Sample Size Determination in Clinical Trials with Time-Dependent Rates of Losses and Noncompliance

Edward Lakatos, Ph.D.

Biometrics Research Branch, National Heart, Lung, and Blood Institute, Bethesda, Maryland

ABSTRACT: Sample size determination is an important part of planning for clinical trials. During the course of a typical clinical trial, people are lost because of competing risks, noncompliance, and the like. Event rates available to the trial designers usually do not take these losses into consideration so that adjustment of these rates is necessary for sample size calculation. This article presents a method of adjusting such rates in the presence of time-dependent rates of losses, noncompliance, and the like. Lag in the effectiveness of medication is also considered.

KEY WORDS: *Markov model, transition matrix, time-dependent losses, lag times*

INTRODUCTION

Sample size determination is an important part of planning for clinical trials. Lachin [1] gives a general discussion of sample size calculations for various situations. Most of these calculations assume that the parameters for calculating the sample sizes, such as event rates or, in the case of survival curves, hazard rates, have already been estimated. Usually, the estimated rates available to the investigators should be modified before being used in the sample size formulas, since the assumptions used in estimating the parameters frequently do not represent what may be expected in the course of the planned trial. For example, if mortality from a given disease is the outcome of interest, previous estimates of mortality may not take into account the fact that people in the control group may start taking medication on their own, i.e., cross over to active treatment (drop-in), thus changing their event rate. Similarly, those on active treatment may discontinue their medication (noncompliers) and thus alter their expected event rate. Loss to follow-up and lag times in the effectiveness of medication further alter the assumed rates.

Address reprints requests to: Dr. Edward Lakatos, Biometrics Research Branch, National Heart, Lung, and Blood Institute, Bethesda, MD 20892.

Received March 24, 1986; revised April 24, 1986.

0197-2456/86/$0.00

Schork and Remington [2], and Halperin, Rogot, Gurian, and Ederer (HRGE) [3] offer methods to adjust the event rates under constant noncompliance rates and specified lag times. Wu, Fisher and DeMets (WFD) [4] generalize the HRGE model to allow different rates in different time intervals and provide a computer program for its implementation. Since the WFD method requires numerical integration, the programming is not simple. Furthermore, although the program is available upon request from the authors, it is not easily modified by a user whose assumptions do not coincide with those of WFD.

The purpose of this article is to introduce a Markov chain model that provides an alternate method of determining sample size under time-dependent rates of losses and noncompliance. The method can be adjusted by the user to accommodate a wide range of assumptions, either parametric, as in WFD, or nonparametric. Under the assumptions of WFD, the Markov model approaches the WFD model as the number of subintervals increases, so the method provides another way to obtain the WFD results (see Appendix 1). Furthermore, the event rates, loss rates, and so on for the intermediate years are readily available and have been found to be quite useful to the investigators. The model also takes into consideration the effects of differential exposure due to staggered entry.

THE GENERAL MODEL

Let P_E and P_C denote the probabilities that patients in the treatment and control groups will experience events by the end of the trial. Assume a binomial model, a null hypothesis of no treatment effect (i.e., $P_E = P_C$) and alternative of $P_E = (1 - r)P_C$ (where r is a prespecified reduction). In order to achieve a two-sided significance level α and a power $1 - \beta$, when the control and experimental groups are the same size, a total sample of approximately

$$2N = [2/(P_C - P_E)^2]\,[z_\alpha\sqrt{2P(1 - P)} + z_\beta\sqrt{P_E(1 - P_E) + P_C(1 - P_C)}]^2 \quad (1)$$

is required, where z_α is the standard normal variate whose absolute value has probability α of being exceeded, and $P = (P_E + P_C)/2$. The problem is to determine appropriate values of P_E and P_C. Assumed values of P_E and P_C are typically based on anticipated control rates and a minimum reduction due to treatment that can be considered meaningful, but do not make adjustments for losses or the like.

