

SAMPLE SIZE AND DURATION FOR COHORT STUDIES  
OF SURVIVAL TIME WITH COVARIABLES

by

John D. Taulbee  
The Proctor and Gamble Company  
Miami Valley Laboratories, Cincinnati, Ohio 45247

and

Michael J. Symons  
Department of Biostatistics  
University of North Carolina  
Chapel Hill, NC 27514

Institute of Statistics Mimeo Series No. 1300

AUGUST 1980

11-13-79

Sample Size and Duration for Cohort Studies  
of Survival Time with Covariables

JOHN D. TAULBEE

The Procter and Gamble Company, Miami Valley Laboratories  
Cincinnati, Ohio 45247, U.S.A.

MICHAEL J. SYMONS

Department of Biostatistics, University of North Carolina  
Chapel Hill, North Carolina 27514, U.S.A.

Summary

Determination of sample size and duration of cohort studies with covariables is addressed. An exponential model, using the Feigl and Zelen (1965) form for the hazard with covariates and asymptotic normality of the maximum likelihood estimators of the parameters are assumed. Emphasis is on applications involving two parameters, an underlying hazard and the coefficient of a single concomitant variable. The results of George and Desu (1974) are reproduced and their model assumptions clarified. An example with more than two dose groups is presented. For situations where the sample size is fixed, extension of the follow-up time is considered for achieving the desired type I and type II errors for a null hypothesis and specified alternative of interest. Further, generalization to situations with other forms for the hazard rate is indicated.

---

Key Words: Sample size; Duration of survival time study; Survival time with covariables.

## 1. Introduction

Most of the work in determining length of follow-up and sample size for cohort studies has been done in regard to controlled clinical trials. In fact, Feinstein (1975) claims the "favorite indoor sport" of statistical theoreticians is contriving new ways to gauge sample size for controlled clinical trials.

Virtually all this work has assumed control and experimental groups of equal size to be compared. For a test of the null hypothesis

$H_0: p_1 = p_2$  versus the alternative  $H_1: p_1 \neq p_2$ , Schlesselman (1974a)

gives the well known formula for sample size of each of two groups

$$N = (p_1 - p_2)^{-2} \left\{ z_\alpha \sqrt{2\bar{p}\bar{q}} + z_\beta \sqrt{p_1(1 - p_1) + p_2(1 - p_2)} \right\}^2,$$

where  $p_1$  is probability of event given no exposure,  $p_2$  is probability of event given exposure,  $\bar{p}$  is the average of  $p_1$  and  $p_2$  and  $\bar{q} = 1 - \bar{p}$ , and  $\alpha$  and  $\beta$  are probabilities of type I and II errors, respectively. In another paper, Schlesselman (1974b) gives the tables of required sample size for various values of  $p_1$ ,  $p_2/p_1$ , and type I and II errors.

Pasternack (1972) assumes a constant hazard for both case and control groups, and an independent constant hazard for loss to follow-up and produces tables which yield sample size for each group given five-year survival rate for control, hypothesized increase in survival rate for cases, and type I and II errors of interest. In regard to length of follow-up, Pasternack and Gilbert (1971) make similar assumptions to those of Pasternack (1972) above, and give tables of required duration of study

based on anticipated median survival time for cases and controls, total number of subjects to be entered each year, and type I and II errors desired.

Schork and Remington (1967) allow for the possibility that individuals may move back and forth between treatment and control groups during the study period. For a study  $L$  time units long, they let  $p_1$  denote the probability of event in a single time unit given no exposure (control);  $p_2 = kp_1$  denote the probability of event in a single time unit given exposure (treatment) and  $R_t$  and  $R'_t$  denote the proportions shifting from the control and treatment groups to the other group, respectively, in time unit  $t$ ,  $t = 1, 2, \dots, L$ . They then provide a formula for determining sample size for the desired values of  $L$ ,  $p_1$ ,  $p_2$ , the  $R_t$  and  $R'_t$ ,  $t = 1, 2, \dots, L$  and type I and II errors.

Halperin, Rogot, Gurian, and Ederer (1968) assume a constant hazard rate and no drop-out for the control group. For the treatment group, they assume a hazard rate which declines linearly over  $f$  years to a constant rate and also a constant drop-out rate. They then produce tables of required sample size for specified values of trial duration,  $f$ , type I and II errors, and the three rates mentioned above.

George and Desu (1974) assume a constant hazard  $\lambda_C$  for the control group and a possibly different hazard  $\lambda_E$  for the experimental (treatment) group. They then provide a table for the sample size required, given values  $\lambda_C / \lambda_E$  and type I and II error probabilities desired.

