

## TABLES OF THE NUMBER OF PATIENTS REQUIRED IN CLINICAL TRIALS USING THE LOGRANK TEST

L. S. FREEDMAN

*MRC Cancer Trials Office, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH England*

### SUMMARY

The logrank test is commonly used in the analysis of clinical trials in chronic diseases such as cancer. Existing tables for the number of patients required in such trials are based on the direct comparison of two proportions. This paper presents tables of numbers required in clinical trials using the logrank test and describes their use. The numbers required are considerably smaller than those in existing tables when the event-free proportions are small, but otherwise comparable.

KEY WORDS Logrank test Clinical trials Power

### INTRODUCTION

The logrank test is now commonly used in the analysis of clinical trials data where outcome becomes manifest after a prolonged time interval.<sup>1</sup> The test is particularly well established in the analysis of results from trials of cancer treatments.

Recently, several authors have pointed out the inadequacy of the numbers of patients entered into many cancer trials and the consequent lack of sensitivity to small treatment differences.<sup>2, 3</sup> Consideration of the number of patients required is important in planning a trial. There are several published tables of the numbers needed in comparative experiments or trials.<sup>4, 5</sup> These entail the assumption that analysis consists of a direct comparison of two proportions, e.g. the survival proportions in two treatment groups. This paper presents more extensive tables of numbers required in trials in which the logrank test constitutes the principal method of analysis.

Before describing the tables and their use, we elaborate the nature of the trials and explain several technical terms. The trials consist of patients who are entered, treated and then followed up. Principal interest during follow-up concerns the occurrence or non-occurrence of a particular 'event' which in many cancer trials is death but could also be local recurrence, metastatic spread or some other clinical observation. Some patients may be *withdrawn* from follow-up. This does not (or should not) mean that the investigator has consciously decided not to keep track of a patient but that despite efforts the patient's follow-up is in some sense incomplete. There are two primary reasons for this. Either the investigator may lose touch with the patient (perhaps because the patient has gone abroad) or some intervening event prevents gathering the necessary information. An example of such an intervening event is a patient who dies in a car accident soon after treatment when death from malignant disease is the event of interest. These concepts appear later in the paper.

### ASSUMPTIONS UNDERLYING THE CALCULATIONS

The tables give two quantities: the number of events needed to be observed and the number of patients needed to be entered. The number of events results from a formula which approximates the exact number required (see Appendix I). The formula assumes 1:1 randomization, i.e. equal numbers entered in the two treatment groups. A comparison of the results obtained from the formula with those obtained from Monte Carlo simulations appears in Appendix II. This comparison assures that the approximation is reasonably accurate and shows that it provides a slight over-estimate of the number of events, which, in practical terms, is an error in the right direction. The number of patients needed can be estimated directly as that required to observe the necessary number of events. The exact form of dependence of the number of accrued events on the patient entry will be determined by the rate of acceptance of patients into the trial, the rate of occurrence of the events and the timing of the definitive analysis of the results.<sup>6</sup>

The tables entail an assumption that analysis occurs at a fixed time  $T$  after the *last* patient has entered the study; information on patient follow-up extending beyond  $T$  is excluded. This assumption commonly (though not universally) accords with practice and is thereby partly motivated. The consequent analysis does not utilize all available information and hence the required number of patients is over-estimated. Only when the great majority of the relevant information has been gathered, however, is it proper to conduct a definitive analysis. This implies choice of a minimum follow-up time  $T$  beyond which the rate of occurrence of events is low. Such a choice of  $T$  ensures that any over-estimation of the required numbers of patients is slight.

The assumption has the positive consequence that the number of required patients is independent of the rates of entry of patients and occurrence of events and depends only on the proportions of event-free patients in the two treatment groups after minimum follow-up time  $T$ . Thus, tabulation of the numbers required becomes feasible. The calculations are described in Appendix I.

One other aspect of a trial which can affect the numbers required is the proportion of patients who are withdrawn (see introduction for the definition of this term). The numbers in the tables assume no withdrawals. Since some withdrawals almost always happen, the investigator must make some allowance for this. For example, if he anticipates  $x$  per cent of patients to withdraw and  $n$  is the required number of patients in the table, then he should actually aim to enter  $100n/(100 - x)$  patients.

