THE DETERMINATION OF SAMPLE SIZE IN TREATMENT-CONTROL COMPARISONS FOR CHRONIC DISEASE STUDIES IN WHICH DROP-OUT OR NON-ADHERENCE IS A PROBLEM

M. Anthony Schork and Richard D. Remington

Department of Biostatistics, School of Public Health, The University of Michigan, Ann Arbor, Michigan

INTRODUCTION

A common feature of many investigations in the chronic disease in the comparison of two groups—for convenience designated here as treatment and control—in which characteristics of one or both of the assigned regimes favor non-adherence or drop-out. For example. the assigned "treatment" may produce undesirable side reactions or may require extensive commitment of time or effort on the part of subjects and may tend to promote voluntary discontinuance of the regime, thus effectively causing such subjects to adopt the "control" regime. On the other hand, in certain situations subjects assigned to the "control" group may hear of assumed benefits of the "treatment" under study and may effectively shift to the other group. This latter case might arise for example when the "treatment" is a dietary change, a change in physical activity level or a change in smoking habits.

An additional complicating feature of studies in the chronic diseases is that a long period of observation of the two groups is frequently necessary to produce any discernible difference in subjects status or to produce sufficient occurrences of an endpoint to allow meaningful comparisons. Furthermore, endpoint determinations are frequently of the "go–no go" two-point type such as the occurrence or non-occurrence of coronary occlusion within a fixed period of observation, say 5 yr.

In this paper we shall consider studies involving a single treatment–control comparison, a relatively long period of observation, and an outcome variable consisting of only two states. Some examples of this type of experimental situation are: 1. Study of the effect of a supervized physical activity program on reduction of coronary heart disease in middle-aged American men. Here the treatment comparison is physical activity vs. relative inactivity, the outcome variable may be the occurrence of a new coronary event within a 5-yr period from the beginning of the study. 2. Investigation of a specific dietary regime for its effect on reducing the frequency of new coronary events. 3. Investigation of the effect of controlling

blood pressure level in patients with cerebral vascular disease. Here the endpoint might be the recurrence or non-recurrence of a cerebral event within a fixed period of observation.

In each of these examples it can be seen that after random assignment to treatment or control groups subjects might well during the course of the study voluntarily shift out of their assigned group to the other group, e.g. subjects assigned to a physical activity program might stop participating in the program and thus effectively assume the characteristics of control subjects. Conversely, it is possible that a subject in the control group of such a study might voluntarily begin exercising at or near the level of the members of the treatment group. Subjects prescribed a particular medication for blood pressure control might discontinue the medication, while subjects assigned to the control group of such a study might seek additional drug treatment even though double blind placebo control was used. This type of voluntary shift will be a more frequently occurring problem in chronic disease epidemiology in which available treatments at present often involve a reasonably long-term alteration in way of life. Epidemiologists and biostatisticians are in general agreement that once a random assignment to treatment or control group has been made it is not good practice to remove subjects from the assigned group if this can possibly be avoided. Such removal at the time of analysis can result in effective derandomization and bias the treatment–control comparison. For most purposes it seems better to take one's losses occasioned by the aforementioned shifting of subjects between groups and retain the subjects for analytical purposes in the group to which they were originally assigned. It is then reasonable to assume that a subject who discontinues a treatment program before completing the assigned study length will spend a portion of the study with the risk assumed for the experimental group and the remainder of the study at the risk assumed for the control group. This results in a dilution of the presumptive difference in risk between treatment and control groups with corresponding increase in sample size required to detect a treatment difference.

We shall show in this paper a quantitative method for determining sample size in the face of such shifting and demonstrate the magnitude of the increase in requisite sample size as a function of the assumed pattern of frequencies with which subjects shift from one group to the other during the course of the study.
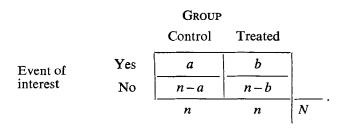
METHODS AND MODEL

Abstracting the common features from the foregoing illustrations the following model evolves:

1. Two groups are to be compared. For conciseness of notation, in this paper, they shall be referred to as control (C) and treated (T), although certainly any situation involving two groups will apply.

2. The outcome or endpoint variable consists of two event states, e.g. death–survival.

3. For each member of the control group there is the same unknown probability that the event of interest will occur during one unit of time (frequently 1 yr).