The WFD and HRGE models assume that all patients are followed from some fixed time to some common closing date of the trial. Analysis is then based on a Z test for the comparison of the two adjusted probabilities. The use of such a model for clinical trials in which accrual takes place over an extended period of time results in conservative sample size estimates. An extended accrual period is incorporated into the Markov model presented here, eliminating this source of bias in the sample size estimation.

The Markov chain model for adjusting P_E and P_C in the presence of time-dependent rates of losses and noncompliance is as follows. Adjustment is performed separately in each treatment group. Without loss of generality

consider only the experimental group. At a given time t in the course of a typical clinical trial, a patient can be in one of the following states:

State L—lost to follow-up so that no further information, including the occurrence of events, is available;

State E—had an event, that is, the endpoint of interest already occurred; or

State A_i—continues to be followed, either on active treatment or as a noncomplier, with event rate p_i. Here p_i will generally depend on whether the patient is on treatment currently, as well as the length of time on that treatment.

The distribution of states at time t is denoted by the column vector

$$\mathcal{D}_t = (P_L(t), P_E(t), P_{A_0}(t), P_{A_1}(t), \ldots, P_{A_n}(t)'),$$

where $P_E(t)$ [respectively $P_L(t)$] denotes the probability a patient will have an event (respectively will be lost to follow-up) by time t and $p_{A_i}(t)$, $i = 0, \ldots, n$, denotes the probability that a patient will be in state A_i at time t. We then define transition matrices between times t_1 and t_2 as follows:

$T_{t_1,t_2}(i,j)$ is the probability of going from state j to state i during the period $[t_1,t_2]$.

The final distribution is given by

$$\mathcal{D}_{t_n} = (\Pi\, T_{t_{i-1},t_i})\mathcal{D}_{t_0},$$

and the adjusted event rate, i.e., the probability of having an event by the time t_n, is given by the second element of $\mathcal{D}_{t_n}$. Since transition probabilities for treatment and control will be different, P_C and P_E are calculated separately.

Although this model indicates a discrete process, a continuous process can be approximated by letting the time interval $[t_1,t_2]$ approach 0.

ASSIGNMENT OF PROBABILITIES TO TRANSITION MATRICES

Where there is an underlying model, such as binomial or exponential losses, the specification of a single parameter can define the losses over the length of the trial. Since the Markov model has no inherent restrictions, specification of the transition probabilities may at first seem perplexing. Certainly one could choose a parametric model and assign probabilities based on a single parameter that would remain fixed over the course of the trial. However, WFD presents a strong case for allowing the parameters to vary over different years. We have taken the WFD approach here. In the absence of evidence for differing rates over different years, one simply assigns the same value throughout.

Assignment of probabilities within a year, say, is also completely unrestricted. Again, we have followed WFD (and HRGE) in assuming that within a given year the probability of an event is the same across intervals of equal length, i.e., the negative exponential applies. This leads to a step function that jumps at the end of each year. Although a step function seems unlikely,

it was chosen because (1) in most instances, results will not differ much from a continuous model, (2) it leads to the same estimates as WFD, (3) it is easy to implement, and (4) there is no strong evidence for a different model. If this restriction within each year seems excessive, one could specify rates at half- or quarter-year intervals. For an arbitrary function, one could fit the function to as many time points during the period of the trial as desired, thus approximating this function.

AN EXAMPLE

Table 1 gives one set of parameters considered in calculating sample sizes for the Systolic Hypertension in the Elderly (SHEP) trial [6]. In this trial patients are randomized to hypertension medication or control and followed for 5 years. The primary outcome is fatal and nonfatal stroke. Those who die of nonstroke causes can no longer be followed for the outcome. Loss to such competing risks is estimated to be 3% in the first year and is expected to increase uniformly to 4% over the next 6 years as this elderly population ages. In contrast, the event (stroke) rate is assumed to depend only on the hypertension treatment currently being taken. Those randomized to control as well as noncompliers to the experimental regimen are assumed to have the rate 1.6 per 100 per year. Treatment is assumed to reduce the rate of stroke by 40% so that those complying with their assigned experimental therapy as well as the drop-ins have a 0.96 per 100 event rate each year. The noncompliance rates in Table 1 arose from experience in several clinical trials. First-year rates are usually twice the rates in successive years. The designers of SHEP felt that because of aggressive national programs for identification and treatment of hypertensives, treatment of controls via their private physicians would increase over the succeeding years.

