CHAPTER 13

Planning and Design of Clinical Trials (II)

In Chapter 12 we introduced the three phases of clinical trials. Among the three, a phase III or comparative trial is by far the most important since its results will affect future decisions concerning patients' lives or welfare. Several considerations in planning and designing comparative trials have been briefly discussed. In this chapter we focus our attention on several additional issues. Section 13.1 discusses the necessary elements of a protocol. Section 13.2 deals with the problem and methods of randomization. Section 13.3 is concerned with the use of prognostic factors. Finally, Section 13.4 discusses randomized versus historical controls in cancer clinical studies.

13.1 PREPARATION OF PROTOCOLS

Any scientific experiment requires a well-prepared plan. A protocol is a plan of scientific experiment or treatment according to Webster's New Collegiate Dictionary. There are no standard rules for the elements of a protocol, but in general, it should outline the purpose of the clinical study, details of experimental design, and the method of administering the treatment. Table 13.1 gives the major ingredients of a protocol for clinical trials.

A protocol should begin with its specific aims, a brief review of the history of the problem, and a rationale for doing the study. If other similar studies are being conducted or have been done, the results should be summarized along with the questions that remain unanswered.

Before undertaking a clinical trial, the investigator must have a clear objective in mind, and this objective must be well justified. A trivial question may not deserve a clinical trial. There should also be a statement of specific and well-defined objectives, which may be divided into primary and secondary. However, there should not be too many objectives, as such studies are too difficult to plan and carry out. The design of the study should ensure that at least the primary variables must be identified.

Table 13.1 Contents of a Protocol

<table>
<thead>
<tr>
<th>No.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Specific aims</td>
</tr>
<tr>
<td>2.</td>
<td>Introduction, scientific study</td>
</tr>
<tr>
<td>3.</td>
<td>Objectives (primary and base)</td>
</tr>
<tr>
<td>4.</td>
<td>Patient population and inclusion criteria</td>
</tr>
<tr>
<td>5.</td>
<td>Experimental design of the trial</td>
</tr>
<tr>
<td>6.</td>
<td>Treatment and administration</td>
</tr>
<tr>
<td>7.</td>
<td>Clinical and laboratory procedures</td>
</tr>
<tr>
<td>8.</td>
<td>Criteria for evaluating response</td>
</tr>
<tr>
<td>9.</td>
<td>Trial monitoring and frequency</td>
</tr>
<tr>
<td>10.</td>
<td>Procedures in the event of an emergency</td>
</tr>
<tr>
<td>11.</td>
<td>Statistical considerations</td>
</tr>
<tr>
<td>12.</td>
<td>Informed consent</td>
</tr>
<tr>
<td>13.</td>
<td>Data collection forms</td>
</tr>
<tr>
<td>14.</td>
<td>References</td>
</tr>
<tr>
<td>15.</td>
<td>Responsible investigator</td>
</tr>
</tbody>
</table>

A protocol should define the relationship of the investigator to the study. For example, if the trial is for melanoma, these stages are usually specified. The protocol must answer a number of questions: What is the purpose of the study? Who are the patients? What is the treatment? What is to be done? A protocol should be written in such a way as to allow the investigator to enter a patient into the study.

The investigator should be clear about the study period who have entered into the study. These patients might be selected based on characteristics and outcomes. Since most data have computerized data
Table 13.1 Contents of a Protocol

1. Specific aims
2. Introduction, scientific background, literature review, and significance of the study
3. Objectives (primary and secondary) of the study
4. Patient population and inclusion and exclusion criteria
5. Experimental design of the study
6. Treatment administration programs
7. Clinical and laboratory procedures and data to be collected
8. Criteria for evaluating response, nonresponse, and toxicity
9. Trial monitoring and frequency of interim analysis
10. Procedures in the event of early significant results
11. Statistical considerations: sample size, interim and final analysis strategies
12. Informed consent
13. Data collection forms
14. References
15. Responsible investigator and telephone number

ensure that at least the primary objectives are attained. The major response variables must be identified according to the primary objectives. If there are multiple response variables, sample size estimation should be done for each variable and the largest estimate used. This could lead to a very large sample size and the trial could be very expensive. This is another good reason for a very careful consideration of the study objectives.

A protocol should define clearly the type of patients that will be entered in the study. This, of course, is related to the objectives of the study. For example, if the trial is to compare two treatments for stage III and IV melanoma, these stages must be precisely defined and only those patients who satisfy the definitions are eligible for the trial. Other questions the protocol must answer are: If the differential diagnosis of the disease is difficult, is a confirmed diagnosis of the disease required and what constitutes acceptable confirmation? If an incorrect diagnosis is discovered later, what is to be done? Are patients who have received previous treatment acceptable or are only untreated patients to be included? If the study involves more than one clinic or hospital, it is even more important to document the inclusion and exclusion criteria so that every participating investigator enters exactly the same kind of patients.