4. A proportionate change in the probability of such events is specified for the second group (treated) relative to the first group (control) is hypothesized.

5. A length of time, denoted by L time units, is specified for the study.

1. *Notation*

1. Let P_C be the probability an individual in the control group has an event in a specified time period.

2. Let d be the hypothesized proportionate change from the control to the treated group per unit of time. d is positive if the change is an increase and negative if it is a decrease.

3. Let P_T be the probability an individual in the treated group has an event in the specified time period. Then $P_T = P_C + d \cdot P_C = P_C(1 + d)$ and e.g. if $P_C = 0.10$, $d = -0.50$, then $P_T = 0.10 - 0.50 \times 0.10 = 0.05$.

The basic format for analysis is the familiar two by two table.

		Control	Treated	
		GROUP		
Event of interest	Yes	a	b	
	No	$n - a$	$n - b$	
		n	n	N

The numbers a, b, $n - a$, $n - b$, represent the number of observations falling into the appropriate category at the conclusion of the experiment.

It is important to note that this paper only considers the situation where the two groups are of the same size. If N is the total sample size then n is the size of sample in each group. where $N = 2n$.

2. *Hypothesis*

The hypothesis to be considered is that the probability an individual in the control group has an event is the same as the probability an individual in the treated group has an event, i.e. $P_C = P_T$, or $d = 0$. The alternative hypothesis is determined by specifying d. The sign of d determines whether the change in P_C induced by the treatment is presumed to be an increase or a decrease and the magnitude of d determines the size of treatment effect the study is designed to detect.

3. *Patterns of subject-shifts*

It is necessary to assume some particular pattern of shift from one group to the other, i.e. the percentage of patients shifting from treatment to control group and from the control to the treatment group within a particular time unit. It is assumed that subjects making the shift during a particular time unit do so on the average at mid-unit. As an example of the type of specification of shift pattern that is needed to determine sample size under this model, suppose that in the 5-yr exercise study described above, 20 per cent of the patients discontinue exercise during the first year while an additional 10 per cent do so in each of the remaining 4 yr of the study. We assume that the 20 per cent of subjects dropping from the exercise group during year one do so on the average after 6 months of exercise and make a similar assumption for other subjects who stop exercizing later in the study.

4. *Effective rates*

Recall the P_C and P_T are expressed as rates per unit of time, frequently 1 yr. For a study L units in length the effective total (study-long) rates for control $(E.P_C)$ and treatment $(E.P_T)$ groups are given in formulae (1) and (2) respectively, assuming no subject-shifts.

(1) $E.P_C = 1 - (1 - P_C)^L$

(2) $E.P_T = 1 - (1 - P_T)^L$.

These results are based on the following reasoning, $1 - P_C$ is the probability a subject assigned to the control group will not have an event within a given time unit, thus the probability he will not exhibit an event through L time units is $(1 - P_C)^L$. The probability therefore, that he will show an event at some point in the study is one minus the probability he will not, thus expression (1). Expression (2) follows similarly.

5. *Dilution of effective rates by subject-shifts*

As noted earlier drop-outs dilute differences between the groups. The effect of such drop-outs is to bring $E.P_C$ and $E.P_T$ closer, causing an increase in sample size if a difference is to be deemed statistically significant.

Since L represents the length of the study, let t be the length of the time the person remains in the group assigned him, and $L - t$ the length of time he spends in the other group.

To compute revised $E.P_C$ and $E.P_T$ (denoted by $R.E.P_C$ and $R.E.P_T$ below) taking into account shift patterns, one must consider the proportion of persons who shift at any time t. Call this proportion R_t for the control group and R'_t for the treated group. For example, if 20 per cent of the persons who start the exercise regime stop within the first year then $R'_{0.5} = 20$ per cent (recalling the assumption that the shift occurs on the average at mid-year). Further if an additional 10 per cent shift in each succeeding year $R'_{i-\frac{1}{2}} = 10\%$ $(i = 2,3,4,5)$. For compactness of shift presentation, we shall assume all shifts occur at mid-unit and define R_L, (and R'_L) as the proportion of subjects completing the entire study in the group to which they were originally assigned. In the example, $R'_L = R'_5 = 40\%$.

