Sample-Size Formula for the Proportional-Hazards Regression Model

David A. Schoenfeld

Biometrics, Vol. 39, No. 2 (Jun., 1983), 499-503.

Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28198306%2939%3A2%3C499%3ASFFTPR%3E2.0.CO%3B2-2

Biometrics is currently published by International Biometric Society.

BIOMETRICS 39, 499–503
June 1983

Sample-Size Formula for the Proportional-Hazards Regression Model

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

SUMMARY

A formula is derived for determining the number of observations necessary to test the equality of two survival distributions when concomitant information is incorporated. This formula should be useful in designing clinical trials with a heterogeneous patient population. Schoenfeld (1981, *Biometrika* **68**, 316–319) derived the asymptotic power of a class of statistics used to test the equality of two survival distributions. That result is extended to the case where concomitant information is available for each individual and where the proportional-hazards model holds. The loss of efficiency caused by ignoring concomitant variables is also computed.

1. Introduction

In the comparison of survival curves of treatment groups, the proportional-hazards regression model (Cox, 1972) is often used to adjust for covariates such as patient age, functional status and disease stage. It is shown that the formula for the sample size required for the comparison of two groups with exponential curves is valid when the proportional-hazards regression model is used to adjust for covariates. A brief tutorial on determining sample size is presented.

The hazard function evaluated at t is the instantaneous probability of death at a time, t, given survival up to that time. A patient's hazard function will depend on the treatment he or she receives as well as on the characteristics of the patient. Sometimes patients have a decreased probability of death after they survive past the first or second year, that is, the hazard function decreases. On the other hand, in long-term studies the hazard function increases as age increases the probability of death.

Suppose that there are two treatments, A and B. The proportional-hazards model specifies that the ratio of the hazard function of a patient given Treatment B to the same patient given Treatment A will be a constant, denoted by Δ, irrespective of time or the characteristics of the patient. Thus, one parameter specifies the effect of treatment. If survival is improved more by Treatment A than by Treatment B, Δ will be greater than 1. The assumption of proportional hazards is reasonable whenever the effect of treatment is constant over time or treatment permanently effects the disease process. If treatment has a transitory effect, then tests based on the proportional-hazards model should not be used and the sample-size formula given here is not valid. Tests appropriate to this situation have been given by Fleming *et al.* (1980).

2. Sample-Size Formula

The sample-size formula for a clinical trial can be simplified if it is expressed as the number of deaths required rather than as the number of patients. Suppose that a one-sided test will be performed with a significance level of α and a power of β when the hazard ratio is Δ_0. Let $z_{1-\alpha}$ and z_β be the $1-\alpha$ and β percentiles of the normal distribution, respectively, and let P_A and P_B be the proportion of the patients randomized to Treatments A and B, respectively.

Key words: Covariates; Logrank test; Power; Survival; Two-sample tests.

Assume that treatment effect is tested by an appropriate test based on the partial likelihood (Cox, 1972). Then the total number of deaths required is given by the following expression which is derived in the Appendix:

$$(z_\beta + z_{1-\alpha})^2/(P_A P_B \log_e^2 \Delta). \tag{1}$$

This is the same formula as that used to calculate sample size when two homogeneous patient groups are compared by using the F test for exponential survival (Bernstein and Lagakos, 1978), or when the logrank test is used to compare treatments with proportional hazards without covariates (Schoenfeld, 1981). However, this does not imply that covariate analysis is without benefit (see §3).

The number of patients required for a study is the same whether the randomization is stratified by covariate values or a simple randomization is used. The formula is not valid in a study where covariates are likely to be extremely unbalanced, such as a nonrandomized study.

2.1 *Determining the Proportion of Patients that Will Die*

Clinical trials have an accrual period, a, the period during which patients enter the study, and a follow-up period, f, the period from the end of accrual until the analysis of the data. The follow-up period substantially reduces the number of patients required in a clinical trial because without it, little information would be provided by patients who entered the trial near the end of the accrual period.