The investigator should keep a log book of all patients seen during the study period who have a diagnosis of the disease under study. If a patient is not entered into the study, the reason should be given. Information from these patients might be valuable in making statements about overall patient characteristics and about any differences between participants and nonparticipants. Since most data analyses are done by computer, it is convenient to have computerized data sheets to record patient information such as medical
history and clinical and laboratory data. Every item on the form should be well defined and clear to the medical technicians who will fill out the forms, to the systems analyst who will build data files on the computer, and to the statisticians who will analyze the data.

The protocol must specify the experimental design. The investigator and the statistician must decide whether the trial is to be sequential or fixed sample, simple randomized or stratified randomized, open or blinded, and so forth. If patients are to be entered sequentially, a stopping rule should be given. If it is going to be fixed sample, sample size and treatment assignments have to be specified. Other questions that must be answered are: If a new treatment is to be compared with a standard treatment, will the controls be randomized or historical? If randomized controls are to be used, what type of randomization will be used? If historical controls are to be used, how will the selection procedure ensure that the historical patients are comparable to the current patients? Some estimate of the number of patients needed and duration of the trial should also be included in the protocol. Statistical considerations should be taken into account in determining sample size and the type of design. Statisticians can advise on these problems in the planning stage.

A comparative clinical trial is always a trial of an agent given according to a particular dosage schedule in a certain way to a certain type of patient; it is not a trial of the agent per se. A protocol should give a detailed description of the treatment program, that is, the regimen of treatment that each patient is to receive. This includes not only the total dose but also the method and administration schedule. If under certain circumstances, for example, severe adverse effect, the dosage is to be altered, the criteria for severity must be well defined and the altered dosage clearly specified. It is usually helpful to have schematic diagram outlining the entire treatment program.

In general, patients on a clinical trial are evaluated using clinical and laboratory procedures. Physical examination procedures as well as laboratory tests must be described in detail in the protocol. For example, if an electrocardiogram (ECG) is to be performed, the protocol should specify whether it will be a resting ECG or exercise ECG. Similarly, if blood will be drawn for glucose measurement, it should be specified whether the patients must fast for a given number of hours prior to the clinic visit. In addition, the schedule for clinical and laboratory examinations must be provided.

A protocol should contain criteria for evaluating treatment effectiveness. This includes response and adverse effects. The definition of response must be stated clearly. For example, complete response is usually defined as disappearance of all objective signs and symptoms of disease. All objective signs and symptoms of disease have to be listed explicitly. Other possible measures of response are time to recurrence, length of survival, time to development of metastasis, proportion of patients surviving a fixed time after treatment, or a certain percentage reduction in white blood cells. Some of these measures may take time for a chronic disease. The time to recurrence or increase should also provide prognostic toxicity. Rules for adjusting treatment are included. Information concerning the duration and level of toxicity on the data form.

To ensure that the protocol is being adhered to, the protocol should be corrected as early as possible after its initiation so that the investigator can report so that the investigator can report the deviation to the monitor. An additional measure is the comparison between groups involved in the study. The statistical considerations should be taken into account in determining sample size and the type of design. Statisticians can advise on these problems in the planning stage.

A comparative clinical trial is always a trial of an agent given according to a particular dosage schedule in a certain way to a certain type of patient; it is not a trial of the agent per se. A protocol should give a detailed description of the treatment program, that is, the regimen of treatment that each patient is to receive. This includes not only the total dose but also the method and administration schedule. If under certain circumstances, for example, severe adverse effect, the dosage is to be altered, the criteria for severity must be well defined and the altered dosage clearly specified. It is usually helpful to have schematic diagram outlining the entire treatment program.

In general, patients on a clinical trial are evaluated using clinical and laboratory procedures. Physical examination procedures as well as laboratory tests must be described in detail in the protocol. For example, if an electrocardiogram (ECG) is to be performed, the protocol should specify whether it will be a resting ECG or exercise ECG. Similarly, if blood will be drawn for glucose measurement, it should be specified whether the patients must fast for a given number of hours prior to the clinic visit. In addition, the schedule for clinical and laboratory examinations must be provided.

A protocol should contain criteria for evaluating treatment effectiveness. This includes response and adverse effects. The definition of response must be stated clearly. For example, complete response is usually defined as disappearance of all objective signs and symptoms of disease. All objective signs and symptoms of disease have to be listed explicitly. Other possible measures of response are time to recurrence, length of survival, time to development of metastasis, proportion of patients surviving a fixed time after treatment, or a certain percentage reduction in white blood cells. Some of these measures may take time for a chronic disease. The time to recurrence or increase should also provide prognostic information concerning the duration and level of toxicity on the data form.

To ensure that the protocol is being adhered to, the protocol should be corrected as early as possible after its initiation so that the investigator can report the deviation to the monitor. An additional measure is the comparison between groups involved in the study. The statistical considerations should be taken into account in determining sample size and the type of design. Statisticians can advise on these problems in the planning stage.