It follows that the revised effective rates (R.E.P.) are found by application of formulae (3) and (4).

(3) $R.E.P_C = \sum_{i=1}^{L} R_{i-\frac{1}{2}} [1 - (1 - P_C)^{i-\frac{1}{2}} (1 - P_T)^{L-i+\frac{1}{2}}] + R_L [1 - (1 - P_C)^L]$.

(4) $R.E.P_T = \sum_{i=1}^{L} R'_{i-\frac{1}{2}} [1 - (1 - P_T)^{i-\frac{1}{2}} (1 - P_C)^{L-i+\frac{1}{2}}] + R'_L [1 - (1 - P_T)^L]$.

6. *Statistical errors*

In a study which involves sampling, one can only find a *statistically* significant difference between $R.E.P_C$ and $R.E.P_T$. Such a result is liable to the foibles of any inductive argument, i.e. erroneous conclusions might be reached. Considering the hypothesis is that $R.E.P_C = R.E.P_T$ there are two classes of possible erroneous conclusions:

1. If $R.E.P_C = R.E.P_T$ is a true statement, the sample data may lead to the rejection of this fact, i.e. a true hypothesis is rejected; or

2. If $R.E.P_C = R.E.P_T$ is a false statement, the sample data may lead to the acceptance of this statement as if it were factual, i.e. a false hypothesis is accepted.

The first type of error is termed the α-error and determines the level of significance, while the second type of error is called the β-error. To estimate sample size we can present α and β. Obviously we want the probability of either an α or β error to be small. It seems apparent that as one reduces these probabilities, there is a corresponding increase in sample size, i.e. the less chance one allows for error, the more information one needs to avoid error.

7. Sample size determination

Once the procedure of computing $R.E.P_C$ and $R.E.P_T$, (3) and (4) is completed, and the probabilities of the α and β errors are stated, one is ready to estimate the sample size needed to detect a statistical difference between $R.E.P_C$ and $R.E.P_T$.

Recall N, the overall sample size is equal to $2n$, where n is the size of sample in each group. This estimating procedure uses a technique developed by PATNAIK. [1] The formulae needed are:

$$(5) \quad K = \frac{\sqrt{(2)}}{4} \left[\sqrt{(P_1 + P_2)(2 - P_1 - P_2)} \right] \log_e \frac{P_1(1 - P_2)}{P_2(1 - P_1)} ,$$

where P_1 is the larger of $R.E.P_T$ or $R.E.P_C$, and P_2 is the smaller; and

$$(6) \quad n = \frac{H^2}{K^2} .$$

H is a number found in Patnaik's tables. Table 1 summarizes the most commonly used values of H.

TABLE 1.   VALUES OF H (FOR A TWO-SIDED TEST)*

α	10%			5%			2%			1%		
$\beta\beta$	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
H	2.928	3.298	3.978	3.246	3.606	4.296	3.609	3.975	4.663	3.863	4.224	4.913

*If a one-sided test is required; let $\dot{a}$ be the type I error, and use that portion of the table for $\alpha^* = 2\dot{a}$, e.g. if $a = 5\%$ for a one-sided test, use $\alpha^* = 2\dot{a} = 10\%$ for H.

It should be noted that the α-error values in Table 1 are presented for a two-sided test. The case considered in this paper employs a one-sided alternative and requires use of the α-value in the table twice that given in the problem, e.g. if $\alpha = 1\%$ in a one-sided test use that portion of Table 1 corresponding to $\alpha = 2\%$.

This procedure yields approximate values for n. The approximation is best for $R.E.P_C$ and $R.E.P_T$ near to 50%, while for $R.E.P_C$ and $R.E.P_T$ near the extremes of 0% or 100% this approximation becomes less accurate. Patnaik discusses a second estimating scheme to be used in such circumstances.

8. Summary of formulae and information needed

The above sections have gone into some detail explaining the development of

the necessary machinery to arrive at sample size estimation. A summary of this information, hopefully, should show the procedure not to be difficult. The information is:

1. P_C and P_T the initial probabilities expressed for one unit of time.
2. L the length of the study, expressed as the number of units of time.
3. Subject-shift pattern including R_t and R'_t, the proportions shifting at time t.
4. The probabilities of the α and β errors.

