



Nomograms for Calculating the Number of Patients Needed for a Clinical Trial with Survival as an Endpoint

David A. Schoenfeld; Jane R. Richter

Biometrics, Vol. 38, No. 1 (Mar., 1982), 163-170.

Stable URL:

<http://links.jstor.org/sici?sici=0006-341X%28198203%2938%3A1%3C163%3ANFCTNO%3E2.0.CO%3B2-G>

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Biometrics is published by International Biometric Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ibs.html>.

Biometrics

©1982 International Biometric Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

Nomograms for Calculating the Number of Patients Needed for a Clinical Trial With Survival as an Endpoint

David A. Schoenfeld and Jane R. Richter

Sidney Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, U.S.A.

SUMMARY

This paper presents nomograms for calculating the sample size for a clinical trial with survival as an endpoint. The nomograms are valid when survival is exponential and patients enter the study uniformly.

1. Considerations Affecting Sample Size

This paper presents nomograms for determining the number of patients necessary for a clinical trial when the endpoint is a time measurement that is exponentially distributed. Nomograms are given for one-sided and two-sided tests with significance level, α , equal to .05, and power, β , equal to .8 and .9. These values are those most often used in designing clinical trials although some authors (Brown, 1980; M. Zelen, personal communication) have suggested that lower values of α be used. A method is also presented for determining sample size when significance levels are .01 and .025, and when β differs from .8 or .9. There are three additional quantities which determine sample size. These are treatment difference, R , accrual time, A , and follow-up time, F .

A clinical trial is designed to have power β at a specified treatment difference. If m_1 and m_2 are the median survival times of patients on the two treatments then this difference can be expressed by $R = m_2/m_1$. Often R is chosen to be the minimum difference between the treatments that would represent substantial patient benefit, or R is chosen on the basis of data from other clinical trials of the same treatments. In many trials of cancer treatments R is chosen to be 1.5, since a 50% increase in survival is regarded as being clinically important and biologically feasible.

The accrual period is the period during which patients enter the trial. We assume that patients enter uniformly during this period. The follow-up period is the period from the entry of the last patient until the data are analyzed. The duration of both these periods determines the proportion of patients who will die before the data are analyzed.

2. Using the Nomograms

Figures 1 and 2 are used when the power is chosen to be .8 or .9, respectively. Figure 3 is a more complicated nomogram which can be used to find the sample size needed to achieve any power and to find power as a function of treatment difference. We will first describe how to use Fig. 1 and Fig. 2, which should be sufficient for most clinical applications.

Key words: Censoring; Failure time; Nomograms; Sample size; Survival.