The particular problem of interest here, however, is for the case where more than two treatment groups are involved. For instance, we might be concerned with an animal study having a zero dose group and two or more non-zero dose groups. We propose that time until the event of interest be modelled as exponential for each group. The difference among groups, if any, is accounted for by the inclusion of a covariate. The hazard rate for any subject would then be

$$\lambda e^{Bz_i} \tag{1}$$

where  $\lambda$  is the underlying hazard and the covariate  $z_i$  reflects the dose or concentration given to subject  $i$ ,  $i = 1, 2, \dots, N$ . The model (1) was originally suggested by Feigl and Zelen (1965) and was generalized by Glasser (1967) and Taulbee (1979).

The test for differences among the survival curves of the three or more groups, i.e. the test of dose effect, would be the test of the null hypothesis  $H_0: B = 0$  versus the alternative  $H_1: B \neq 0$ . If we are interested in two groups only, the covariate  $z_i$  becomes a group labeling device and our model simplifies to produce the sample size result of George and Desu (1974).

The present method utilizes the dependency of the Fisher information matrix upon the total sample size, the censoring time, and the asymptotic distribution of the maximum likelihood estimators of model parameters to obtain the needed results. The test of hypothesis of interest is the test for equality of the survival curves, which is available by testing whether  $B$  in (1) is zero or not.

Extension of the theory to accommodate tests of hypothesis involving more than one parameter is given by Taulbee and Symons (1979).

## 2. Asymptotic Distribution of the Estimators of the Model Parameters

The approach here is to use model (1) for the hazard rate with covariables and to assume the asymptotic normality of the maximum likelihood estimators (MLEs) of the required model parameters in order to obtain sample size requirements for testing specific hypotheses about the parameter of interest. The MLEs of the model parameters are then asymptotically unbiased with the variance-covariance matrix given by the inverse of the Fisher information matrix. It will be shown that the variance of the MLE of the model parameter of interest can be expressed as a multiple of  $1/N$ , where  $N$  is the cohort size. With probability of type I error,  $\alpha$ , and probability of type II error,  $\beta$ , specified, the value of  $N$  is determined as the smallest integer resulting in  $\alpha$  and  $\beta$  being no larger than specified, presuming the asymptotic distribution results are adequate approximations. Alternatively, if sample size is fixed, one may be able to alter the distribution of censoring times iteratively until the power requirement is satisfied.

This approach is useful for any situation in which estimation and hypothesis testing for a parametric model is done by maximum likelihood techniques. There is no restriction as to the number of parameters involved nor must the model necessarily be for survival time. However, for the specific problem at hand, we use an exponential model with hazard given by (1). The logarithm of the likelihood for the model is then

$$\ln L = \sum_{i=1}^N [I_i(\ln \lambda + Bz_i) - \lambda \exp(Bz_i)t_i] , \quad (2)$$

where for individual  $i$ ,  $i = 1, 2, \dots, N$ ,  $z_i$  is the covariate value,  $t_i$  is the censoring time for survivors, for whom  $I_i = 0$ , and  $t_i$  is the time of death for failures, for whom  $I_i = 1$ . Evaluating the corresponding first and second partial derivatives with respect to the parameters  $\lambda$  and  $B$ , we find that the Fisher information matrix is given by

$$V^{-1} = E \begin{bmatrix} \sum_{i=1}^N I_i \lambda^{-2} & \sum_{i=1}^N \exp(Bz_i) t_i z_i \\ \sum_{i=1}^N \exp(Bz_i) t_i z_i & \sum_{i=1}^N \lambda \exp(Bz_i) t_i z_i^2 \end{bmatrix}, \quad (3)$$

where the expectation is with respect to the random variables  $t_i$ ,  $i = 1, 2, \dots, N$ , and  $N$  is the number of individuals in the cohort, or for our purposes, the sample size.

Let

$$D = E [\lambda \exp(Bz_i) t_i z_i^2]. \quad (4)$$

We may now express the lower right-hand element of the Fisher information matrix (3) as  $ND$ , so  $1/(ND)$  is the variance of the MLE of  $B$  of (1). Note that  $D$  will depend on  $\lambda$ , the underlying hazard. Values to be used for  $\lambda$  in the sample size determination must be estimated from prior experience. For example, one could use the fact that the inverse of  $\lambda$  is the mean time until occurrence of the event of interest in the zero dose group.

The general method for evaluation of  $D$  of (4) is demonstrated in the Appendix for two cases of interest. In both cases, the covariate is assumed to be discretely distributed with possible values  $Z_j$ ,  $j = 1, 2, \dots, J$  and  $\Pr(Z_j) = p_j$ . In the first case, censoring time is also

assumed to be discretely distributed with possible values  $T_k$ ,  $k = 1, 2, \dots, K$  and  $\Pr(T_k | Z_j) = p_{jk}$ . In this case, we obtain

$$D = \sum_{j=1}^J p_j Z_j^2 \sum_{k=1}^K p_{jk} [1 - \exp(-\lambda e^{BZ_j} T_k)] . \quad (5)$$

In the second case, censoring time  $T_k$  is assumed to be continuous and exponentially distributed, with

$$f(T_k | Z_j) = \mu e^{CZ_j} \exp(-\mu e^{CZ_j} T_k)$$

and we obtain

$$D = \sum_{j=1}^J p_j Z_j^2 [\lambda \exp(BZ_j)] / [\lambda \exp(BZ_j) + \mu \exp(CZ_j)] . \quad (6)$$

Expressions (5) and (6) are derived in the Appendix.