### DESCRIPTION OF THE TABLES AND THEIR USE

The form of the tables is similar to that of Tables 3A and 3B of Casagrande *et al.*<sup>5</sup> which, in turn, are based on Table 2.1a of Cochran and Cox.<sup>4</sup> To use these tables one must first 'guess' the proportion of patients event-free at the minimum follow-up interval in the less favourable of the two groups. When the event is death this proportion is simply the survival rate. Many clinical trials compare a 'new' with a 'standard' treatment with the accompanying hope that the 'new' leads to improvement. Here, the less favourable group consists of those patients treated by the 'standard'; previous experience usually provides a reasonable guess at the event-free rate. One must then specify the smallest improvement in event-free rate one wishes the trial to be able to detect reliably. At this point one chooses Table 1A or 1B depending, respectively, on interest in a one- or two-tailed test of significance. Generally, one-tailed tests apply to trials of a standard with a new treatment when the new is more toxic or more expensive. Interest focuses on differences in response which are favourable to the new treatment group. This is particularly relevant to trials comparing a combination of treatments with one component of the combination (e.g. surgery and radiotherapy

vs. surgery alone). Two-tailed tests apply when there are no strong *a priori* grounds to favour one or the other of the treatments such as a comparison of two single drugs with similar toxicities.

Now one turns to the appropriate table and looks along the row corresponding to the smallest improvement. Each cell in the table contains 6 numbers. The numbers in parentheses are the total number of events needed to be observed; the numbers without parentheses are the total numbers of patients required. The 3 sets of numbers of patients and events correspond to three combinations of significance levels and 'power'. 'Power' is the chance of finding a significant difference if it exists. The three combinations are: test of significance at 5 per cent level, power 80 per cent; test of significance at 5 per cent level, power 90 per cent; and test of significance at 1 per cent level, power 95 per cent. These are the same combinations as chosen by Casagrande *et al.*<sup>5</sup> and Cochran and Cox<sup>4</sup> and facilitate comparison between the different tables. We emphasize that the final choice of the number of patients entails a compromise between the expected rate of patient entry and the statistical ideal. Thus, if only 100 patients per annum are likely to be available it may be better to plan a study of 300 patients with a power of 80 per cent than a trial of 500 patients with a power of 95 per cent. Although the latter trial is more sensitive to any differences between treatments, it may not be realistic to expect to maintain enthusiasm for the trial beyond 3 years. On the other hand, if expected patient entry rates are insufficient to provide adequate power, it may be better not to embark on the study. In the absence of obvious restrictions imposed by patient entry rates one might, as a general rule of thumb, recommend a power of 90 per cent at a significance level of 5 per cent for comparative studies of two treatments.

### Example

Consider the planning of a trial of superficial bladder cancer. With the current method of treatment (resection of tumour at cystoscopy) the recurrence-free rate is 50 per cent at 2 years. One hopes to increase this to at least 70 per cent using intravesical chemotherapy immediately after surgery at the time of cystoscopy. Referring to Table 1A (one-tailed test), the appropriate cell indicates a sample size of 153, 211, or 386 patients according to the particular combination of power and level of significance. Allowance for a possible 20 per cent withdrawal rate increases these numbers to 190, 264 and 482 respectively. Thus, between 250 and 300 patients seems a reasonable size for this trial. The relatively high incidence of this tumour (around 7000 new cases per annum in England and Wales) and the high level of interest among urological surgeons make this sample size a realistic goal for a trial.

## DISCUSSION

The assumption that analysis excludes information beyond the minimum follow-up time may be unattractive, particularly when patient accrual is extended over several years. In such circumstances, a large proportion (80 per cent or more) of the total events expected may have already occurred a short time after patient entry has closed. If so, an analysis at this juncture may be reasonable. For example, consider a trial in which patients are entered at a constant rate over 3 years. Suppose that the average survival rate at 1 year after treatment is 50 per cent and that there is an exponential distribution of survival times up to 4 years beyond which time the death rate is negligible. The total proportion of deaths in the trial is 94 per cent. One year after patient entry closes the proportion of deaths is 79 per cent which is 84 per cent of the total deaths expected. Therefore an analysis at one year after the last patient has entered, but including information beyond one year's follow-up is justified.