In general, patients on a clinical trial are evaluated using clinical and laboratory procedures. Physical examination procedures as well as laboratory tests must be described in detail in the protocol. For example, if an electrocardiogram (ECG) is to be performed, the protocol should specify whether it will be a resting ECG or exercise ECG. Similarly, if blood will be drawn for glucose measurement, it should be specified whether the patients must fast for a given number of hours prior to the clinic visit. In addition, the schedule for clinical and laboratory examinations must be provided.

A protocol should contain criteria for evaluating treatment effectiveness. This includes response and adverse effects. The definition of response must be stated clearly. For example, complete response is usually defined as disappearance of all objective signs and symptoms of disease. All objective signs and symptoms of disease have to be listed explicitly. Other possible measures of response are time to recurrence, length of survival, time to development of metastasis, proportion of patients surviving a fixed time after treatment, or a certain percentage reduction in white blood cells. Some of these measures may take time for a chronic disease. The time to recurrence or increase should also provide prognostic information concerning the duration and level of toxicity on the data form.

To ensure that the protocol is being adhered to, the protocol should be corrected as early as possible after its initiation so that the investigator can report the deviation to the monitor. An additional measure is the comparison between groups involved in the study. The statistical considerations should be taken into account in determining sample size and the type of design. Statisticians can advise on these problems in the planning stage.

In general, patients on a clinical trial are evaluated using clinical and laboratory procedures. Physical examination procedures as well as laboratory tests must be described in detail in the protocol. For example, if an electrocardiogram (ECG) is to be performed, the protocol should specify whether it will be a resting ECG or exercise ECG. Similarly, if blood will be drawn for glucose measurement, it should be specified whether the patients must fast for a given number of hours prior to the clinic visit. In addition, the schedule for clinical and laboratory examinations must be provided.

A protocol should contain criteria for evaluating treatment effectiveness. This includes response and adverse effects. The definition of response must be stated clearly. For example, complete response is usually defined as disappearance of all objective signs and symptoms of disease. All objective signs and symptoms of disease have to be listed explicitly. Other possible measures of response are time to recurrence, length of survival, time to development of metastasis, proportion of patients surviving a fixed time after treatment, or a certain percentage reduction in white blood cells. Some of these measures may take time for a chronic disease. The time to recurrence or increase should also provide prognostic information concerning the duration and level of toxicity on the data form.

To ensure that the protocol is being adhered to, the protocol should be corrected as early as possible after its initiation so that the investigator can report the deviation to the monitor. An additional measure is the comparison between groups involved in the study. The statistical considerations should be taken into account in determining sample size and the type of design. Statisticians can advise on these problems in the planning stage.

13.2 RANDOMIZATION

As expected, the result of the randomized design is a set of data. Results from a
of these measures may take a long time to observe, for example, survival time for a chronic disease, and thus require a lengthy trial time. In this case, time to recurrence or increase of disease may be more feasible. A protocol should also provide procedures in the event of severe side-effects and toxicity. Rules for adjusting dosage or stopping treatment should be given. Information concerning the type of effects and degrees of toxicity should be recorded on the data forms.

To ensure that the protocol is being followed during the trial, all clinical trials should be monitored closely. Any deviations from the protocol should be corrected as early as possible. Adverse effects need to be monitored and reported so that the investigator can take prompt action in case of severe toxic reaction. In addition, whether interim analysis for treatment comparison is to be performed should be specified in the protocol. This is an issue involving ethical, practical, and statistical considerations. If the trial lasts a long duration, methods for retaining subjects’ interest and for maintaining good compliance should also be discussed.

Statistical considerations include sample size determination and data analysis strategies. If interim analysis is to be performed, the protocol should describe when and how the data will be organized and how often it will be carried out. Decisions on the interpretation and actions to be taken should be well thought out and described in the protocol. The major objective for performing interim analysis is to avoid undue prolongation of a trial if the comparison between treatments is clear-cut so that more patients can be treated with a superior treatment. Therefore, it is critically important to determine which response variables are to be analyzed and what the stopping rules are. The stopping rules usually involve statistical significance level. However, it should not be the only rule.

In addition to the measure of response, required laboratory and clinical data should be specified in a protocol. Data collection forms may then be designed accordingly.

Other elements that should be included in a protocol are references, responsible investigator’s name and telephone number, form of informed consent, and personnel who are to coordinate the execution of the protocol. The last item is especially important in a collaborative clinical trial. The coordinator should know the study thoroughly, be able to prevent trouble or catch it quickly when it happens, and always be ready to help when needed. The coordinator’s duties often fall upon an applied statistician.