Now, if we wish to test a hypothesis about  $B$ , the Fisher information is  $ND$ . So, if  $\hat{B}$  is the MLE of  $B$ , we have

$$(ND)^{1/2} (\hat{B} - B) \sim N(0, 1),$$

from large sample theory. This result can be used to determine approximately the required sample size for a test of size  $\alpha$  ( $0 < \alpha < 1$ ) requiring power of at least  $1 - \beta$  ( $0 < \beta < 1$ ) against a specific alternative. In all of the determinations,  $B$  and  $D$  will be subscripted with a zero (0) or one (1) to denote that the value is that under the null or specific alternative hypothesis of interest, respectively.



3. Formulas for Determination of Sample Size for Tests of Hypothesis about the Parameter of Interest

We wish to have a test of size  $\alpha$  for the hypothesis that  $B = B_0$  with power at least  $1-\beta$  when  $B = B_1$ , a one-sided or two-sided alternative hypothesized value. We now define  $z_\gamma$  to be such that for  $0 < \gamma < 1$ ,  $\Pr(z \leq z_\gamma) = \gamma$ , where  $z$  is a standard normal variate.

Suppose we want to test the null hypothesis  $H_0: B = B_0$  against the alternative  $H_1: B > B_0$ . We reject  $H_0$  if  $(ND_0)^{1/2} (\hat{B} - B_0) > z_{1-\alpha}$ . We wish to reject  $H_0$  with probability at least  $1-\beta$  if, in fact,  $B = B_1, B_1 > B_0$ . The power requirement can be stated as

$$\Pr[(ND_0)^{1/2}(\hat{B} - B_0) > z_{1-\alpha} | B = B_1] \geq 1 - \beta . \quad (7)$$

Algebraic manipulation of (7) gives the required sample size as

$$N \geq (B_1 - B_0)^{-2} (z_{1-\alpha} D_0^{-1/2} + z_{1-\beta} D_1^{-1/2})^2 . \quad (8)$$

If we want the alternative hypothesis to be  $H_1: B < B_0$  with specific alternative of interest  $B = B_1, B_1 < B_0$ , then the sample size requirement is still given by (8).

Now, for the two-sided alternative hypothesis, specifically  $H_1: B \neq B_0$ , we also require a test of size  $\alpha$  with power at least  $1-\beta$  against the specific alternative  $B = B_1$ . We reject  $H_0$  if

$$(ND_0)^{1/2} (\hat{B} - B_0) < z_{\alpha/2} \text{ or } (ND_0)^{1/2} (\hat{B} - B_0) > z_{1-\alpha/2} .$$

If we let

$$x_L = z_{\alpha/2} (D_1/D_0)^{1/2} + (ND_1)^{1/2} (B_0 - B_1) \quad (9)$$

and

$$x_U = z_1 - \alpha/2 (D_1/D_0)^{1/2} + (ND_1)^{1/2} (B_0 - B_1) , \quad (10)$$

then the power requirement can be shown to be that N must be of sufficient size so that

$$\Pr(z < x_L) + \Pr(z > x_U) \geq 1 - \beta \quad (11)$$

where  $x_L$  is given by (9),  $x_U$  by (10), and  $z$  is a standard normal variate.

Note that an explicit solution of (11) for N is not given. The proper value of N is the smallest positive integer resulting in values of  $x_L$  and  $x_U$  which satisfy (11), and this value must be found by trial and error in the general case. However, we will frequently find in practice that either  $\Pr[z < x_L]$  or  $\Pr[z > x_U]$  will be virtually zero, in which case an explicit solution for N can be found.

#### 4. Examples

Let us consider the situation given in Sections 2 and 3 of George and Desu (1974). There, constant hazard rates  $\lambda_C$  for the control group and  $\lambda_E$  for the experimental group are assumed with  $\lambda_C/\lambda_E = \Delta$  and no censoring. There are at least two distinct cases leading to  $\lambda_C/\lambda_E = \Delta$ . We first consider the situation implicitly assumed in George and Desu (1974), referred to as case 1. Let  $\lambda_C = \lambda \Delta^{1/2}$  and  $\lambda_E = \lambda \Delta^{-1/2}$ , where  $\lambda > 0$  need not be specified. Then  $\lambda_C/\lambda_E = \Delta$  and under  $H_0: \Delta = 1$ ,  $\lambda_C = \lambda_E = \lambda$ . Under  $H_1: \Delta > 1$ , we have  $\lambda_E < \lambda < \lambda_C$ . Required sample sizes in each group for various levels of  $\Delta$ ,  $\alpha$ , and  $1 - \beta = \text{power}$  are given in Table 1 of George and Desu, reproduced as Table 1 below.