To estimate the required number of patients under this policy, it would be *wrong* to enter the

Table IA. Number of patients required to detect an improvement ( $P_2 - P_1$ ) in survival rate over a baseline survival rate ( $P_1$ ), when (i)  $\alpha = 5$  per cent,  $1 - \beta = 80$  per cent, (ii)  $\alpha = 5$  per cent,  $1 - \beta = 90$  per cent, (iii)  $\alpha = 1$  per cent,  $1 - \beta = 95$  per cent

1-tailed test  
 $P_2 - P_1$

$P_1$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0.05	391 (362) 511 (511) 997 (932)	177 (123) 189 (170) 348 (313)	79 (69) 109 (95) 200 (175)	55 (46) 75 (68) 138 (117)	42 (36) 95 (87) 106 (87)	34 (27) 47 (38) 86 (69)	29 (23) 40 (31) 73 (56)	25 (19) 35 (26) 64 (48)	23 (16) 31 (22) 57 (41)	21 (14) 28 (20) 52 (36)
0.10	759 (644) 1051 (920) 1934 (1692)	332 (273) 322 (273) 592 (503)	122 (101) 169 (139) 311 (256)	80 (64) 110 (88) 202 (161)	58 (45) 80 (62) 146 (113)	45 (34) 62 (47) 114 (86)	37 (27) 51 (37) 93 (69)	32 (22) 43 (30) 79 (58)	28 (18) 38 (25) 69 (46)	24 (16) 33 (23) 61 (41)
0.15	1115 (973) 1640 (1377) 3840 (3277)	379 (326) 415 (353) 815 (652)	161 (134) 206 (172) 406 (312)	101 (79) 135 (107) 255 (197)	76 (51) 100 (76) 180 (130)	54 (38) 75 (53) 137 (95)	43 (30) 56 (40) 109 (73)	36 (23) 46 (32) 90 (58)	31 (19) 42 (28) 77 (48)	27 (15) 35 (22) 67 (40)
0.20	1439 (1155) 1993 (1545) 3668 (2843)	398 (318) 551 (413) 1013 (760)	194 (140) 268 (194) 492 (357)	118 (83) 167 (114) 300 (210)	82 (55) 113 (76) 207 (140)	61 (40) 84 (55) 154 (100)	48 (30) 66 (41) 131 (76)	40 (24) 54 (32) 99 (59)	33 (19) 46 (26) 83 (48)	29 (16) 40 (22) 72 (37)
0.25	1723 (1249) 2386 (1730) 4391 (3183)	464 (325) 643 (450) 1182 (827)	321 (149) 306 (206) 562 (379)	133 (86) 119 (78) 336 (218)	90 (56) 124 (78) 228 (142)	66 (40) 92 (55) 168 (100)	52 (30) 71 (41) 130 (75)	42 (23) 58 (31) 105 (58)	35 (18) 48 (25) 87 (46)	30 (15) 41 (20) 75 (37)
0.30	1969 (1323) 2675 (1932) 4765 (3513)	518 (336) 717 (465) 1317 (857)	342 (151) 509 (329) 889 (573)	143 (86) 198 (119) 364 (218)	96 (55) 133 (76) 244 (140)	70 (38) 97 (55) 177 (100)	54 (28) 74 (41) 140 (75)	43 (21) 60 (30) 109 (59)	36 (17) 50 (25) 90 (46)	31 (14) 41 (18) 72 (34)
0.35	2147 (1341) 2973 (1858) 5471 (3419)	559 (325) 774 (463) 1423 (853)	358 (148) 507 (305) 856 (577)	151 (83) 208 (114) 362 (210)	100 (52) 138 (76) 253 (135)	72 (36) 100 (50) 182 (91)	55 (26) 76 (36) 138 (65)	44 (20) 60 (27) 110 (46)	36 (15) 50 (21) 90 (38)	31 (12) 42 (16) 76 (30)
0.40	2281 (1312) 3159 (1816) 5814 (3343)	586 (322) 812 (446) 1473 (821)	368 (140) 370 (194) 680 (357)	155 (77) 214 (107) 392 (196)	102 (48) 140 (66) 257 (122)	73 (33) 100 (45) 183 (82)	55 (23) 76 (32) 138 (58)	44 (17) 60 (24) 109 (43)	36 (13) 49 (18) 88 (33)	30 (10) 41 (14) 74 (28)