13.2 RANDOMIZATION

As expected, the results of a statistical analysis depend on the nature of the data. Results from a comparative trial depend on how patients were
assigned to the treatments. It is easy to conduct a comparative trial in such a way that the results are useless. For example, suppose treatments A and B are assigned to two groups of children with leukemia. Patients receiving A are two to six years old and patients receiving B are either younger or older. It has been shown by several investigators that “middle” age children have a better prognosis than the others. So, if the trial is conducted in this way, the observed difference between groups A and B would be an estimate of treatment difference plus the difference due to age. If results of the trial show a significant difference between groups A and B, there is no way to find out whether the difference is due to treatment alone, age alone, or a combined effect of treatment and age. In this case, we have a bias that is not difficult to discover. In other trials where little is known about prognostic factors or type of variability existing, unexpected biases can lead to erroneous conclusions. Thus, in a comparative clinical trial, it is essential to have comparable groups of patients. Randomization is one way to achieve comparability.

From a statistical viewpoint, to reach valid conclusions about populations by inference from samples, statistical tests and procedures typically assume that the sample are obtained in a random fashion. In clinical trials the term randomization refers to the assignment of patients to treatments by a random process in such a way that each patient has an equal and independent chance of receiving any of the treatments under study. That is, not only must each member in the population have an equal chance of being selected, but the selection of any member of the population must in no way influence the selection of any other member. An essential feature of randomization is that it should be objective and impersonal.

Despite the above advantages, the role and appropriateness of randomization continue to be among the most controversial in clinical trials. Randomization is considered a requirement for a scientific experiment by the classically trained statistician but is seen frequently as an unethical act by the classically trained clinician. This controversial issue will be discussed more in Section 13.4. In this section we assume that a randomized controlled design is chosen and discuss several most commonly used procedures of randomization: simple, restricted, and stratified randomization. For simplicity, we will assume that there are two treatments under study, A and B. We will also assume that patients enter a study sequentially.

13.2.1 Simple Randomization

Simple randomization is the most elementary and simple to apply. There are many informal ways to assign treatments to patients randomly such as flipping a coin and shuffling numbered cards. However, the most common technique is to use a table of random numbers such as Table C-20. In this table of digits 0–9, each digit has an equal and independent chance of appearing in every entry.
appearing in every entry, and each digit occurs with approximately equal frequency with no systematic pattern.

One acceptable scheme using a table of random numbers is to assign the even numbers to treatment A and the odd numbers of treatment B or numbers less than 5 to A and numbers greater than 4 to B. To illustrate the procedure, let us use the first row of Table C-20 to assign treatment to 20 patients. If the random number is even, assign treatment A; otherwise assign treatment B. The assignments are listed in Table 13.2. Of the 20 patients, 9 are assigned treatment A and 11 treatment B. In this case the ratio of A patients to B patients is very close to 1. However, simple randomization does not guarantee this, although the ratio approaches 1 as the number of patients increases. The main advantage of simple randomization is its simplicity to apply. Its main criticism is the possibility of unbalanced assignments.

13.2.2 Restricted Randomization (or Block Randomization)

One way to avoid unbalanced assignments is to use restricted randomization or block randomization. In this scheme, patients are grouped into several blocks of equal size according to their chronological entry time. Within each group of patients, the treatments are assigned so that there is a balanced allocation for each treatment. For example, if two treatments are to be assigned to 24 patients, it is reasonable to require balance after every four patients, assigning two to each treatment. We can divide the patients into six groups (or blocks) according to their entry times, the first four patients in group 1, the next four in group 2, and so forth. Then within each group we randomly assign two patients to each treatment. This will ensure that after

<table>
<thead>
<tr>
<th>Chronological Patient Number</th>
<th>Random Number</th>
<th>Treatment Assignment</th>
<th>Chronological Patient Number</th>
<th>Random Number</th>
<th>Treatment Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>B</td>
<td>11</td>
<td>5</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>A</td>
<td>12</td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>A</td>
<td>13</td>
<td>1</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>B</td>
<td>14</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>B</td>
<td>15</td>
<td>3</td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>B</td>
<td>16</td>
<td>9</td>
<td>B</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>A</td>
<td>17</td>
<td>7</td>
<td>B</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>B</td>
<td>18</td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>A</td>
<td>19</td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>A</td>
<td>20</td>
<td>8</td>
<td>A</td>
</tr>
</tbody>
</table>
every forth patient assignment, there will be an equal number of patients on each treatment.

To implement restricted randomization, we consider all the six possible arrangements of assigning two treatments to four patients:

1. AABB
2. BBAA
3. ABAB
4. BABA
5. ABBA
6. BAAB

First, we assign randomly these six arrangements to the six groups of patients. This can be done by rolling a die. If the number has appeared previously, roll again until each side of the die appears only once. Suppose the results of the die rolling are 2, 5, 4, 3, 1, and 6; the first group of four patients will be assigned the second arrangement (BBAA), the second group of four patients will be assigned the fifth arrangement (ABBA), and so forth. The overall assignment is given in the first two columns of Table 13.3. The 24 patients are equally assigned to treatments A and B. The advantage of the restricted randomization is obvious for a study involving only one clinic or hospital. However, if there are several hospitals participating in the study and the restricted randomization is made from the headquarters, it is possible that although the treatment assignment over all hospitals would be balanced, within a hospital there might be a serious imbalance.