Table 1. Number of Patients (d) Required to Detect a Significant Difference in Two Survival Distributions, from George and Desu (1974).

$1 - \beta$	$\Delta$													
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.5	3.0	3.5	4.0
0.95	3479	951	460	280	193	144	113	92	77	66	38	27	21	17
	2384	652	315	192	132	98	77	63	53	46	26	18	14	12
0.90	2870	785	379	231	159	118	93	76	64	55	32	22	17	14
	1884	515	249	152	105	78	61	50	42	36	21	15	11	9
0.80	2213	605	292	178	123	91	72	59	49	42	24	17	13	11
	1360	372	180	110	76	56	44	36	30	26	15	11	8	7

Notes. (1)  $\Delta$  is the ratio of the median survival for the experimental treatment group to the median survival for the control treatment group. (The ratio of means could be substituted for the ratio of medians).

(2)  $1 - \beta$  is the probability of detecting a significant difference at the 0.01 level (upper figure) or the 0.05 level (lower figure) when the true ratio is  $\Delta$ .

(3) d is the number of patients required in each treatment group, all followed until failure.

Model (1) contains this as a special case. Let  $Z_1 = 1$  for the control group,  $Z_2 = -1$  for the experimental group. Then  $\lambda_C = \lambda e^B$  and  $\lambda_E = \lambda e^{-B}$ , so  $B = (\ln \Delta)/2$ . Also, we have the null hypothesis

$H_0: B = 0$  and the alternative  $H_1: B > 0$ .

Since there is no censoring, we set  $T_k$  of (5) at infinity. We desire equal sample sizes in each group, so  $p_1 = p_2 = 1/2$ . We have  $Z_j^2 = 1$ ,  $j = 1, 2$  so from (5)  $D_0 = D_1 = 1/2(1) + 1/2(1) = 1$ . We have  $B_0 = 0$  and  $B_1 = (\ln \Delta)/2$ , so from (8)

$$N \geq (2/\ln \Delta)^2 (Z_{1-\alpha} + Z_{1-\beta})^2. \quad (12)$$

Therefore, the sample size required in each of the two equal sized groups for case 1 is

$$d = 2 (\ln \Delta)^{-2} (Z_{1-\alpha} + Z_{1-\beta})^2. \quad (13)$$

In Table 2, sample sizes computed using (13) are given in the same format as Table 1. Numbers in Table 2 which are underlined differ from those given in Table 1. Comparison of Tables 1 and 2 reveals that these differences are not large, and that the approximate values given by (13) are accurate enough so that there is little need to resort to more precise iterative techniques. Note that (13) is equivalent to (4) of George and Desu (1974).

Table 2

Number of Patients (d) Required per Group to Detect a Significant Difference in Two Survival Distributions, as Calculated from (13).

$1 - \beta$	$\Delta$													
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.5	3.0	3.5	4.0
0.95	<u>3472</u>	<u>949</u>	<u>459</u>	<u>279</u>	<u>192</u>	<u>143</u>	113	92	77	66	38	27	21	17
	2384	652	315	192	132	98	77	63	53	46	26	18	14	12
0.90	<u>2867</u>	<u>784</u>	379	<u>230</u>	159	118	93	76	64	55	32	22	17	14
	<u>1887</u>	<u>516</u>	249	152	105	78	61	50	42	36	21	15	11	9
0.80	<u>2210</u>	<u>604</u>	292	178	123	91	72	59	49	42	24	17	13	11
	<u>1362</u>	<u>373</u>	180	110	76	56	44	36	<u>31</u>	26	15	11	8	7

Notes. (1)  $\Delta$  is the ratio of the median survival for the experimental treatment group to the median survival for the control treatment group. (The ratio of means could be substituted for the ratio of medians).

(2)  $1 - \beta$  is the probability of detecting a significant difference at the 0.01 level (upper figure) or the 0.05 level (lower figure) when the true ratio is  $\Delta$ .

(3) d is the number of patients required in each treatment group, all followed until failure.

(4) Numbers underlined differ from those in Table 1.

Notice that for case 1 above, under  $H_0$ ,  $\lambda_C = \lambda_E = \lambda$  while under  $H_1$ ,  $\lambda_E < \lambda < \lambda_C$ . Thus the control group survival curve is not the same under  $H_0$  and  $H_1$ . In many studies, it would be of greater interest to have under  $H_1$ ,  $\lambda_E < \lambda = \lambda_C$ . That is, for case 2, the control group survival curve is the same under  $H_0$  and  $H_1$ . The experimental group survival curve is the same as for control under  $H_0$  but differs from control under  $H_1$ . The case 2 situation is applicable where with two groups, randomly chosen, one is given treatment such as a placebo not expected to alter survival (control group) and the other is given treatment which may alter survival (experimental group). Case 2 is then an experimental situation such as an animal experiment or clinical trial, while case 1 assumptions are more relevant to observational studies.