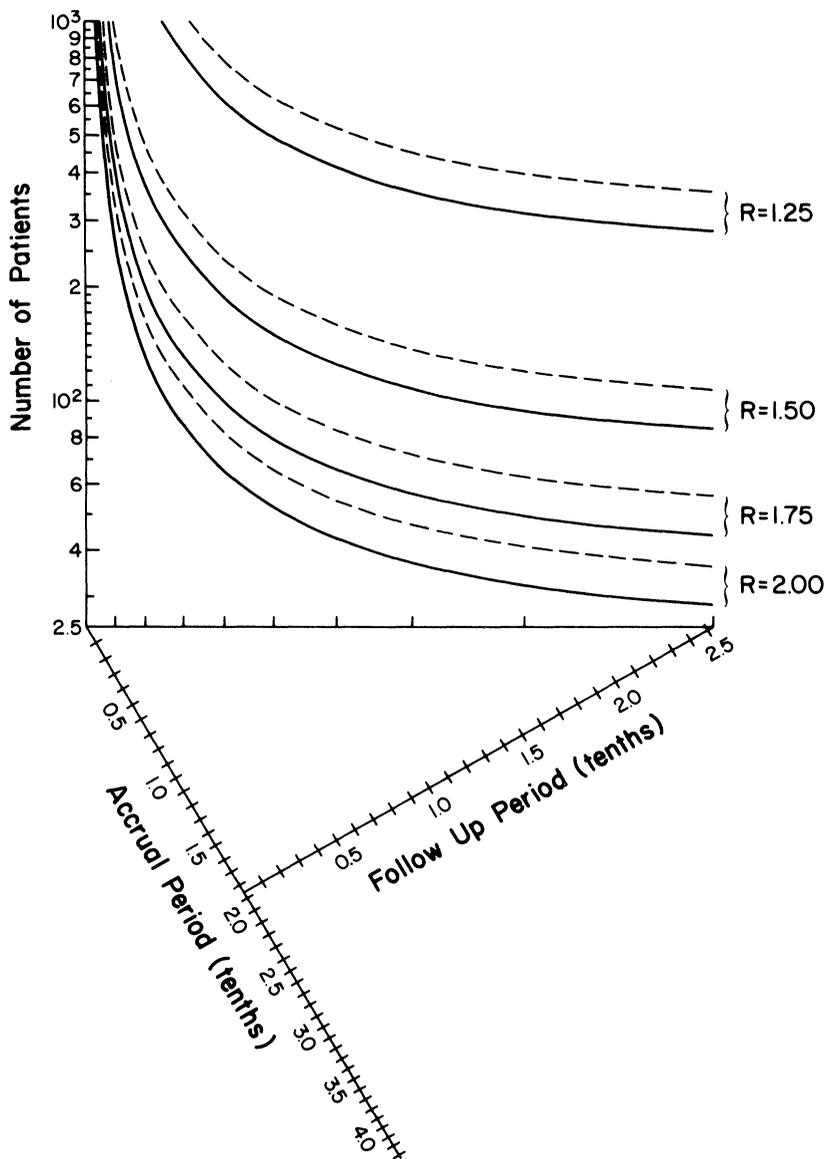


Figure 1. Number of patients per treatment group; $\alpha = .05$, $\beta = .80$. Dashed line: two-sided tests.
Solid line: one-sided tests.

To determine sample size, first calculate the average median survival of all patients on the trial under the alternative hypothesis, $\frac{1}{2}(m_1 + m_2)$, and divide the follow-up time and the accrual time by this number. Then mark these numbers on the 'accrual period' axis and 'follow-up period' axis in the lower part of the nomogram. Draw a line through the mark on the 'accrual period' axis parallel to the 'follow-up period' axis, and another line through the mark on the 'follow-up period' axis parallel to the 'accrual period' axis. These lines are shown in the example in Fig. 4. Draw a vertical line through the point where these two lines cross. Choose the appropriate treatment difference R and the direction of the test (one-sided or two-sided) and note where the vertical line that was just drawn