In this case there are four states for the distribution of patients (losses, events, active with rate P_E, active with rate P_C). For the experimental group, the initial distribution in these four states is

$$\mathfrak{D}_{t_0} = (0,0,1,0)'.$$

The transition matrix at year 3 is

	L	E	A_E	A_C
L	0	1	0.036	0.036
E	1	0	0.0096	0.016
A_E	0	0	$1-\Sigma_1$	0.050
A_C	0	0	0.035	$1-\Sigma_2$

Table 1 Loss, Noncompliance, and Drop-in Rates for a Clinical Trial

State	Year 1	Year 2	Year 3	Year 4	Year 5
Lost	0.03	0.032	0.034	0.036	0.038
Event (P_E)	0.0096	0.0096	0.0096	0.0096	0.0096
Noncompliance	0.07	0.035	0.035	0.035	0.035
Drop-in	0.09	0.045	0.050	0.055	0.060
Event (P_c)	0.016	0.016	0.016	0.016	0.016

where $\Sigma_1 = 0.036 + 0.0096 + 0.035$ and $\Sigma_2 = 0.036 + 0.016 + 0.050$. Column A_E indicates that those active and on the experimental treatment become losses with probability 0.036, become noncompliers (noncompliance) with probability 0.035, and otherwise remain on assigned therapy.

In this example we assume (cf. WFD) that the probability of an event is constant across intervals of equal length within a given year and thus approximate the continuous negative exponential distribution by replacing each $T_{i,j}$ by a product of n transition matrices in which each entry x is replaced by $x' = 1 - (1 - x)^{1/n}$. A program implementing this setup is given in Appendix 2.

Using the parameters specified above and the program of Appendix 2 yields 5-year distributions of

$$\mathcal{D}_{t_{5\text{year}}} = (0.1528, 0.0677, 0.5875, 0.1920)$$

and

$$\mathcal{D}_{t_{5\text{year}}} = (0.1548, 0.0463, 0.6683, 0.1306)$$

for the control and experimental groups, respectively. The rates are thus $P_C = 0.0677$ and $P_E = 0.0463$, leading to a sample size of $N = 4928$ from equation (1). If sample size were calculated ignoring losses, noncompliance and drop-in, the control and treatment rates would be $1 - (1 - 0.016)^5 = 0.0775$ and $1 - (1 - 0.0096)^5 = 0.0471$, respectively. The sample size would be calculated at 2652, leading to a grossly underpowered study.

THE PATIENT ACCRUAL PERIOD

The general model just presented assumes all patients enter simultaneously at the beginning of the trial. Usually, patients enter over a period of time and if they do not have an event and are still being followed for the outcome of interest at the common closing date of the trial, the observations on these patients are administratively censored. This extended accrual period could be incorporated into the model by setting the complying active state to zero at time zero and adding according to the accrual pattern over the course of the trial. One disadvantage of such an approach is that at any fixed point in time, different people have different exposure times. Thus, the setup would exclude modeling the failure rate changing over time. Because of this, the approach taken here is to enter everyone at the beginning (i.e., set the active complying state to one at time zero) and administratively censor observations at a rate in concordance with the accrual rate. In particular, let p_i be the probability of being recruited during the ith subinterval, $i = 1, \ldots, N$. At the time of minimum follow-up, all active participants have been exposed for the same length of time. The probability of a participant being administratively censored during this interval is $p_N/\Sigma^N p_i$. In the next interval it is $p_{N-1}/\Sigma^{N-1} p_i$, etc.

The SHEP trial was designed for a 2-year recruitment period and each patient who was not lost to follow-up or did not have an event would be observed for a minimum of 4 years. Thus, the average observation time would be 5 years. Under uniform recruitment, the event rates are $P_E = 0.0457$ and