0.45	2355 (1240) 3072 (1719) 5625 (3181)	600 (300) 829 (413) 1524 (784)	374 (129) 536 (329) 889 (573)	155 (70) 214 (107) 392 (196)	101 (43) 140 (66) 257 (122)	72 (38) 100 (50) 182 (91)	54 (28) 74 (41) 140 (75)	43 (21) 60 (30) 109 (59)	36 (17) 50 (25) 90 (46)	30 (11) 41 (14) 72 (30)
0.50	2381 (1175) 3111 (1722) 6092 (2854)	601 (374) 817 (471) 1530 (688)	379 (144) 522 (305) 863 (590)	153 (81) 214 (107) 392 (196)	98 (41) 138 (76) 253 (135)	70 (34) 100 (50) 182 (91)	52 (26) 72 (36) 136 (65)	41 (17) 56 (23) 100 (43)	33 (12) 45 (16) 80 (30)	28 (9) 38 (13) 66 (22)
0.55	2345 (1005) 3225 (1391) 6025 (2561)	588 (335) 814 (425) 1497 (598)	361 (97) 360 (135) 661 (248)	146 (51) 202 (70) 370 (129)	94 (30) 129 (42) 236 (76)	66 (19) 90 (27) 163 (49)	49 (13) 67 (18) 120 (33)	38 (8) 52 (13) 92 (23)	30 (9) 41 (12) 76 (18)	24 (7) 34 (10) 61 (14)
0.60	2285 (856) 3164 (1186) 5821 (2185)	562 (196) 778 (422) 1429 (500)	346 (80) 500 (110) 824 (202)	137 (41) 189 (56) 345 (103)	87 (24) 120 (33) 217 (59)	61 (15) 87 (30) 160 (37)	45 (10) 60 (16) 106 (25)	30 (7) 41 (10) 76 (18)	22 (6) 31 (8) 57 (13)	16 (4) 22 (6) 43 (9)
0.65	2151 (688) 2976 (948) 5481 (1781)	523 (156) 723 (216) 1328 (598)	326 (62) 457 (157) 873 (457)	125 (30) 171 (57) 312 (78)	77 (21) 107 (24) 194 (43)	54 (10) 74 (14) 132 (26)	38 (7) 52 (11) 92 (19)	28 (6) 38 (8) 69 (14)	20 (5) 27 (6) 52 (11)	14 (3) 19 (4) 37 (8)
0.70	1945 (540) 2721 (748) 5065 (1376)	470 (117) 650 (162) 1193 (298)	201 (45) 276 (62) 505 (113)	109 (21) 149 (29) 271 (54)	68 (12) 98 (21) 166 (29)	45 (8) 62 (11) 114 (16)	30 (5) 41 (6) 76 (11)	20 (4) 27 (5) 52 (8)	14 (2) 19 (3) 37 (6)	9 (1) 12 (2) 24 (4)
0.75	1726 (388) 2392 (567) 4392 (986)	405 (81) 559 (111) 1025 (205)	167 (29) 233 (40) 424 (74)	91 (13) 124 (18) 222 (33)	66 (9) 91 (13) 166 (24)	45 (5) 60 (8) 106 (14)	30 (3) 41 (5) 76 (10)	20 (2) 27 (3) 52 (7)	14 (1) 19 (2) 37 (5)	9 (0) 12 (1) 24 (3)
0.80	1476 (251) 1986 (377) 3652 (639)	327 (49) 450 (77) 824 (123)	174 (22) 238 (36) 411 (57)	87 (11) 118 (16) 206 (28)	54 (6) 74 (9) 132 (18)	38 (4) 52 (6) 92 (12)	28 (3) 38 (5) 69 (9)	20 (2) 27 (3) 52 (7)	14 (1) 19 (2) 37 (5)	9 (0) 12 (1) 24 (3)
0.85	1094 (136) 1513 (189) 2778 (347)	238 (23) 326 (32) 592 (58)	137 (13) 187 (18) 336 (32)	66 (6) 91 (9) 166 (16)	45 (4) 60 (6) 106 (10)	30 (3) 41 (4) 76 (7)	20 (2) 27 (2) 52 (5)	14 (1) 19 (2) 37 (4)	9 (0) 12 (1) 24 (3)	6 (0) 8 (1) 16 (2)
0.90	705 (52) 1776 (133)	52 (2) 133 (13)	22 (1) 33 (3)	11 (1) 15 (1)	7 (0) 10 (0)	5 (0) 7 (0)	4 (0) 5 (0)	3 (0) 4 (0)	2 (0) 3 (0)	1 (0) 1 (0)

