Special Article

Sample Size for Randomized Trials in Stroke Prevention. How Many Patients Do We Need?


The question of sample size is important, both in the planning of new investigations and in the evaluation of published results. Problems in the interpretation of negative trials are now widely recognized. When sample size is small, a negative result cannot be interpreted as providing evidence that the trial therapies are for all practical purposes therapeutically equivalent. In many instances the number of patients included in published trials is simply too small to provide an adequate comparison of the therapies under study. It is therefore imperative that careful thought be given to sample size calculations for future investigations.

This communication will provide a concise, nonmathematical guide to the decisions involved in planning the size of randomized clinical trials. Although the issues to be described are widely applicable, our clinical focus will rest on trials of new treatments for stroke prevention. Such trials typically employ the following design. Patients meeting specified inclusion/exclusion criteria are assembled over several months or years. As each patient is identified, he or she is randomly assigned to one of two possible treatment plans. The occurrence and date of important outcome events, usually stroke and death, are recorded during regularly scheduled follow-up visits, typically at 3 or 4 monthly intervals. All patients are followed to a predetermined final assessment date, typically one or more years past entry of the last case. When the trial is over, the difference between treatments is assessed by a class of statistical methods, known as survival analysis. These analyses compare treatment groups on the rate and timing of study outcome events, while taking into account the fact that some study patients have been followed much longer than others, and many patients have not suffered an event by the time the trial is completed.

Before attempting to determine the number of patients required for a trial employing this design, it is crucial that a final decision be made on the principal clinical question to be addressed by the trial. This decision requires definition of the patients to be enrolled, the treatments to be compared, and the events that will constitute treatment failure. Although consideration of the issues involved in making these decisions is beyond the scope of this paper, asking the right question at the right time is without doubt the most important element of any research program. Once the principal research question is clear, sample size calculations can proceed as follows:

Step 1. Anticipated Outcome in the Control Group

The first task is to estimate the rate of patient failure in the control or standard treatment group. If the hazard or yearly risk of failure is small, a large number of patients will be required in order to detect clinically interesting differences between treatments. Published series are usually the best source for such estimates, but careful consideration should be given to whether trial patients are likely to have a different level of risk due to the inclusion/exclusion criteria defined for the trial. Because uncertainty regarding this estimate exists during the planning of any trial, sample size calculations should be performed to explore the effect of reasonable departures from the best estimate that can be made from published series or other available data. Because the sample size requirement increases as the failure rate in the control group falls toward zero, overestimating the failure rate in the control group will lead to an underestimate of the required sample size. It is therefore wise to err on the side of being overly optimistic rather than pessimistic about the prognosis of patients in the control group.

Frequently there is interest in more than one patient outcome, and thus in more than one definition of patient failure. Death, stroke, and transient ischemic attack (TIA), have all been examined in "stroke prevention" trials. Typically, sample size is planned for the trials primary definition of patient failure. If for example a patient is considered to have failed if he or she suffers a stroke or dies from any cause, the date of failure is set once one of these occurs, and it is the annual risk of this happening that is required for sample size calculations.

Step 2. Therapeutic Effect Considered Clinically Important

The next task is to decide what degree difference in outcome, between the treatments under study, would be considered clinically important. Decisions made at this step have an enormous effect on the calculation of sample size. A good chance of reaching a statistically significant conclusion with a small number of patients will exist, if it is reasonable to expect a large difference in outcome between the study treatments. This fact

From the Department of Clinical Epidemiology and Biostatistics, McMaster University Medical School, Hamilton, Ontario L8S 4J9. Address correspondence to: Wayne Taylor, Department of Clinical Epidemiology and Biostatistics, McMaster University Medical School, Hamilton, Ontario L8S 4J9.

Received September 28, 1983; revision #1 accepted March 14, 1984.
frequently tempts investigators to propose that they are only interested in being able to detect large therapeutic effects, and that the trial is therefore feasible with a small sample size. However, if the true treatment effect is smaller, but still large enough to be considered clinically important, a trial with a small sample size will have a low probability of producing a statistically significant conclusion. Failure to establish a statistically significant difference between treatments in a small trial does not mean that the treatments are therapeutically equivalent. Negative results from small trials may prove adequate to rule out large treatment effects, but they are usually quite incapable of ruling out the possibility of a smaller but still clinically worthwhile therapeutic benefit.