For example, suppose there are four hospitals participating. The hospitals are numbered 1, 2, 3, and 4 and are given in column 3 of Table 13.3. Then the treatment assignment in column 2 results in an imbalance within hospitals 2 and 4 as shown below.

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Number Assigned to A</th>
<th>Number Assigned to B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>12</td>
<td>24</td>
</tr>
</tbody>
</table>

To ensure balance within a hospital, one could use the so-called balanced restricted randomization (or balanced randomization) in which treatment assignment balance within a hospital is ensured before the treatment is assigned. To implement it, we use a restricted randomization scheme with the help of an auxiliary table of random integers. These integers represent the difference in the numbers tolerated by the investigator in the integers 1 and 2 of each patient in the set. When a patient is entered, tentatively assigned a treatment, and assignment scheme is to be used, the rules is that a balance between assignment is less than a balanced randomization.

Suppose an auxiliary table was used, as follows. Suppose, for the example given, the integers obtained from the randomization scheme for the first eight patients are 1, 2, 4, 3, 5, 6, 7, and 8. The first eight assignments of patients to treatment A and B could then be determined by the randomization scheme.
### Table 13.3 Treatment Assignment by Restricted Randomization

<table>
<thead>
<tr>
<th>(1) Patient Number in order of Entrance Time</th>
<th>(2) Tentative Treatment Assigned (3) Hospital</th>
<th>(4) n</th>
<th>(5) Final Assignment Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>1</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>3</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>1</td>
<td>B</td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>3</td>
<td>B</td>
</tr>
<tr>
<td>11</td>
<td>B</td>
<td>3</td>
<td>B</td>
</tr>
<tr>
<td>12</td>
<td>A</td>
<td>4</td>
<td>B</td>
</tr>
<tr>
<td>13</td>
<td>A</td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>14</td>
<td>B</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>15</td>
<td>A</td>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>16</td>
<td>B</td>
<td>1</td>
<td>B</td>
</tr>
<tr>
<td>17</td>
<td>A</td>
<td>4</td>
<td>B</td>
</tr>
<tr>
<td>18</td>
<td>A</td>
<td>3</td>
<td>A</td>
</tr>
<tr>
<td>19</td>
<td>B</td>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>20</td>
<td>B</td>
<td>3</td>
<td>B</td>
</tr>
<tr>
<td>21</td>
<td>B</td>
<td>1</td>
<td>B</td>
</tr>
<tr>
<td>22</td>
<td>A</td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>23</td>
<td>A</td>
<td>3</td>
<td>A</td>
</tr>
<tr>
<td>24</td>
<td>B</td>
<td>2</td>
<td>A</td>
</tr>
</tbody>
</table>

The difference in the number of treatments assigned to A and B that can be tolerated by the investigator. In practice, this auxiliary table may have only the integers 1 and 2 or 1, 2, and 3. The procedure is as follows: When a patient is entered, tentatively assign a treatment according to the restricted randomization scheme. Calculate the difference in number of patients assigned to A and B with the tentative assignment. Then choose a random integer from the auxiliary table. If the difference $D$ in the treatment assignment is less than or equal to the random integer, the tentative assignment is to be used; otherwise the alternate treatment will be assigned.

Suppose an auxiliary random number table of integers 1, 2, and 3 is used for the example given above. Column 4 in Table 13.3 lists the random integers obtained from the auxiliary table. Column 5 gives the balanced restricted assignments of treatment for patients from the four hospitals. The first eight patients are assigned the same treatments as in the restricted
randomization (column 2). For the ninth patient in hospital 2, the tentative assignment was B in column 2. However, since the previous two patients in hospital 2 were assigned treatment B, the difference in treatment assignment D is 2 - 0 = 2, which is larger than the random integer 1 obtained from the auxiliary table. Therefore, the ninth patient is assigned the alternate treatment A. Similarly, for example, the fourteenth patient from hospital 2 is assigned treatment A since the difference would be 3 - 1 = 2 > 1 if the tentative treatment B were assigned. The resulting treatment allocation is balanced as shown below.

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Number Assigned to A</th>
<th>Number Assigned to B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

When preparing a balanced randomization list, it is also reasonable to require a balance at every sixth patient. This can be done by numbering all 20 possible arrangements of assigning two treatments to six patients as follows:

00–04 AABABB 50–54 BAAABB
05–09 AABABB 55–59 BAABAB
10–14 AABABB 60–64 BAABBA
15–19 AABABB 65–69 BBAAB
20–24 ABAABB 70–74 BAABA
25–29 ABAABB 75–79 BBAAB
30–34 ABAABB 80–84 BABA
35–39 ABAABB 85–89 BABA
40–44 ABBBAA 90–94 BABB
45–49 ABBBAA 95–99 BBAA

Then select randomly a two-digit random number from the random-number table. This can be done by closing your eyes and pointing to the random-number table haphazardly and starting with the nearest number. Suppose the random numbers so obtained are 70, 12, 91, 69, . . . ; the first six patients will receive treatments BBAABA in sequence, the second six will receive treatments AABBB, and so on.