To adapt the George and Desu (1974) model to case 2, we let  $\lambda_C = \lambda$  and  $\lambda_E = \lambda/\Delta$ . Then we have  $\lambda_C/\lambda_E = \Delta$  and under  $H_0$ :  $\Delta = 1$ ,  $\lambda_C = \lambda_E = \lambda$ . Under  $H_1$ :  $\Delta > 1$ , we have  $\lambda_E < \lambda = \lambda_C$ .

Model (1) can also accommodate this situation. We let  $Z_1 = 0$  for the control group,  $Z_2 = 1$  for the experimental group. Then  $\lambda_C = \lambda$  and  $\lambda_E = \lambda e^B$ , so  $B = -\ln\Delta$  in this case. Also, we have the null hypothesis  $H_0$ :  $B = 0$  versus the alternative  $H_1$ :  $B < 0$ . We again set  $T_k = \infty$  and  $p_1 = p_2 = 1/2$  for the same reasons previously described. Since  $Z_1 = 0$  and  $Z_2 = 1$ , we have from (5) that  $D_0 = D_1 = 1/2$ . We have  $B_0 = 0$  and  $B_1 = -\ln\Delta$ , so from (8)

$$N \geq 2 (\ln\Delta)^{-2} (z_{1-\alpha} + z_{1-\beta})^2. \quad (14)$$

Therefore, the sample size required in each of two equal sized groups for case 2 is

$$d = (\ln\Delta)^{-2} (z_{1-\alpha} + z_{1-\beta})^2. \quad (15)$$

Comparison of (13) and (15) reveals that for the same values of  $\Delta$ ,  $\alpha$ , and  $1-\beta$ , the case 1 situation as given in George and Desu (1974) requires twice the number of subjects as does the case 2 situation. It appears that the case 1 assumptions are more appropriate for observational studies, while the case 2 assumptions are more appropriate to randomized trials or experiments having a control group. For case 1, Table 1 from George and Desu (1974) or Table 2 can be used directly. For case 2, the table entry should be divided by two.

Often, the two-sided alternative is more appropriate and censoring will be required because of the need to have a study of fixed length. Evaluation of D of (5) with  $T_k < \infty$  and use of (9), (10), and (11) would be used for this circumstance. Rather than continuing with two groups as above, we illustrate this situation with an example having more than two groups and specifying  $\alpha$  and  $\beta$  at 0.05.

In this example, the time,  $t$ , is weeks until death after tumor implant in mice. Taking  $t$  to be exponentially distributed, suppose the experimenter has three groups, all to be given an implant. The groups will receive, respectively, 0, 10, and 20 unit doses of a compound which is believed to slow the spread of the tumor. The experimenter expects 15 week survival in the zero dose group to be about 20%, based on prior experience. He would like to be able to detect an increase in 15 week survival to 35% in the 10 unit dose group and to 50% in the 20 unit dose group. With units of dose as the covariate,  $Z_j$ , the hazard is  $\lambda e^{BZ_j}$ . The survival probability at time  $t$  is then  $\exp(-\lambda e^{BZ_j} t)$ . If we let  $\lambda = 0.1$ ,  $B = -0.04$  and  $t = 15$ , we find survival probabilities of 22%, 37% and 51%, respectively for  $Z_j = 0, 10, \text{ and } 20$ . So, in terms of model (1),

the investigator's hypotheses can be stated as  $H_0: B = 0$  versus  $H_1: B \neq 0$  to be tested against the specific alternative  $B = -0.04$ . Censoring time  $T_k$  will be the same for all individuals and equal group sizes are desired so that  $p_j = 1/3$ ,  $j = 1, 2, 3$  where  $Z_1 = 0$ ,  $Z_2 = 10$ ,  $Z_3 = 20$ .

From (5), then we have that

$$D_0 = [1 - \exp(-0.1 T_k)] [100 + 400] / 3 \quad (16)$$

and

$$D_1 = [1 - \exp(-0.067 T_k)] (100) / 3 + [1 - \exp(-0.045 T_k)] (400) / 3. \quad (17)$$

For  $T_k = 15$  weeks,  $D_0 = 129.48$ ,  $D_1 = 86.58$ ,  $B_0 = 0$ , and  $B_1 = -0.04$ ,

we have from (9) and (10)

$$x_L = -1.6027 + N^{1/2} (.3722)$$

and

$$x_U = 1.6027 + N^{1/2} (.3722).$$

Since for  $N$  of any reasonable size,  $\Pr(z > x_U) = 0$ , we ignore this term in (11). So we require

$$\Pr(z < x_L) \geq 0.95,$$

or

$$x_L > z_{0.95}.$$

Solving this last inequality for  $N$ , we find  $N = 77$ . Actually, we must have  $N = 78$  so that each of the three groups would contain 26 mice.