(i) Example: To detect an improvement from 0.5 to 0.7 using a 1-tailed test consult the cell in the table which is shaded.  
 (ii) Note: For numbers to detect smaller but realistic differences in the bottom left-hand corner use formulae (2) and (3):  
 $d = (z_{\alpha} + z_{\beta})^2 (1 + \theta)^2 / (1 - \theta)^2$ ;  $n = 2d / (2 - P_1 - P_2)$



table with event-free rates appropriate to the minimum follow-up time (in our example, 1 year). The number of patients needed would thereby be over-estimated, often seriously. Instead, as an approximate device, the table should be entered with event-free rates appropriate to the *average* follow-up time (in our example 2.5 years). The number of patients estimated in this way is approximate but adequate for practical purposes.

Some investigators are more used to thinking of treatment differences in terms of median survival than survival rates. The following is meant to give a rough feeling for their relationship. Survival times are assumed exponentially distributed. A treatment which increases the median survival time by 50 per cent changes a survival rate of 50 per cent to 63 per cent, a rate of 25 per cent to 40 per cent or a rate of 10 per cent to 22 per cent. A treatment which doubles the median survival time increases the survival rate from 50 to 71 per cent, from 25 to 50 per cent or from 10 to 32 per cent.

Table 1A and 1B are similar in form to those of Casagrande *et al.*<sup>5</sup> and Cochran and Cox<sup>4</sup> but differ in two notable ways. Firstly the values of the lower event-free rates are extended in our tables up to 90 per cent instead of stopping at 50 per cent. The extension is necessary because, whereas the numbers required in a direct comparison of proportions are symmetric around the 50 per cent point (i.e. one needs the same numbers for detecting a difference between proportions of 30 per cent and 40 per cent as between 70 per cent and 60 per cent), this is *not* true of the logrank test. Secondly the numbers in the tables are total numbers of patients (or events required in a trial) whereas Casagrande *et al.*<sup>5</sup> give numbers required in each group. We made this change to *total* numbers purposely because we believe this figure most directly interests investigators.

For the greater part of our tables the numbers of patients required are similar to those given by Casagrande *et al.*<sup>5</sup> although slightly smaller. This is of interest since the results from Appendix II suggest that the numbers in Tables 1A and 1B are slight over-estimates. In one part of the table the differences are more important. When event-free rates are very low, namely under 25 per cent, then it appears that substantially smaller numbers of patients are required with use of the logrank test than one would have thought had one consulted the tables by Casagrande *et al.*<sup>5</sup> or similar tables. For example, at a significance level of 5 per cent and power 90 per cent the number of patients required to detect an improvement from a baseline survival rate of 10 per cent to a new survival rate of 20 per cent is, from Table 1A, 322. The equivalent number according to the tables of Casagrande *et al.*<sup>5</sup> is 464, a figure about half as large again. This is not very surprising when one considers that the logrank test takes account of the order in which events occur and not just simply the occurrence or non-occurrence of events. Thus when only a small minority of patients remains event-free the gain in information using the logrank test is considerable. In view of the important difference, we recommend that the tables in this paper should be used when designing clinical trials that will employ the logrank test as the principal analytic method.

#### ACKNOWLEDGEMENTS

I thank Dr. I. Sutherland and Dr. S. Gore who gave helpful advice and Mrs. B. Smith who typed the manuscript. I also thank the referee who suggested the use of average follow-up time.

#### APPENDIX I

Suppose that two treatments give rise to survival rates of  $P_1$  and  $P_2$ , respectively at some chosen time point. If the ratio of the hazards (i.e. risks of death) in the two groups does not change with time and is  $\theta:1$  then the quantities  $P_1$ ,  $P_2$  and  $\theta$  are related by:

$$\theta = \log_e(P_1)/\log_e(P_2) \quad (1)$$

The total number of events  $d$  in both series, needed to be observed in a trial is:

$$d = (z_1 + z_2)^2 \left( \frac{1 + \theta}{1 - \theta} \right)^2 \tag{2}$$

where  $z_1$  is the normal deviate corresponding to the particular significance level employed in the logrank test and  $z_2$  is the normal deviate corresponding to the required power.