A careful approach to this decision requires weighing a realistic appraisal of the potential benefits of the treatments under study against their potential risks and costs. A relatively small therapeutic advantage might be considered clinically important for a treatment which was cheap and entailed few medical risks. Prior clinical trials with treatments and clinical conditions comparable to those proposed for study, may help to provide a realistic estimate of the therapeutic benefit that might be achieved. In the absence of clinical data careful consideration should be given to the biologic rationale for the new treatment plan. If the pathogenic mechanism is multifarious and poorly understood, and the new treatment plan addresses a single component of this mechanism, it may be unrealistic to hope for a large therapeutic effect.

Step 3. Trial Duration

Next a decision must be made on the number of years that will be devoted to conducting the trial. If the prognosis of control patients is poor and a large risk reduction is anticipated with the experimental treatment, the trial can be completed in a relatively short period of time, and with a relatively small number of patients. When control prognosis is already good or a modest risk reduction is anticipated, longer followup and larger numbers of patients are clearly required. Thus, this decision depends in part on decisions already made in steps 1 and 2. However, it also depends on the speed with which patients can be enrolled and on the maximum number of patients that can be assembled. To achieve a statistically significant result, enough events must be accumulated for the difference between treatments to become apparent. When patients are at a relatively constant and continuing risk of suffering an outcome event, as is the case for the risk of stroke among patients with TIA, sufficient events may be accumulated by following many patients for a short period of time or a smaller number of patients for a longer period of time. Thus patient numbers and length of followup are, to a certain extent, interchangeable. Of course, if the trial is to determine whether a difference exists between short term and long term treatment effects, the trial must be planned for long term patient followup.

Step 4. Type 1 and 2 Error Rates

Finally, we must recognize that all data sets are subject to "random error" or noise that is unmeasured, unknown and unavoidable. Random error is due to all those factors over which we have no direct control, and is responsible for the "random variation" in results that is observed when exactly the same experiment is repeated several times. The act of randomization guarantees that "on average," these errors will cancel out between treatment groups. Occasionally however, random variation will produce a result that differs substantially from the truth leading to an incorrect conclusion from the trial at hand. We might incorrectly conclude that a treatment difference exists when in fact the treatments are equivalent (the type 1 or alpha error), or that the treatments do not differ when in fact one of them is therapeutically superior (the type 2 or beta error). Although the risk of these statistical errors cannot be reduced to zero, we do have control over the degree of risk that we are willing to take in any single trial. Typically these risks are set in advance at the conventional levels of 5% for the type 1 error and 20% or less for the type 2 error. Smaller risks can be selected but only at the expense of increasing the required sample size.

Confusion and debate exists over whether the type 1 error should be defined as "1 tailed" or "2 tailed". A 1 tailed test requires a larger sample size but permits the identification of therapeutic superiority, if it exists, in either of 2 directions (experimental treatment superior to control, or control treatment superior to experimental). Two tailed tests are most appropriate when the 2 treatments under study are roughly equivalent in terms of cost, immediate risk and side effects, because in such cases we want to identify the superior treatment plan whether it be experimental or control. In contrast, a 1 tailed test only allows determination of whether experimental treatment is superior to control treatment. A one tailed test is inappropriate when the experimental treatment entails greater cost, immediate risk or side effects, because in this situation the experimental treatment would only be recommended if its therapeutic advantage outweighed the accompanying disadvantages. Since the standard or control treatment would be preferred whether it was superior or equal to experimental treatment, it is not necessary to plan a trial large enough to demonstrate a therapeutic advantage with standard treatment. A one tailed test will be appropriate for example, with most trials of surgical vs medical treatment, because medical treatment would be preferred unless a trial had demonstrated that the immediate risks and costs of surgery were outweighed by its long term benefits.