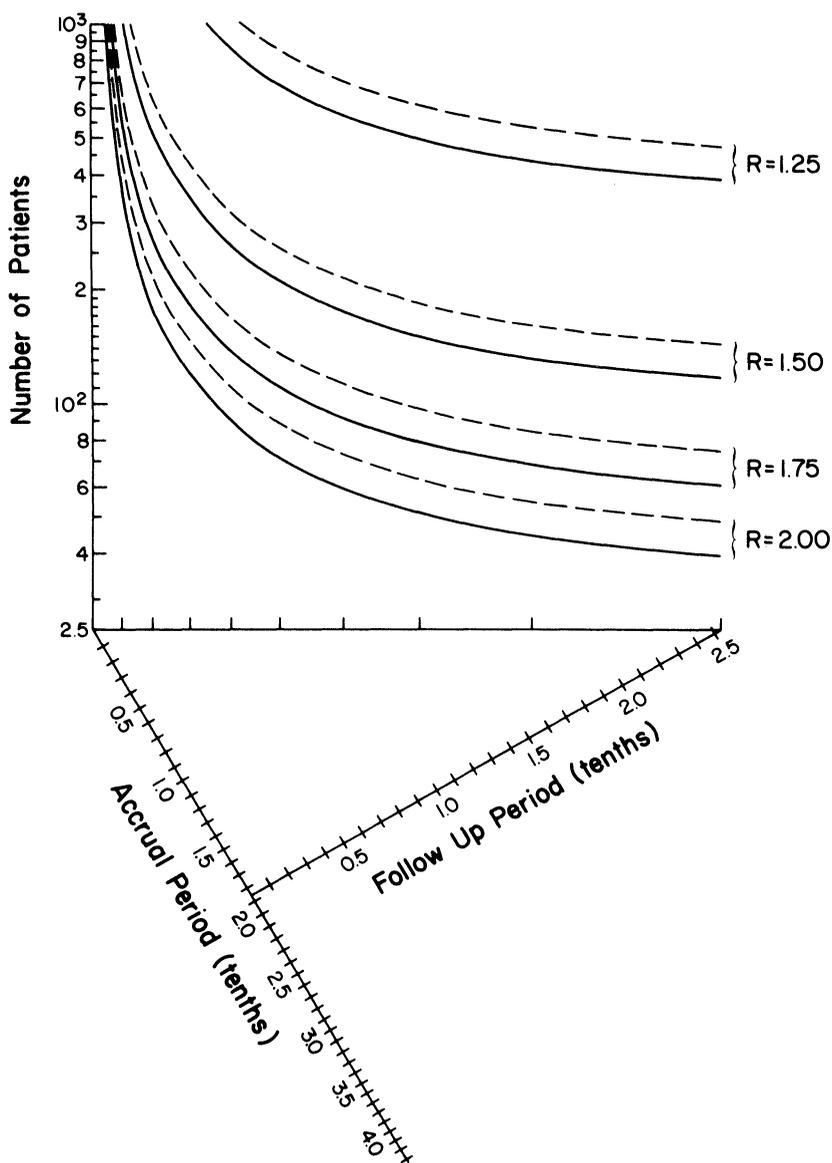


Figure 2. Number of patients per treatment group; $\alpha = .05$, $\beta = .09$. Dashed line: two-sided tests. Solid line: one-sided tests.

crosses the appropriate curve. The number of cases needed for each treatment group can then be read from the 'number of patients' scale directly across from this point. It is difficult to determine accurately sample sizes over 300 by use of the nomogram. However, the results beyond 300 should be accurate enough to determine the feasibility of a study.

As an example, suppose that a trial will permit accrual for two years and have one additional year of follow-up. Suppose that the median survival on the standard therapy is 11.0 months and that we want to detect an improvement of 50%, to 16.5 months. The accrual period is $24.0/13.8 = 1.7$ and the follow-up period is $12.0/13.8 = .87$. We mark these points on the triangle as shown in Fig. 4, and draw the lines indicated there.