Alternatively if the investigator must limit the size of the study to 15 or 20 animals per group, the question becomes what study length (censoring time) will be needed to obtain the required power of 0.95? The first step is to see whether any study length will be adequate. This is accomplished by examining power for a specified N with  $T_k = \infty$ . For  $T_k = \infty$ , we have from (16) and (17) that  $D_0 = D_1 = 166.67$ . So, from (9) and (10), for  $N = 45$ , we have  $x_L = 1.50$  and  $x_U = 5.42$ . The power is then  $\Pr(z < 1.50) + \Pr(z > 5.42) = 0.93$ . Therefore, for 15 animals per group there is no study length long enough to provide the required power. For 20 animals per group, we proceed in exactly the same fashion as before and find that for  $T_k = \infty$ ,  $x_L = 2.04$  and the power is thus 0.98.

The task now is to find the smallest  $T_k$  for which the power is at least 0.95 for 20 animals per group. This requires iteration using (9), (10), and (11). We know that  $T_k$  must be greater than 15 since this study length required 26 animals per group. For  $T_k = 25$  weeks, (16) and (17) give  $D_0 = 152.99$  and  $D_1 = 117.14$ . Using these values along with  $N = 60$  in (9) and (10), we find  $x_L = 1.638$  and  $x_U = 5.068$ . The power,  $\Pr(z < x_L) + \Pr(z > x_U) = 0.949$ . For  $T_k = 26$  weeks, we find from (16) and (17) that  $D_0 = 154.29$  and  $D_1 = 119.45$ . From (9) and (10), we find  $x_L = 1.662$  and  $x_U = 5.111$ . Here, the power is  $\Pr(z < x_L) + \Pr(z > x_U) = 0.951$ . Therefore, for 20 animals per group, a 26 week study period would be required.

## 5. Discussion

The key assumption underlying this method of sample size determination is that the maximum likelihood estimators of the parameters of the survival time distribution are asymptotically normal with the variance-covariance matrix given by the inverse of the Fisher information matrix. Using four distribution functions and expressing the Fisher information matrix as a multiple of cohort size,  $N$ , the required sample size for a given test size and power requirement can be determined for the hypotheses of interest in practice. The four distribution functions, which can be chosen by the investigator to reflect properties of the cohort being studied, are as follows: 1) the distribution of observation times, conditional on whether the observation was an event of interest or censored, censoring time and the covariate; 2) the distribution of the variable indicating whether the observation was an event or censored, conditional on censoring time and the covariate; 3) the distribution of censoring time conditional on the covariate; and 4) the distribution of the covariate.

The first two distributions are specified by the model chosen for the hazard rate with covariables. Here, we used  $\lambda \exp(Bz_1)$ . Any suitable hazard function could be substituted for it in the method used here. The latter two distributions, for censoring time and covariate, can be made as simple as desired by use of discrete distributions, each with only a small number of possible values. There are no restrictions on the distributions which may be assumed, however.

Provision for the distribution of censoring times to be conditional on the covariate is made because censoring time may be quite different for different covariate values in some cases. For instance, suppose that the covariate indicates whether the subject in a clinical trial is in the experimental or control group. If the experimental treatment is much more rigorous than the control treatment, one might reasonably expect considerably earlier censoring time in the experimental group. However, it is not required that censoring time depend on covariate values, as was demonstrated in the examples in Section 4.

The method does not depend on choosing the particular form used here for the four distributions, it can be used with whatever forms of the distributions the investigator feels are appropriate. We emphasize that the form for the underlying hazard for the survival time distribution is not restricted. The constant hazard was used in this case, but the method can be used with any underlying hazard. However, it must be emphasized that the distributions chosen should reflect the conditions in the cohort to be studied insofar as possible. If the actual distributions in the cohort differ greatly from those assumed, then it is possible that the sample size chosen will not produce the desired power for the hypothesis test of interest.

The results of George and Desu (1974) are well approximated by a special case of model (1). In addition, the implicit problem formulation of George and Desu is clarified and two distinct cases are described. The first seems more appropriate for observational studies and the second for designed clinical trial investigations. In particular this suggests that for a two group randomized clinical trial with a placebo treatment, sample sizes taken from Table 1 of George and Desu (1974) are twice the minimum size required.

We note that the method developed here can be used to select required follow-up time for a given sample size. This can be done by fixing  $N$  and allowing the distribution of censoring times to vary upward until the power requirement is satisfied. Of course, if  $N$  is too small, then no censoring time, even infinity, will be adequate.