This approach can be extended to three treatment cases (Peto et al. 1976). In order to obtain a random order of two A's, two B's, and two C's, we proceed as above but after selecting one of the 20 sequences, use the next two-digit random number to change one of the A's and one of the B's into C's. The following rules may be used.

<table>
<thead>
<tr>
<th>Random Number</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>01–11</td>
<td>First A, first B into C's</td>
</tr>
<tr>
<td>12–22</td>
<td>First A, second A into C's</td>
</tr>
<tr>
<td>23–33</td>
<td>First A, third A into C's, and use the next two-digit random number of ABAABB. Suppose the next A and third B into C's, and</td>
</tr>
<tr>
<td>34–44</td>
<td>Second A, second A into C's</td>
</tr>
<tr>
<td>45–55</td>
<td>Second A, third A into C's, and</td>
</tr>
</tbody>
</table>

That is, if the random number is the first B into C's; if the random number is the first A and second B into C's, and so on. If unbalanced allocation procedure can easily be extended if a 1 : 2 ratio is needed. After the allocation of C's to A's (for a 2 : 1 ratio).

13.2.3 Stratified Randomization

Prognostic factors that might influence the outcomes need to be taken into account in the design and conduct of randomized trials. However, the factors are not always given in Sec.

Taking account of these factors ensures that the distribution of each of the factors is approximately the same in the two treatment groups. For example, if some factors are important, a stratified randomization will ensure that the two treatment groups are well balanced in the proportion of patients with each factor.

Suppose one decides to stratify by sex, with favorable prognoses for males and unfavorable prognoses for females. The simplest stratification for sex is to allow only one sex to be randomized in a single clinic or multicenter trial. If randomization is made within institutions as well.

If there are several important prognostic factors in a study, several subcategories, then the number of strata for randomization will be very large. Patients in each stratum are randomized separately, and this will be repeated in each strata.

For example, if sex and age are the important prognostic factors, then the study is stratified by sex and age. Each stratum is randomized separately, and this will be repeated in each strata.
That is, if the random number is between 01 and 11, change the first A and the first B into C’s; if the random number is between 12 and 22, change the first A and second B into C’s; and so on. If the random number is 00, ignore it and use the next two-digit number. For example, suppose the first two-digit random number obtained is 21. The sequence selected would be ABAABB. Suppose the next random number is 57. We change the second A and third B into C’s, and thus the treatment assignment is ABCABC.

If unbalanced allocation of two treatments is desired, the above procedure can easily be extended. Suppose a treatment allocation ratio of 2 : 1 or 1 : 2 is needed. After the above A, B, and C assignment, either change all the C’s to A’s (for a 2 : 1 ratio) or change all the C’s to B’s (for a 1 : 2 ratio).

### 13.2.3 Stratified Randomization

Prognostic factors that might influence response, when known, should be taken into account in the initial randomization. Not to do so may introduce biases into the data and thus lead to incorrect conclusions about the treatments under study. Discussions of the use and importance of prognostic factors will be given in Section 13.3.

Taking account of prognostic factors in the initial treatment assignment ensures that the distributions of patients with respect to the important prognostic factors are equal. For example, it is known that age is an important factor in childhood leukemia. In a study of two treatments for childhood leukemia, this fact should be seriously considered. Stratified randomization will ensure equal distributions of patients with regard to age in the two treatment groups. In other words, there should be an equal proportion of patients with good prognosis in each treatment group.

Suppose one decides to consider two strata, one consisting of patients with favorable prognoses and the other consisting of patients with poor prognoses. The simplest stratified randomization is to make up a separate restricted randomization schedule for each stratum. This can be done either in a single-clinic or multiclinic trial. In a multiclinic trial, adjustments are made within institutions as described earlier to prevent imbalances.

If there are several important prognostic factors involved and each has several subcategories, then the number of combinations and consequently the number of strata for randomization may be enormous. There may be very few patients in each stratum. Such a randomization is equivalent to one...
in which there is a single list for all patients or simple randomization. Sometimes, if there are several important prognostic factors, a summary prognostic index can be used as a single stratifying variable to achieve balance among the separate variables on which it is based. An example is the hazard index or hazard ratio $h_i(t)/h_0(t)$ or $\log[h_i(t)/h_0(t)]$ discussed in Section 10.2.2.

Zelen (1979) also proposed a new design for randomized trials. Patients are randomly assigned to two groups, A and B. Patients in group A receive the standard treatment; those in group B are asked if they will accept the experimental treatment; if they decline, they receive the standard treatment.