In a clinical trial with an accrual period, a, and a follow-up period, f, the proportion of patients that will survive is the average of the survival curve from Time f to Time $a + f$, provided that patients enter the trial at a constant rate. Thus if one has conducted a previous trial using Treatment B and has an estimate of the survival curve $S_B(t)$, one can use Simpson's rule to approximate the proportion of patients that will die on Treatment B:

$$d_B = 1 - \tfrac{1}{6}\{S_B(f) + 4S_B(f + .5a) + S_B(f + a)\}.$$

The proportion that will die on Treatment A can be approximated by $d_A = 1 - (1 - d_B)^{1/\Delta}$. Finally, the proportion dying in the trial is given by

$$d = P_A d_A + P_B d_B.$$

The number of patients required for the trial is equal to the number of deaths given by (1), divided by d.

This approximation for d_A is slightly conservative in that it underestimates the proportion of deaths on Treatment A. If the covariates divide the patient population into J groups with frequencies $U_1, U_2, \ldots, U_J$ and survival curves $S_{B1}(t), \ldots, S_{BJ}(t)$, then a better approximation would be

$$d_A = 1 - \tfrac{1}{6}\sum_{j=1}^{J} U_j\{S_{Bj}(f)^{1/\Delta} + 4S_{Bj}(f + .5a)^{1/\Delta} + S_{Bj}(f + a)^{1/\Delta}\}. \tag{2}$$

However, if $\Delta \leqslant 2$, $.1 \leqslant S_{Bj}(a + f)$ and $S_{Bj}(f) \leqslant .9$, then the error in the approximate formula is less than $.16\, d_B$.

If the survival is exponential within each group with the same covariates, a more precise calculation is possible. If the median survival of a subgroup on Treatment B is t, the proportion expected to die will be $1 - \{\exp(-.69f/t)\}\{1 - \exp(-.69a/t)\}/(.69a/t)$. Then d_B will be the average of these values weighted by the proportions of patients in each prognostic subgroup. The median survival on Treatment A in a subgroup will be Δt, and d_A can be calculated by using the same formula.

2.2. *Using the Nomograms to Determine Sample Size*

The nomograms given by Schoenfeld and Richter (1982) can be used to determine sample size when covariates are to be included in the analysis. To use the nomograms, simply calculate the proportion of the patients predicted to die, as in §2.1. The tick marks on the horizontal scale of the nomogram correspond to the proportions of patients that will die in a clinical trial. The first corresponds to 10%, the second to 20%, and so on. Therefore, mark the percentage that will die on the horizontal scale of the proper nomogram; draw a vertical line from this point to the graph with $R = \Delta$; then draw a horizontal line from this point of intersection to the vertical scale and read the number of patients required per arm.

3. The Power Advantages of Adjusting for Covariates

Even though the formula for sample size is the same whether covariates are adjusted for or not, the powers of the two procedures are different. If the two treatment groups follow the proportional-hazards regression model, then, if covariates are ignored, the ratio of the hazard functions of the two groups will be nonproportional. This ratio will be less than Δ at every value of $t > 0$ and the power of any test without covariates will be less than that of the test that uses covariates.

As an example, suppose that there is a binary covariate which divides the patient population into two equal groups. Patients in Treatment Group B have a median survival of two years in one group and of six months in the other. Survival is exponential and $\Delta = 1.5$, so patients in Group A have a median survival of three years in one group and of nine months in the other. The ratio of the hazard functions at the start of the study will be 1.5, however by the second year, the hazard ratio will have dropped to 1.27. In a clinical trial conducted with two years of accrual and two years of follow-up, the logrank test will have an efficiency of 61% when compared to a test that uses the covariate. For the calculation involved, see Schoenfeld (1981). If the hazard ratio were not constant and if we used the optimal rank test, its efficiency would only be 63%. Thus, the use of covariates can substantially increase power when the proportional-hazards model holds.

4. Example

The Radiation Therapy Oncology Group is conducting a series of trials on the treatment of primary brain cancer. The first study, completed in 1978, showed an advantage of chemotherapy and radiation therapy over radiation therapy alone. Subsequent studies are testing whether the addition of a radiation sensitizer or the use of neutron radiotherapy improves survival. In this example, it is shown how the data from the first study can be used to plan a new study.