Assumptions

Discussion of the mathematical formula used to calculate sample size, incorporating the decisions made in each of the preceding 4 steps, is beyond the scope of this paper. However, these calculations are not difficult to perform, and the formula can be found in most biomedical libraries. These formula involve two addi-
tional assumptions. First, that within each treatment group patients are at a constant level of risk (ie. that their event-free survival follows an exponential model); and second, that the rate of patient entry is constant during the accrual phase of the trial. Our experience in both the Canadian Cooperative Study of Threatened Stroke and the EC/IC Bypass trial indicates that the first assumption is met for the risk of stroke or death among patients with TIAs. If patient accrual is not constant and turns out to be slow at the start of the trial, these prior calculations will have underestimated the required sample size. This problem can be overcome by lengthening the trial.

The calculations performed for the tables contained in this paper also assume that the two treatment groups are to contain an equal number of patients. Although this is usually the best policy, it may sometimes be desirable to place a larger proportion of patients on the experimental treatment. This may be justified by the desire to gain more information on that treatment plan about which the least is already known, or to gain the cooperation of those investigators who are unwilling to have 50% of their patients placed on the control treatment. Provided all patients from all investigators are randomized in the same ratio no bias will be introduced by this proposal. A larger total sample size will however be required than for equal allocation.

Results

The following tables provide some useful examples. Tables 1a, 1b and 1c show the number of patients required in each of 2 treatment groups when step 1 produced an anticipated 5 year event rate in the standard or control treatment group of 20%, 30% and 40% respectively; step 2 led to the decision that a clinically important benefit would cut these event rates in half (a relative risk reduction of 50%); step 3 led to a maximum trial duration of 3 to 8 years, with all patients entered over 1 to 5 years; and step 4 led to the selection of an alpha (type 1) error of 5% (one tailed), and a beta (step 2) error of 10%. Tables 2a, 2b and 2c show sample sizes for the same decisions in steps 1, 3 and 4 but with a 30% risk reduction selected in step 2. Inspection of these tables illustrates the following general conclusions regarding the impact on sample size of the decisions made in steps 1 to 3.

1. Increasing the length of a trial reduces the required sample size because it increases the total number of events that will be observed among those patients included in the trial. Thus longer followup can be used to improve the power of small trials. Note the reduction in sample size as you read across the rows of these tables.

2. For trials of the same lengths, the faster patients can be accrued, the smaller the required sample size. This is because rapid accrual results in a longer average patient followup and, once again, increases the number of events that will be observed. Note the reduction in sample size as you read up the columns of these tables.

3. Trials among high risk patients require a smaller sample size to detect the same degree of relative risk reduction. Note the decrease in sample size as you move from table 1a, in which control patients have a 5 year event rate of 20%, to table 1c, in which control patients have a 5 year event rate of 40%.

4. Trials designed to detect small relative risk reductions will require more patients than trials designed to detect large relative risk reductions. Note the increase in required sample size between tables 1a, 1b and 1c, in which a 50% risk reduction is considered both clinically important and realistic, and tables 2a, 2b and 2c, in which we seek to demonstrate a 30% risk reduction.

As an example consider the design of a placebo controlled clinical trial to test a new drug regimen for patients with TIA. Suppose the risk of stroke among these patients is estimated at 7% per year. This corresponds to a 5 year event rate of 30% (1.0-EXP(-.35)) (step 1). We recognize from prior trials in this area that the risk reduction is likely to lie in the neighborhood of 30%, which is judged clinically important if true (step...
TABLE 2a Number of Patients Required in Each of 2 Treatment Groups. IF: —5 year failure rate = 20% in the control group, —5 year failure rate = 14% in the experimental group. —alpha = 0.05 (1 tailed), and beta = 0.10.