Formula (2) is derived by considering the expected value ( $E$ ) and variance ( $V$ ) of the logrank statistic when the true hazard ratio is  $\theta$ . By arguing conditionally on the set of patients at risk before each event and letting  $\phi_i$  denote the ratio of patients at risk in the two groups before event  $i$  ( $i = 1, \dots, d$ ), then

$$E = \sum_{i=1}^d \left[ \frac{\phi_i \theta}{1 + \phi_i \theta} - \frac{\phi_i}{1 + \phi_i} \right] / \sqrt{\left( \sum_{i=1}^d \frac{\phi_i}{(1 + \phi_i)^2} \right)}$$

and

$$V = \sum_{i=1}^d \left[ \frac{\phi_i \theta}{(1 + \phi_i \theta)^2} / \sum_{i=1}^d \left[ \frac{\phi_i}{(1 + \phi_i)^2} \right] \right]$$

Assuming  $\phi_i = 1$  these reduce to

$$E = d^{1/2}(\theta - 1)/(\theta + 1)$$

$$V = 4\theta/(\theta + 1)^2.$$

and

By treating the logrank statistic as a Normal variable with mean  $E$  and variance  $V$  one may then show that

$$d = \frac{(\theta + 1)^2}{(\theta - 1)^2} \left[ z_1 + \frac{2z_2 \sqrt{\theta}}{\theta + 1} \right]^2$$

where, as mentioned earlier,  $z_1$  and  $z_2$  are normal deviates corresponding to the required significance level and power. Finally, taking the coefficient of  $z_2$ ,  $2\sqrt{\theta}/(\theta + 1)$ , as approximately equal to 1, we obtain Formula (2).

As explained, this formula is an approximation and relies on the simplification that the ratio of the number of patients in each group at risk just before each death is equal to 1. In a trial with equal numbers of patients in each group this ratio will indeed be very near 1 at the start of treatment but will increasingly diverge from 1 as the time from treatment increases, if there is a difference in the survival rates. In addition, the true coefficient of  $z_2$  will increasingly diverge from unity as  $\theta$  differs from unity. The effect of such departures on the accuracy of formula (2) is examined in Appendix II. Unequal withdrawals from the two groups will also affect the accuracy of the approximation but this is not examined further in the present paper. Note that formula (2) relates the power of the test directly to the number of events and implies that power will be *independent of the number of patients* given that the number of events is kept fixed. This is verified separately in Appendix II.

Once the number of events,  $d$ , has been estimated the total number of patients required in the trial can be estimated by

$$n = 2d/(2 - P_1 - P_2) \tag{3}$$

assuming equal numbers in the two treatment groups.

The equivalent of formula (2) when the ratio of patients in the two groups is  $\phi : 1$  rather than  $1 : 1$  is

$$d = \frac{(z_1 + z_2)^2 (1 + \theta \phi)^2}{\phi (1 - \theta)^2} \tag{4}$$

This formula provides a basis for approximating the numbers required when randomization is, say, 2:1.<sup>7</sup> Having calculated the total number of events required from (4) the total number of patients required is

$$n = \frac{d(1 + \phi)}{\phi(1 - P_1) + (1 - P_2)} \quad (5)$$

Formula (2) corresponds to a formula (No. 22) given by Lachin<sup>8</sup> which arose as an approximation for the special case where event times have a negative exponential distribution. Other formulae for this special case are given by George and Desu.<sup>6</sup> The use of one such formula,  $d = 4(z_1 + z_2)^2 / (\ln\theta)^2$ , in the more general context of the logrank test is justified by Schoenfeld.<sup>9</sup> His method gives estimates which are similar but slightly smaller than those derived from formula (2).

## APPENDIX II

To check the formula (2) relating the number of events to the power of the logrank test, Monte Carlo simulations were carried out. A set of 1000 clinical trials with equal numbers of patients in each group, a given total number of events  $d$  and a constant hazard-ratio  $\theta$  between the groups were generated for 5 values of  $d$  and 4 values of  $\theta$ . The results appear in Table II. They indicate quite good agreement between the power predicted by formula (2) and the Monte Carlo estimates. In general the power is slightly under-estimated by the formula, the discrepancy increasing as the hazard-ratio departs further from unity. However, these differences are of little practical consequence and, as stated in the text, are in the preferred direction of slightly overestimating the number required.

Formula (2) relates the power to the total number of *events* observed, but not to the total number of patients entered. Thus the formula predicts that when the number of events remains constant, while the number of patients entered varies, the power of the logrank test remains unchanged.