#### Acknowledgments

We would like to thank Regina C. Elandt-Johnson, P. K. Sen, Kerry L. Lee, Robert D. Bruce and Scott E. Michaels for helpful suggestions during the preparation of this paper. The early part of the research was partially supported by a National Institute for Occupational Safety and Health training grant and by a National Institute of Environmental Health Sciences Young Environmental Scientist Health Research Grant while J. D. Taulbee was in the Department of Biostatistics at the University of North Carolina at Chapel Hill. The work was completed with the support of the Procter and Gamble Company through the encouragement of Brandon H. Wiers for J. D. Taulbee and under an Intergovernmental Personnel Act appointment with the U.S. Environmental Protection Agency in North Carolina for M. J. Symons.

References

- Feigl, P., and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information, Biometrics 21, 826-838.
- Feinstein, A. R. (1975). Clinical Biostatistics XXXIV. The other side of statistical significance: alpha, beta, delta and the calculation of sample size, Clinical Pharmacology and Therapeutics 18, 491-505.
- George, S. L., and Desu, M. M. (1974). Planning the size and duration of a clinical trial studying the time to some critical event. Journal of Chronic Diseases 27, 15-24.
- Glasser, M. (1967). Exponential survival with covariance, Journal of the American Statistical Association 62, 561-568.
- Halperin, M., Rogot, E., Gurian, J., and Ederer, F. (1968). Sample sizes for medical trials with special reference to long-term therapy, Journal of Chronic Diseases 21, 13-24.
- Pasternack, B. S. (1972). Sample sizes for clinical trials designed for patient accrual by cohorts, Journal of Chronic Diseases 25, 673-681.
- Pasternack, B. S., and Gilbert, H. S. (1971). Planning the duration of long-term survival time studies designed for accrual by cohorts, Journal of Chronic Diseases 24, 681-700.
- Schlesselman, J. J. (1974a). Sample size requirements in cohort and case-control studies of disease, American Journal of Epidemiology 99, 381-384.
- Schlesselman, J. J. (1974b). Tables of the Sample Size Requirement for Cohort and Case-Control Studies of Disease, (Unpublished paper).

Schork, M. A., and Remington, R. D. (1967). The determination of sample size in treatment-control comparisons for chronic disease studies in which drop-out or non-adherence is a problem, Journal of Chronic Diseases 20, 233-239.

Taulbee, J. D. (1979). A general model for the hazard rate with covariables. Biometrics 35, 439-450.

Taulbee, J. D. and Symons, M. J. (1979). Approximate sample size requirements for cohort studies with covariables. Presented at the Joint Meeting of the American Statistical Association, Biometric Society, and Institute for Mathematical Statistics, Washington, D.C., August.

Appendix

Evaluation of D where  $(ND)^{-1} = \text{Var}(\hat{B})$

To evaluate  $D = E[\lambda \exp(Bz_i) t_i z_i^2]$ , we need the distribution of the random variable  $t_i$ , time observed for individual  $i$ . In general,  $t_i$  depends upon the values of the parameters  $\lambda$  and  $B$ ; whether the individual was censored or not ( $I_i = 0$  for censored observations,  $I_i = 1$  otherwise); the censoring time  $T_k$ ; and the covariate,  $Z_j$ . (Different subscripts are used for censoring time and covariate to make explicit the possibility that more than one individual may have the same value.)

It is possible that dependencies of one random variable upon the other may occur. For instance, observation time will certainly depend upon censoring time. In addition, censoring time may depend upon the value of the covariate in some cases. Therefore, we will specify the joint distribution of  $t_i$  with the other random variables by means of conditional distributions, as follows: the conditional cumulative distribution function (CDF) of  $t_i$  is  $F_0(t_i | Z_j, T_k, I_i)$ ; the conditional CDF of  $I_i$  is  $F_1(I_i | Z_j, T_k)$ ; the conditional CDF of  $T_k$  is  $F_2(T_k | Z_j)$ ; and the CDF of  $Z_j$  is  $F_3(Z_j)$ . Of course, the CDF's will depend upon the values specified for the parameters of the model. Using these distributions, we have that

$$D = \iiint \int \lambda e^{BZ_j} t_i z_j^2 dF_0(t_i | Z_j, T_k, I_i) dF_1(I_i | Z_j, T_k) dF_2(T_k | Z_j) dF_3(Z_j) \quad (A1)$$

where the integrals of (A1) are Riemann-Stieltjes integrals over the appropriate region. We note that if  $F(t_i, I_i, T_k, Z_j)$  is the joint CDF of  $t_i$ ,  $I_i$ ,  $T_k$ , and  $Z_j$ , then (A1) could also be expressed as

$$D = \iiint \lambda e^{BZ_j} t_i Z_j^2 dF(t_i, I_i, T_k, Z_j).$$

While this general formulation (A1) may seem complicated at first, it is not difficult to evaluate in cases of interest. We model the hazard for an individual with covariate  $Z_j$  as  $\lambda \exp(BZ_j)$ . For the distribution of  $I_i$  given covariate  $Z_j$  and censoring time  $T_k$ , we find