Statistical properties of various randomization methods in clinical trials have been discussed by many. For example, the December 1988 issue of the journal Controlled Clinical Trials is devoted to this topic. General recommendations are given regarding the use of several randomization methods. Among the recommendations offered, one is that the simple unrestricted complete randomization is desirable for large trials ($n > 200$ overall and within each subgroup) especially in unblinded trials. On the other hand, for small trials with $n < 100$ overall or within any subgroup, restricted randomization is better. The interested reader is referred to this issue of the journal.

### 13.3 THE USE OF PROGNOSTIC FACTORS IN CLINICAL TRIALS

The primary objective of most clinical trials is to obtain a precise comparison of treatments, usually a new treatment and a standard treatment. An important requirement for a precise comparison is that patients in the different treatment groups are comparable with respect to prognostic characteristics except the treatments being assigned. If a small number of prognostic variables have been identified, how can this information be used in a clinical trial so that treatment comparisons can be more precise? The information can be used to stratify patients at the stage of randomization (stratified randomization or prestratification) in order to achieve comparability in the treatment groups. It can also be used at the analysis stage (poststratification) to adjust for and minimize the effects of group differences. The use of prognostic factors in prestratification and poststratification has been discussed by many, for example, Armitage and Gahan (1974), Pocock and Simon (1975), Peto et al. (1976), Feinstein (1977), Simon (1979), Feinstein and Landis (1976), Armitage (1981), Meier (1981), Grizzle (1982), and McHugh and Matts (1983). In this section we briefly discuss the advantages and limitations of prestratification and poststratification. For simplicity, we assume that two treatments are to be compared.

As mentioned in Section 12.3.2, knowledge of prognostic variables could be used to group patients into prognostic strata. Patients within a stratum have similar values of the prognostic variable used to define the strata and patients among strata are substantially different. For example, if previous studies indicate that female patients, patients could be randomized. Within each stratum, treatment or block randomization or block randomization by sex distribution in the two treatment comparisons. A comparison of treatments A and B can be carried out within stratum and within the male stratum alone. If B is better than A within the male stratum alone (in the extreme case), then there is no need for sex (Armitage 1981). This is a basic of treatment selection.

However, prestratification in randomization, values of the variable at the time of randomization for treatments can easily be met if available at the time of randomization has no absolute definition. Stress or degree of stenosis in stratum is high. It happens on the information about the patient's age, the time of stratified randomization may be found erroneous for all strata.

In large clinical trials the chance of achieving group in small trials. So, prestratification is needed. However, the advantage may be large. For example, age, sex, three age levels and four race Caucasians.

Age 40-49
Sex Male
Race Caucasian

The total number of strata will be more than one and less than or equal to 100, each stratum will be compared within stratum. The interaction would not be made to it for patients that fit in some of the.

In addition, some of the will be insignificantly related to pre
studies indicate that female patients tends to survive longer than male patients, patients could be divided into two strata, males and females. Within each strata, treatments A and B are assigned using simple randomization or block randomization. This would guarantee a comparable overall sex distribution in the two treatment groups and increase the precision of treatment comparisons. Another advantage is that in addition to an overall comparison of treatments A and B, comparisons of the two treatments could be carried out within stratum. For example, we could compare A and B within the male stratum and within the female stratum. If the results show that A is better than B in males and B is better than A in females (an extreme case), then there exists a strong interaction between treatment and sex (Armitage 1981). This information would be useful to physicians in treatment selection.

However, prestratification is not problem free. In order to use stratified randomization, values of the prognostic factors must be known exactly at the time of randomization. For variables such as age and sex, this requirements can easily be met with little risk of error. If any time-consuming laboratory test result is used to classify, the information may not be readily available at the time of randomization. Or if the variable used for stratification has no absolute definition or involves subjective judgment, for example, stress or degree of stenosis of a coronary artery, the chance of misclassification is high. It happens often that the physician does not have complete information about the patient or may have to make a personal judgment at the time of stratified randomization. The classification made at that time may be found erroneous later and the patient ineligible or being in the wrong stratum.

In large clinical trials that involve hundreds or thousands of subjects, the chance of achieving group comparability in prognostic factors is higher than in small trials. So, prestratification is more likely to be used in small studies. However, the advantage may be limited if the total number of strata is large. For example, age, sex, and race are used to stratify and there are three age levels and four race groups:

- Age: 40–49 years, 50–59 years, ≥60 years
- Sex: Male, female
- Race: Caucasian, black, Hispanic, other

The total number of strata is $3 \times 2 \times 4 = 24$. For each of the 24 strata, we assign one-half the patients to A and one-half to B. If the total sample size is less than 100, each stratum will have no more than five patients. In this case, comparisons within stratum and the study of treatment–prognostic variable interaction would not be meaningful. Sometimes it may be difficult to find patients that fit in some of the strata.

In addition, some of the variables used to stratify may turn out to be insignificantly related to prognosis and other more important variables may
be discovered. The effort of stratification at randomization may become totally unworthy. Therefore, the merit of stratified allocation is debatable and most statisticians agree that poststratification may be more desirable.

