as a modified Mantel-Ciminera logrank procedure.
John Tukey has noted that, because treatment with a potential tumorigen can cause deaths, hence censored observations, as well as tumors, the use of a permutation approach for obtaining variances of total logrank scores of treated rats could be invalid. The distribution of observations for treated animals will differ from that for controls even if no tumorigenic effect existed. This difficulty would also attend use of other recognized procedures for handling time-to-response data, e.g., those of Gehan (4, 5, 9), when treatment can modify the censoring mechanism.

In the present instance, involving 50 litters, an alternative estimate of the total variance can be based, with 49 d.f., on litter-to-litter variation in the departure of the logrank score for the treated rat from its expectation, i.e., its litter average. This approach yields an estimated variance of 7.912 (cf. 7.799 and 8.863 in Table 3) for a t value of -3.057, t^2 = 9.34, p (single-tail) ca. 0.002. The limited change in the estimated variance would reflect the small intralitter correlation in the occurrence of tumors. Use of between-litters variation is adaptable to the case of experiments at several dose levels. Also, the analysis is readily modifiable so as to take into account tumors identifiable only after death, an issue raised by Peto (14). Tukey has further suggested the possibility for obtaining a jackknifed estimate of the desired variance, using the entire results for each litter as the unit of observation in such calculation.

With use of litter-to-litter variation as the basis for calculating the variance of logrank score analyses, the advantage of permitting variation in litter sizes would no longer exist, and for uniform litter sizes our earlier method (10) would be applicable without being subject to the effect of differential mortality.

For clinical trials and other studies in which treatment should not modify the censoring mechanism, the permutational approach would remain appropriate for determining variances.

We note that all methods for analyzing time-to-tumor data are subject to an unavoidable assumption which may not be fully tenable. The assumption is that, where observations are censored because of death, the future course of the animals would not otherwise have been different from those of surviving animals.

Acknowledgment

The valuable suggestions of Richard Peto are gratefully acknowledged.

References

Use of Logrank Scores in the Analysis of Litter-matched Data on Time to Tumor Appearance

Nathan Mantel and Joseph L. Ciminera

*Cancer Res* 1979;39:4308-4315.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/39/11/4308

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.