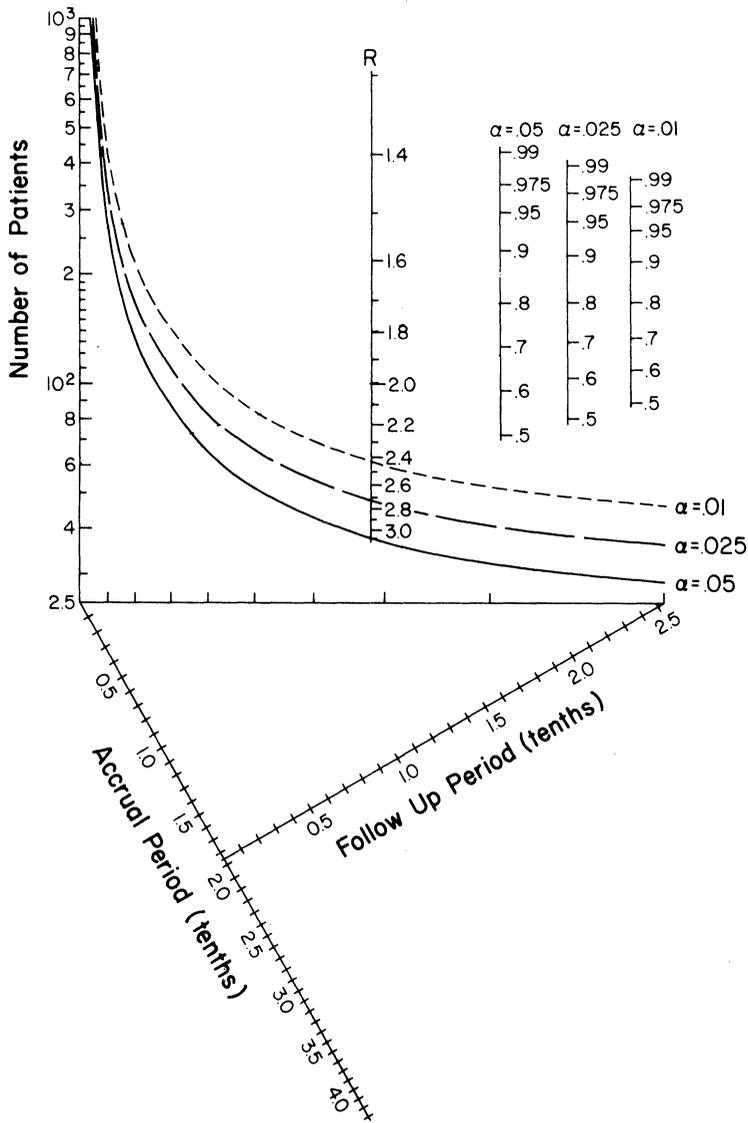


Figure 3. Number of patients per treatment group. The three vertical scales on the right give values of β corresponding to one-sided significance levels $\alpha = .05, .025$ and $.01$.

Assuming that we plan to perform a one-sided test, we use the solid curve labeled 'R = 1.5' and read off the sample size of 110 patients directly to the left of the point where it crosses the vertical line that has just been drawn.

To use Fig. 3 we start by following the previous procedure to determine sample size. The choice of curve is dictated by the significance level that we desire. The lowest curve is used for a one-sided level of .05 or a two-sided level of .10, the second curve for a one-sided level of .025 or a two-sided level of .05, and the third curve for a one-sided level of .01 or a two-sided level of .02. The number of patients found will be appropriate for $R = 2.0$ and $\beta = .8$. To find the number of patients for a different value of R we measure the distance between the new value of R and 2.0 on the R -scale, and then

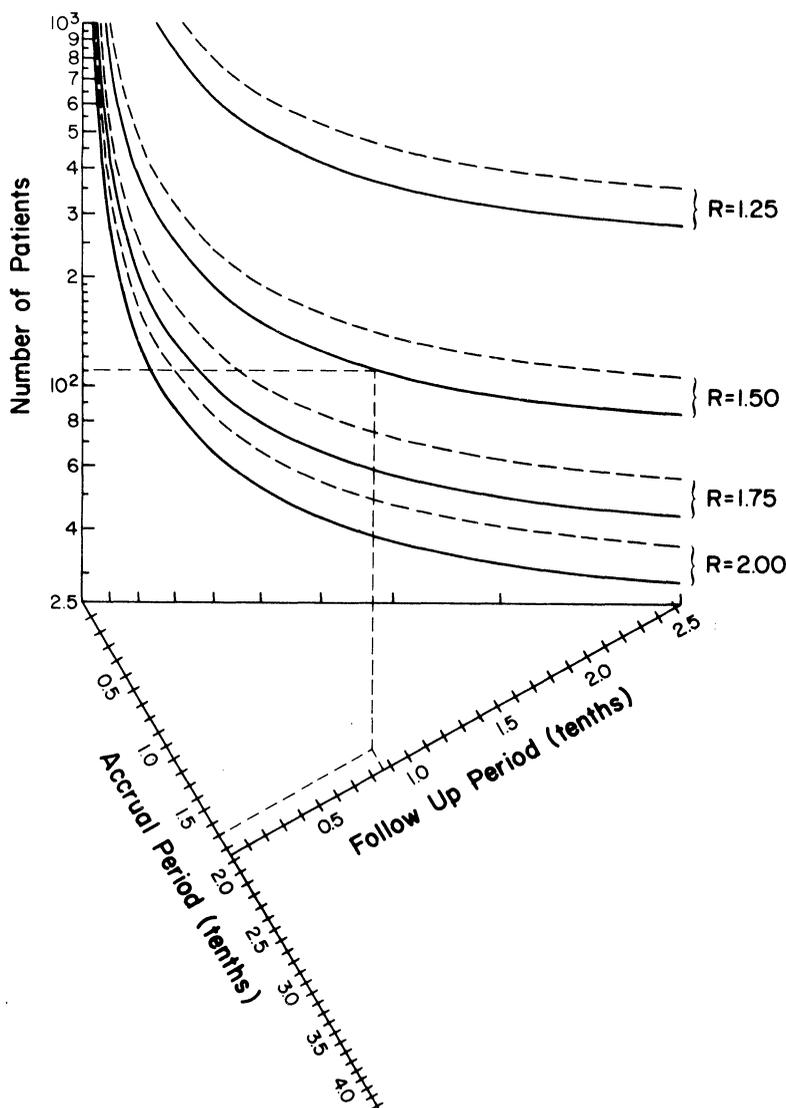


Figure 4. Example using Fig. 1.