$$P_{jkl} = \Pr(I_i = 1 | Z_j, T_k) = \int_0^{T_k} \lambda e^{BZ_j} \exp[-\lambda e^{BZ_j} t] dt$$

or

$$P_{jkl} = 1 - \exp[-\lambda e^{BZ_j} T_k]$$

is the probability that an individual with covariate  $Z_j$  and censoring time  $T_k$  is observed to fail, that is, has  $I_i = 1$ . Therefore the probability that an individual with covariate  $Z_j$  and censoring time  $T_k$  is not observed to fail, that is, has  $I_i = 0$  is given by

$$P_{jk0} = \exp[-\lambda e^{BZ_j} T_k].$$

Now, let us examine the observation time  $t_i$  for individual  $i$ . For an individual with covariate  $Z_j$  and censoring time  $T_k$ , if  $I_i = 0$ , that is, the individual has not failed, we find

$$\Pr(t_i = t | Z_j, T_k, I_i = 0) = \begin{cases} 1 & \text{for } t = T_k \\ 0 & \text{for } t \neq T_k \end{cases}.$$



If an individual with covariate  $Z_j$  and censoring time  $T_k$  has  $I_i = 1$ , that is, the individual has failed, then the probability density function of  $t_i$  is given by

$$f_0(t_i | Z_j, T_k, I_i = 1) = \begin{cases} \frac{\lambda e^{-BZ_j t_i} \exp[-\lambda e^{BZ_j t_i}]}{1 - \exp[-\lambda e^{BZ_j T_k}]}, & 0 \leq t_i \leq T_k \\ 0 & \text{elsewhere} \end{cases}$$

In the evaluation of D of (4), it is necessary to evaluate the following quantity:

$$\iint t_i dF_0(t_i | Z_j, T_k, I_i) dF_1(I_i | Z_j, T_k)$$

This quantity is given by the following expression:

$$P_{jk0} T_k + P_{jk1} \int_0^{T_k} t f_0(t | Z_j, T_k, I_i = 1) dt$$

Evaluating the integral in the above expression by parts, we find that

$$\iint t dF_0(t | Z_j, T_k, I_i) dF_1(I_i | Z_j, T_k) = P_{jk0} T_k + P_{jk1} \left\{ \frac{\lambda^{-1} e^{-BZ_j} [1 - (1 + \lambda e^{BZ_j T_k}) \exp(-\lambda e^{BZ_j T_k})]}{1 - \exp(-\lambda e^{BZ_j T_k})} \right\}$$

Upon substitution of the values for  $p_{jk0}$  and  $p_{jk1}$ , the above expression becomes

$$\lambda^{-1} e^{-BZ_j} [1 - \exp(-\lambda e^{BZ_j T_k})]. \quad (A2)$$

We have from (A1) and (A2) that

$$D = \iint Z_j^2 [1 - \exp(-\lambda e^{BZ_j} T_k)] dF_2(T_k | Z_j) dF_3(Z_j). \quad (A3)$$

In all cases, the  $Z_j$  will be assumed to be discretely distributed. The evaluation of  $D$  will be done assuming that the  $T_k$  are discretely distributed (case I) and again assuming that the  $T_k$  are exponentially distributed (case II).

The covariate can assume any one of  $J$  values denoted by  $Z_j$ ,  $j = 1, 2, \dots, J$ . We take  $\Pr(Z_j) = p_j$  and note that the condition

$$\sum_{j=1}^J p_j = 1 \text{ must hold.}$$

Case I:  $T_k$  discretely distributed.

The censoring time can assume any one of  $K$  values denoted by  $T_k$ ,  $k = 1, 2, \dots, K$ . We take  $\Pr(T_k | Z_j) = p_{jk}$  and note that we must have

$$\sum_{k=1}^K p_{jk} = 1 \text{ for each } j,$$

$j = 1, 2, \dots, J$ . For case I, then, we have from (A3) that

$$D = \sum_{j=1}^J p_j Z_j^2 \sum_{k=1}^K p_{jk} [1 - \exp(-\lambda e^{BZ_j} T_k)]. \quad (A4)$$

Case II:  $T_k$  distributed exponentially.

To allow for dependence of censoring time upon the covariate  $Z_j$ , we let

$$f_2(T_k | Z_j) = \mu e^{CZ_j} \exp(-\mu e^{CZ_j} T_k), \text{ so}$$

$$\int [1 - \exp(-\lambda e^{BZ_j} T_k)] f_2(T_k | Z_j) dT_k = \frac{[\lambda \exp(BZ_j)]}{[\lambda \exp(BZ_j) + \mu \exp(CZ_j)]}.$$

So, from (A3) we have that for case II

$$D = \sum_{j=1}^J p_j z_j^2 [\lambda \exp(BZ_j)] / [\lambda \exp(BZ_j) + \mu \exp(CZ_j)]. \quad (A5)$$

Formulas (A4) and (A5) for D are given in (5) and (6), respectively, in Section 2.