<table>
<thead>
<tr>
<th>Length of the patient accrual phase (yrs)</th>
<th>Total trial duration (in years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1285 935 741 618 533 470</td>
</tr>
<tr>
<td>2</td>
<td>1594 1082 827 674 572 499</td>
</tr>
<tr>
<td>3</td>
<td>2119 1291 938 743 619 534</td>
</tr>
<tr>
<td>4</td>
<td>— 1609 1089 831 676 574</td>
</tr>
<tr>
<td>5</td>
<td>— — 1304 945 747 622</td>
</tr>
</tbody>
</table>

TABLE 2b IF: —5 year failure rate = 30% in the control group, —5 year failure rate = 21% in the experimental group. —alpha = 0.05 (1 tailed), and beta = 0.10.

<table>
<thead>
<tr>
<th>Length of the patient accrual phase (yrs)</th>
<th>Total trial duration (in years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>748 551 441 372 324 289</td>
</tr>
<tr>
<td>2</td>
<td>925 635 490 403 346 305</td>
</tr>
<tr>
<td>3</td>
<td>1227 754 554 443 373 325</td>
</tr>
<tr>
<td>4</td>
<td>— 939 641 493 406 348</td>
</tr>
<tr>
<td>5</td>
<td>— — 766 560 446 375</td>
</tr>
</tbody>
</table>

TABLE 2c IF: —5 year failure rate = 40% in the control group, —5 year failure rate = 28% in the experimental group. —alpha = 0.05 (1 tailed), and beta = 0.10.

<table>
<thead>
<tr>
<th>Length of the patient accrual phase (yrs)</th>
<th>Total trial duration (in years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>482 359 291 248 218 197</td>
</tr>
<tr>
<td>2</td>
<td>593 411 321 268 232 207</td>
</tr>
<tr>
<td>3</td>
<td>785 487 362 292 249 219</td>
</tr>
<tr>
<td>4</td>
<td>— 605 417 324 270 234</td>
</tr>
<tr>
<td>5</td>
<td>— — 498 367 296 251</td>
</tr>
</tbody>
</table>

2) We wish to complete the trial in 5 years with all patients assembled in the first 3 years (step 3). And we select .05(1 tailed) and .10 as the acceptable risk of making type 1 and type 2 errors respectively (step 4). These specifications give a total sample size of 1108, 554 per treatment group (table 2b).

Discussion

When investigators are faced with numbers such as these, they are frequently thrown into despair at the size of the trials required to test their hypotheses. But as advances are made and it becomes necessary to test new treatments against regimens known to have some degree of efficacy, even larger sample sizes will be required. Unfortunately this realization frequently leads to unrealistic modifications of prior step 1 estimates of the anticipated event rate in the control group; of prior step 2 decisions about the size of an important and realistic therapeutic effect; of prior step 3 decisions about the speed with which patients can be enrolled, and the duration over which enthusiasm can realistically be maintained for continuing the trial; and even to the acceptance of dangerously high risks of making a type 2 error in step 4. It is of course possible to appropriately modify these estimates if we elect to ask a different clinical question. For example, we might restrict our interest to high risk patients, only enroll highly compliant patients and the most skilled clinicians, or include a less serious outcome such as TIA in the definition of treatment failure. However, if we wish to remain with the original clinical question the best solution is to cooperate in the conduct of large scale multicenter trials. It is our view that much more will be learned by such investigations than by a host of smaller trials incapable of answering clinically important questions. Although the administrative and logistical challenges of large scale trials must not be underestimated, cooperative trials have a number of advantages which justify the effort. They not only make large sample sizes attainable, but also increase the generalizability of the trials conclusions and permit an estimation of center to center variation in the results. A consistent result across many centers is both more compelling and useful than the same result obtained in a single institution. The frequent desire to perform exploratory analyses with clinically interesting patient subgroups, will only provide useful information if a large number of patients are enrolled in future trials. Cooperative trials also have the advantage of focusing the attention of the medical community on the clinical problem and treatments under study, and may thus prepare the way for more rapid dissemination of the trial’s results.

References

Sample size for randomized trials in stroke prevention. How many patients do we need?

D W Taylor, D L Sackett and R B Haynes

*Stroke*. 1984;15:968-971
doi: 10.1161/01.STR.15.6.968

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1984 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/15/6/968.citation

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org/subscriptions/