transfer this distance to the 'number of patients' scale. For example, suppose 60 patients were required when $R = 2.0$. If R were 1.4, then 250 patients would be needed. This is shown by the dashed lines from the R -scale to the 'number of patients' scale in Fig. 5. To change the value of β we measure the distance from .8 on the α -scale (corresponding to the value of α chosen) to the desired value of β , then we transfer this distance to the 'number of patients' scale. This procedure is not illustrated in Fig. 5. To change β and R we apply these procedures sequentially.

The vertical scales in Fig. 3 can be used to investigate how changes in R will affect power. First, mark on a piece of paper the distance between the old value of R and the new value. Then set the point corresponding to the new value on the old value of β and read the new value off the scale. Remember that a decrease in R will yield a decrease in

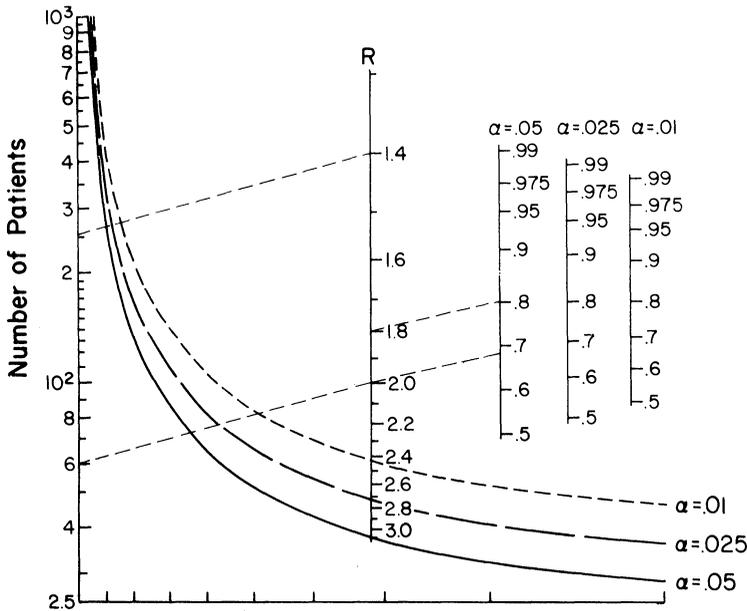


Figure 5. Example using Fig. 3 (without ‘accrual’ and ‘follow-up’ scales).

power. The dashed lines from the R -scale in Fig. 5 to the .05 scale show how a decrease in R from 2.0 to 1.8 will cause the power to drop from .8 to .68.

3. Justification of the Nomograms

Assume that each treatment has exponential survival with parameter θ_j . Let t_j be the total time on study and a_j be the number of deaths on Treatment j . Then the maximum likelihood estimate of θ_j is a_j/t_j and its variance is approximately θ_j^2/a_j (see Kalbfleisch and Prentice, 1980, pp. 48–54). Thus the variance of $\ln(a_j/t_j)$ is approximately $1/a_j$ (see Bickel and Doksum, 1977, p. 28) and $T = \ln(t_2 a_1 / t_1 a_2) (a_1^{-1} + a_2^{-1})^{-\frac{1}{2}}$ can be used to test whether Treatment 2 has better survival than Treatment 1.

Under the null hypothesis T will have a unit normal distribution. Let p_j be the probability that a randomly-selected patient accruing on Treatment j will die before the trial is analyzed. Then under the alternative hypothesis, for large n , the variance of T is one and the mean of T is $n^{\frac{1}{2}}(\ln R)(p_1^{-1} + p_2^{-1})^{-\frac{1}{2}}$. Let Φ be the normal distribution function. Then the power of the level- α test based on T , with n patients on each treatment, is approximately

$$\beta = \Phi\{n^{\frac{1}{2}}(\ln R)(p_1^{-1} + p_2^{-1})^{-\frac{1}{2}} - \Phi^{-1}(1 - \alpha)\}.$$