The new study would use chemotherapy and radiation therapy for its standard arm. A

Table 1
Prognostic subgroups in primary brain tumors

Group	Number of patients	Median survival (months)
Age $\leq$ 40, no necrosis	21	29.1
Age $\leq$ 40, necrosis	22	15.7
40 < Age $\leq$ 60, no necrosis	24	26.5
40 < Age $\leq$ 60, necrosis	125	9.3
Age > 60, no necrosis	5	7.9
Age > 60, necrosis	75	5.0

total of 272 patients were treated with this combination in the first study. The most important prognostic factors were age and whether the patients had necrosis in their surgical specimens. Table 1 shows the numbers of patients in three age groups and two necrosis groups. Of the 272 patients, 43% survived past one year, 20% past two years, and 11% past three years. For a study with an accrual period of two years and one year of additional follow-up, the proportion of deaths is computed to be $1 - \frac{1}{6}\{.43 + 4\,(.2) + .11\} = .78$. If the hazard ratio of the old treatment to the new treatment were 1.5, approximately $1 - .22^{(1/1.5)} = .64$ of the patients on the new treatment would die. Thus, $d = \frac{1}{2}(.64 + .78) = .71$. If one uses the approximation (2), one finds $d = .73$. To achieve a power of 80% with a one-sided significance level of 5%, $(1.645 + .841)^2/[.25\ \{\ln(1.5)\}^2\ .71] = 212$ patients would be required.

If covariates were ignored in subsequent studies, the efficiency of a logrank test for comparing treatments would only be 67%. Even if the optimal rank test were used, its efficiency would only be 70%. Thus, the use of these covariates substantially reduces the sample size needed to achieve adequate power.

Acknowledgements

This work was funded by Grants CA-25162 and CA-23415 of the National Cancer Institute. I wish to express my appreciation to the Radiation Therapy Oncology Group for providing the example.

Résumé

Il s'agit de déterminer le nombre d'observations nécessaires pour tester l'égalité de deux distributions de survie quand on incorpore l'information concomitante. Cette formule peut être utile à la planification d'essais cliniques quand la population des patients est hétérogène. Schoenfeld (1981, *Biometrika* **68,** 316–319) a calculé la puissance asymptotique d'une classe de statistiques utilisées pour tester l'égalité de deux distributions de survie. Ce résultat est étendu au cas où on peut connaître l'information concomitante pour chaque individu et dans le cas d'un model de hasard proportionnel. On détermine la perte d'efficacité quand on ignore les variables concomitantes.

References

Bernstein, D. and Lagakos, S. W. (1978). Sample size and power determination for stratified clinical trials. *Journal of Statistical Computing and Simulation* **8,** 65–73.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–279.

Fleming, T. R., O'Fallon, J. R., O'Brien, P. C. and Harrington, D. P. (1980). Modified Kolmogorov–Smirnov test procedures with application to arbitrarily right-censored data. *Biometrics* **36,** 607–625.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* New York: Wiley.

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

Schoenfeld, D. and Richter, J. (1982). Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics* **38,** 163–170.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Annals of Statistics* **9,** 93–108.

Received June 1981; *revised December* 1981 *and May* 1982

Appendix

Derivation of Sample-Size Formula for the Score Test

Since the usual tests of treatment effect, when covariates are adjusted for, are asymptotically equivalent, I shall derive a sample-size formula for the score test. The argument follows Schoenfeld (1981).

Let $j = 1, \ldots, n$ identify the patients in the trial. For each patient assume that there is a vector $\mathbf{y}_j = (y_{j1}, \ldots, y_{jp})'$ of covariates. For convenience, relabel the treatments 0 and 1 and let x be the treatment label. Assume that the probability that Patient j receives Treatment x is P_x which is independent of $\mathbf{y}_j$.

This allows stratification based on y_j, as long as the proportion randomized to each treatment is the same in each stratum. Assume that the hazard function for the jth patient is given by

$$\lambda_j(t) = \lambda_0(t)\exp(a_0 x_j + \sum a_i y_{ji}),$$

where the summation is over the set $i = 1, p$, and $a_0 = \log_e \Delta$.