Table II. Power of logrank test estimated by Monte-Carlo simulation ( $n = 1000$ ) compared with the power predicted by formula (2)

Total number of events, $d$	Total number of patients	Source	Hazard-ratio ( $\theta$ )							
			1.33		1.5		2.0		3.0	
			Sig. level ( $\alpha$ )		Sig. level ( $\alpha$ )		Sig. level ( $\alpha$ )		Sig. level ( $\alpha$ )	
			0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
20	40	Monte Carlo (2)	0.088	0.029	0.147	0.046	0.362	0.148	0.659	0.419
			0.092	0.026	0.143	0.046	0.320	0.139	0.609	0.367
50	100	Monte Carlo (2)	0.171	0.054	0.293	0.115	0.678	0.442	0.972	0.895
			0.169	0.058	0.293	0.123	0.654	0.413	0.942	0.831
100	200	Monte Carlo (2)	0.301	0.129	0.523	0.303	0.929	0.796	0.999	0.992
			0.293	0.123	0.516	0.282	0.915	0.775	0.999	0.992
200	400	Monte Carlo (2)	0.522	0.288	0.808	0.609	0.996	0.983	—	—
			0.517	0.283	0.807	0.600	0.997	0.984	—	—
500	1000	Monte Carlo (2)	0.890	0.746	0.993	0.981	—	—	—	—
			0.886	0.723	0.994	0.971	—	—	—	—



Table III. Power of logrank test estimated by Monte-Carlo simulation ( $n = 1000$ ) for varying number of patients entered

Total number of events, $d$	Total number of patients	Hazard-ratio							
		1.33		1.5		2.0		3.0	
		Sig. level ( $\alpha$ )		Sig. level ( $\alpha$ )		Sig. level ( $\alpha$ )		Sig. level ( $\alpha$ )	
		0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
20	22	0.114	0.032	0.167	0.050	0.341	0.161	0.675	0.419
	40	0.088	0.029	0.147	0.046	0.362	0.148	0.659	0.419
	400	0.080	0.028	0.111	0.046	0.285	0.161	0.589	0.394
50	56	0.190	0.061	0.283	0.111	0.656	0.427	0.975	0.907
	100	0.171	0.054	0.293	0.115	0.678	0.442	0.972	0.895
	1000	0.190	0.073	0.301	0.135	0.697	0.462	0.970	0.903
100	110	0.304	0.119	0.543	0.284	0.922	0.811	1.000	0.998
	200	0.301	0.129	0.523	0.303	0.929	0.796	0.999	0.992
	2000	0.299	0.139	0.522	0.291	0.917	0.798	0.999	0.993

Results of further Monte Carlo simulations to verify this appear in Table III. There is no obvious trend in power, thus confirming the use of formula (2).

## REFERENCES

1. Peto, R., Pike, M. C. *et al.* 'Design and analysis of clinical trials requiring prolonged observation of each patient. II: Analysis', *Br. J. Cancer*, **35**, 1-39 (1977).
2. Pocock, S. J. 'The size of cancer clinical trials and stopping rules', *Br. J. Cancer*, **38**, 757-766 (1978).
3. Tate, H., Rawlinson, J. and Freedman, L. S. 'Randomised comparative studies in the treatment of cancer in the United Kingdom: Room for improvement', *Lancet*, **2**, 623-624 (1979).
4. Cochran, W. G. and Cox, G. M. *Experimental Design*, (2nd ed), Wiley, New York, 1957, pp 24-25.
5. Casagrande, J. T., Pike, M. C. and Smith, P. G. 'The power function of the "exact" test for comparing two binomial distributions', *Applied Statistics*, **27**, 176-180 (1978).
6. George, S. L. and Desu, M. M. 'Planning the size and duration of a clinical trial studying the time to some critical event', *J. Chron. Dis*, **27**, 15-24 (1974).
7. Peto, R., Pike, M. C. *et al.* 'Design and analysis of clinical trials requiring prolonged observation of each patient. I: Introduction and design', *Br. J. Cancer*, **34**, 585-612 (1976).
8. Lachin, J. M. 'Introduction to sample size determination and power analysis for clinical trials', *Controlled Clinical Trials*, **2**, 93-113 (1981).
9. Schoenfeld, D. 'The asymptotic properties of comparative tests for comparing survival distributions', *Biometrika*, **68**, 316-319 (1981).