Thus, sample size is computed by

$$n = \{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\beta)\}^2 (\ln R)^{-2} (p_1^{-1} + p_2^{-1}).$$

Furthermore,

$$p_i = 1 - P_i(A)G_i(F),$$

where $P_j(A)$ is the probability that a patient lives to the end of the accrual period and $G_j(F)$ is the probability that a patient lives from the end of the accrual period to the end

of the follow-up period. These functions can be found by double integration with respect to an exponential density function for survival and uniform density function for censoring. They are

$$P_j(A) = \{1 - \exp(-.69A/m_j)\}/(.69A/m_j), \quad (1)$$

$$G_j(F) = \exp(-.69F/m_j). \quad (2)$$

Let $P(A)$ and $G(F)$ be (1) and (2) with $\frac{1}{2}(m_1 + m_2)$ substituted for m_j , and let p be $1 - P(A)G(F)$. The value of p will be between p_1 and p_2 , so we will approximate n by using $2p^{-1}$ as an approximation for $p_1^{-1} + p_2^{-1}$. These two numbers differ by less than 2% in the range of the nomogram. Thus

$$n \approx \{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\beta)\}^2 (\ln R)^{-2} (2p^{-1}).$$

The nomograms are constructed as follows: The bottom left scale is $-\ln P(A) \times 2$ and the right scale is $-\ln G(F) \times 2/\sqrt{3}$. Since the lower triangle is a 30° right triangle, the vertical line that is constructed will intersect the horizontal axis at the point

$$x = -\ln\{P(A)G(F)\} = -\ln(1 - p).$$

The tick marks on the horizontal axes are .1, .2, . . . , .9 and indicate this approximation to the probability of death during the trial. The curves in the figures are then computed by

$$n = \{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\beta)\}^2 (\ln R)^{-2} \times 2 \times (1 - e^{-x})^{-1}.$$

Figure 3 has an additional vertical scale which is $(\ln 2/\ln R)^2$ on the same scale as n . It has three horizontal scales which are $\{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\beta)\}^2 / \{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(.8)\}^2$ for $\alpha = .05$, .025 and .01. Multiplication by these factors is accomplished by addition, since n is plotted on a log scale.

4. Discussion

Since the logrank test (Peto and Peto, 1972; Mantel, 1966) is fully efficient, the nomograms are approximately correct when exponential survival data are to be analyzed by the logrank test, as well as when the data are to be analyzed parametrically.

Two other methods are often used to determine the sample size necessary for a clinical trial. One can calculate the number of deaths in each treatment group that will be required, rather than the number of patients. The formula is

$$d \approx 2\{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\beta)\}^2 (\ln R)^{-2}.$$

If one has decided to wait until 80% of the patients are dead to analyze the data, then one can divide d by .8 and use the resulting value as the sample size. This method is appropriate whenever the survival on both treatments is poor.

Often the treatments are compared in terms of the proportion of patients who die before a fixed time t_0 after their entry in the trial. This method has the advantage of not requiring the assumption of exponentiality. Letting p_j be the death rate on Treatment j , the required sample size is

$$n \approx \{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(\beta)\}^2 (p_1 - p_2)^{-2} \{p_1(1 - p_1) + p_2(1 - p_2)\}$$

if there is little censoring before t_0 . In exponential samples this procedure will require larger sample sizes than the method previously described, although the loss of efficiency will be slight if t_0 is near the end of the trial and the proportion of deaths at t_0 is less than .5.

ACKNOWLEDGEMENTS

This investigation was supported by Grants Numbers CA-25162 and CA-06516, awarded by the National Cancer Institute. The authors wish to thank Dr Marvin Zelen for suggesting this problem.

RÉSUMÉ

Cet article présente des nomogrammes pour le calcul de l'effectif de l'échantillon pour un essai clinique portant sur la survie. Les nomogrammes sont valides quand la survie est exponentielle et que les malades entrent dans l'étude uniformément.

REFERENCES

- Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day.
- Brown, B. W. (1980). Statistical controversies in the design of clinical trials—some personal views. *Controlled Clinical Trials* **1**, 13–27.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports* **50**, 163–170.
- Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society, Series A* **135**, 185–207.

Received June 1980; revised March 1981