To define the score statistic (Rao, 1973, p. 417), let D be the set of identifiers of those patients who die, let t_j be the death time for the jth patient in D, and let x_j denote the patient's treatment label. Assume the t_j are distinct. Let $R(t_j)$ be the set of identifiers of patients being observed at Time $t_j - 0$. For any function $g(x, \mathbf{y})$, define

$$E_j\{g(x, y)\} = \frac{\sum\limits_{k\in R(t_j)} g(x_k, \mathbf{y}_k)\exp(\Sigma a_m y_{km})}{\sum\limits_{k\in R(t_j)} \exp(\Sigma a_m y_{km})}$$

and let $\hat{E}_j$ be E_j with maximum likelihood estimates (assuming $a_0 = 0$) replacing the parameters $\{a_m\}$.

Letting y_i be the ith compound of $\mathbf{y}$, we define the elements of the $p \times 1$ vector $\mathbf{B}$ by

$$\hat{B}_i = n^{-1} \sum_{j\in D} \{\hat{E}_j(xy_i) - \hat{E}_j(x)\hat{E}_j(y_i)\}$$

and we define the elements of the $p \times p$ matrix $\mathbf{M}$ by

$$\hat{M}_{ik} = n^{-1} \sum_{j\in D} \{\hat{E}_j(y_i y_k) - \hat{E}_j(y_i)\hat{E}_j(y_k)\}.$$

The score statistic can then be expressed as

$$\frac{n^{-\frac{1}{2}}\left[\sum\limits_{j\in D} \{x_j - \hat{E}_j(x)\}\right]}{\left(n^{-1}\left[\sum\limits_{j\in D} \hat{E}_j(x)\{1 - \hat{E}_j(x)\}\right] - \hat{\mathbf{B}}'\hat{\mathbf{M}}^{-1}\hat{\mathbf{B}}\right)^{\frac{1}{2}}}.$$

The term $\hat{\mathbf{B}}'\hat{\mathbf{M}}^{-1}\hat{\mathbf{B}}$ is the effect of the estimation of $a_1, a_2, \ldots, a_p$ on the variance of $x_j - \hat{E}_j(x)$.

Assume that a_0 is $O(n^{-\frac{1}{2}})$. At the start of the trial, the distribution of vectors $\mathbf{y}$ will be the same in the two treatment groups. Since $a_0 \to 0$, this will remain true for any time, t, so $\hat{E}_j(xy_i) \xrightarrow{p} \hat{E}_j(x)\hat{E}_j(y_i)$. Thus $\hat{\mathbf{B}} \xrightarrow{p} 0$ and the second term in the denominator of S can be ignored. The term $\hat{\mathbf{B}}$ appears in the Taylor expansion of $\hat{E}(x)$ about $a_1, a_2, \ldots, a_m$, which implies that

$$n^{-\frac{1}{2}} \sum_{j\in D} \{\hat{E}_j(x) - E_j(x)\} \xrightarrow{p} 0.$$

Thus S can be written as

$$S = \frac{n^{-\frac{1}{2}} \sum\limits_{j\in D} \{x_j - E_j(x)\}}{\left[n^{-1} \sum\limits_{j\in D} E_j(x)\{1 - E_j(x)\}\right]^{\frac{1}{2}}}.$$

Define

$$e_j = \left\{\sum_{k\in R(t_j)} x_k \exp\left(a_0 x_k + \sum_{m=1}^{p} a_m y_{km}\right)\right\} \Big/ \left\{\sum_{k\in R(t_j)} \exp\left(a_0 x_k + \sum_{m=1}^{p} a_m y_{km}\right)\right\}.$$

The numerator of S can be written

$$n^{-\frac{1}{2}} \sum_{j\in D} \{x_j - E_j(x)\} = n^{-\frac{1}{2}} \sum_{j\in D} (x_j - e_j) + n^{-\frac{1}{2}} \sum_{j\in D} \{e_j - E_j(x)\}.$$

The first term is asymptotically normal with mean 0 and variance $n^{-1} \sum e_j(1 - e_j)$, where the summation is over D (Cox, 1975; Tsiatis, 1981). Expanding the second term in a Taylor series about $a_0 = 0$, we find that this term approaches $a_0 n^{-\frac{1}{2}} \sum E_j(x)\{1 - E_j(x)\}$. However, since $a_0 \to 0$, both e_j and $E_j(x)$ approach P_1. Thus S is asymptotically normal with unit variance and mean equal to $a_0(P_1 P_2)^{\frac{1}{2}}$ times the square root of the expected number of deaths on the trial. This yields (1).