ILLUSTRATION

For the heart disease–exercise example presented in the introductory section consider:

$P_C=2\%$ per annum

$d=25\%$ reduction per annum, so

$P_T=2\% -0.25 \times 2\% =1.5\%$ per annum; one unit of time is 1 yr.

$L=5$ yr.

Subject-shift pattern—suppose 1% of the original group per year shift from control to treated by initiating an exercise program on their own, while of the original exercise group 20% the first year and 10% each of the succeeding 4 years stop exercising. Further suppose all shifts occur on the average at mid-year, thus

$R_{i-\frac{1}{2}}=1\%$, $i=1, 2, 3, 4, 5$

$R_5=95\%$,

$R'_{0.5}=20\%$

$R'_i-\frac{1}{2}=10\%$, $i=2, 3, 4, 5$

$R'_5=40\%$

$\alpha=\beta=5\%$,

now,

$$
\begin{aligned}
R.E.P_C=\ &0.01 \left[1-(1-0.02)^{0.5} (1-0.015)^{4.5}\right] \\
&+0.01 \left[1-(1-0.02)^{1.5} (1-0.015)^{3.5}\right] \\
&+0.01 \left[1-(1-0.02)^{2.5} (1-0.015)^{2.5}\right] \\
&+0.01 \left[1-(1-0.02)^{3.5} (1-0.015)^{1.5}\right] \\
&+0.01 \left[1-(1-0.02)^{4.5} (1-0.015)^{0.5}\right] \\
&+0.95 \left[1-(1-0.02)^5\right]=0.0955
\end{aligned}
$$

and

$$
\begin{aligned}
R.E.P_T=\ &0.20 \left[1-(1-0.015)^{0.5} (1-0.02)^{4.5}\right] \\
&+0.10 \left[1-(1-0.015)^{1.5} (1-0.02)^{3.5}\right] \\
&+0.10 \left[1-(1-0.015)^{2.5} (1-0.02)^{2.5}\right] \\
&+0.10 \left[1-(1-0.015)^{3.5} (1-0.02)^{1.5}\right] \\
&+0.10 \left[1-(1-0.015)^{4.5} (1-0.02)^{0.5}\right] \\
&+0.40 \left[1-(1-0.015)^5\right]=0.0807,
\end{aligned}
$$

$H=3.298$

$$K = \frac{\sqrt{(2)}}{4}[\sqrt{(0.1762 \times 1.8238)}] \log_e \frac{0.0955 \times 0.9193}{0.0807 \times 0.9045}$$

$$= 0.0369.$$

$$n = \frac{(3.298)^2}{(0.0369)^2} 2 = 7969; \quad N = 15,938.$$

The sample size (almost 8000 per group) indicates that such an undertaking would be expensive and time consuming, but it is important to recognize these facts prior to initiating a study with samples which could not lead to a conclusion.

DISCUSSION

The foregoing series of manipulations may seem tedious every time an estimate of sample size is desired. It is presented first to help point out that there is no simple answer to the oft-asked statistical question: "How many observations are needed to . . . .?" and second to demonstrate a technique which incorporates many facets of sample size computation.

One result the example illustrates is that if one wishes to show a statistical difference between two groups, many observations may be needed. It is also noted that if d—the hypothetical difference between the groups—is not large many observations are needed to discern a difference. Subject-shift plays an important role in over-all sample size; as a rule, the more shifts, the larger the sample.

One should also consider subject "lost" to follow-up (as opposed to those shifting, "lost" subjects can not be followed, e.g. moved, died, records lost, etc.). For such subjects it is suggested that a number be estimated for the length of study and this be added to the n as computed in this paper, e.g. in the example suppose it is assumed 10 percent of the observations are lost to follow-up. Therefore, n should be increased from about 8000 to 8900.

It might be felt that the procedure presented here for estimating n is so basic that it does not relate to all the information wanted from a study. However, if the most gross difference between the groups is not statistically significant for the given sample sizes, then more subtle differences will prove a fruitless venture. As a rule: estimate requisite sample size for the broadest, most basic variable under consideration. It should be the best indicator in determining statistical results.

REFERENCE
1. PATNAIK, P . B.: The power function of the test for the difference between two proportions in a 2×2 table, *Biometrika* 35, 157–175, 1948.