By poststratification we mean that the recognized prognostic variables are adjusted using statistical methods at the analysis stage. This method, often referred to as stratified analysis or covariate adjustment, is recommended for both randomized and nonrandomized studies. It not only minimizes the effect of imbalance between treatment groups and improves efficiency in comparison but also avoids erroneous conclusions due to confounding factors.

In order to minimize the effects of prognostic heterogeneity, the treatment groups can be divided into subgroups, or strata, the prognostic variable at the time of analysis. This is usually done if the prognostic variable is categorical, discrete, or convertible into intervals. For example, systolic blood pressure may be converted into two intervals (or strata): \(<160\) mm Hg, \(\geq 160\) mm Hg. A comparison is then made within each stratum and then over all strata to obtain an overall comparison. If the outcome variable is also categorical (e.g., response or nonresponse), the stratified analysis can be performed using the Mantel–Haenszel chi-square test described in Chapter 5. The following example illustrates the importance of stratified analysis.

**Example 13.1**
Suppose that two treatments A and B are assigned randomly to 100 patients and the outcome variable of interest is response. At the end of the clinical trial, the following results are obtained:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Response</th>
<th>Nonresponse</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30 (60%)</td>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td>B</td>
<td>15 (30%)</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>55</td>
<td>100</td>
</tr>
</tbody>
</table>

Sixty percent of the patients receiving treatment A and 30% of the patients receiving B responded. We might conclude that the two treatments are distinctly different. However, suppose that previous studies showed a better response rate in females and that sex might be a prognostic variable. When the patients are stratified by sex, we find the following results:

<table>
<thead>
<tr>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Response</td>
<td>Nonresponse</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>A</td>
<td>2 (13%)</td>
<td>13</td>
</tr>
<tr>
<td>B</td>
<td>4 (11%)</td>
<td>31</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Response</th>
<th>Nonresponse</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>29 (80%)</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>B</td>
<td>11 (70%)</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>11</td>
<td>50</td>
</tr>
</tbody>
</table>

Cancer, one of the most feared diseases in the United States, has entailed four decades. In cancer clinical trials, the standard treatment is usually the only option available. In order to evaluate effectiveness, how should one compare two randomized controlled clinical trials? Patients are treated with each of the two treatments and a comparison is made. The latter is usually to control for prognostic factors.

**13.4 CONTROLS IN CANCER CLINICAL TRIALS**

The Mantel–Haenszel chi-square test can be used to determine if there is a significant difference between the two groups. The overall significant result obtains if the sex distribution between the two groups differs. For example, 30% males and 30% females in group A and 30% males and 30% females in group B. The difference between the two groups is statistically significant if the Mantel–Haenszel test is significant. The Mantel–Haenszel test is a useful tool for detecting if the treatment effect is consistent across the strata. If the treatment effect is consistent across the strata, the Mantel–Haenszel test is significant. If the treatment effect is not consistent across the strata, the Mantel–Haenszel test is not significant.

Although poststratification can be used to adjust for covariates in a sample size in a given stratum, the overall sample size is generally not affected. Although poststratification can be used to adjust for covariates in a sample size in a given stratum, the overall sample size is generally not affected. Although poststratification can be used to adjust for covariates in a sample size in a given stratum, the overall sample size is generally not affected. Although poststratification can be used to adjust for covariates in a sample size in a given stratum, the overall sample size is generally not affected.
The Mantel-Haenszel chi-square test shows that there is no significant difference between the two treatments after adjusting for sex. Thus, the overall significant result obtained earlier is mainly due to the imbalance in sex distribution between the two treatment groups (70% females in group A and 30% females in group B). After the effect of sex is controlled, the difference between the two treatments becomes negligible. Similarly, stratified analysis can also be used to avoid false-negative results and to detect if the treatment difference exists only in a certain subgroup. Stratified analysis can be performed using variables that have not been reported as prognostic factors. Thus, it can be used to detect and identify important confounding variables as well as prognostic factors.

Although poststratification can improve comparison efficiency, when the sample size in a given stratum is small, the disproportion can cause average loss in efficiency. Meier (1981), in comparing the relative efficiency between a balanced and unbalanced design, concludes that the loss is not big. For sample sizes as large as 20 (10 in each group), the expected relative efficiency is close to 100%. Thus, for moderate sample sizes, poststratification is as efficient as prestratification.

In addition to poststratification, the linear logistic regression method discussed in Chapter 11 can be used to adjust for the effect of prognostic factors when the outcome variable is dichotomous and the prognostic variables are either categorical or continuous.

If the outcome variable is continuous, the analysis-of-covariance technique can be used to adjust for prognostic factors. This topic is not discussed in this book but can be found in most standard statistics textbooks, for example, Snedecor and Cochran (1967), Kleinbaum, Kupper, and Muller (1988), and Howell (1987). If time to a given event (e.g., survival time) is the outcome variable of interest and censored observations are involved, Cox’s proportional hazards model discussed in Chapter 10 can be used to adjust for prognostic factors.

### 13.4 CONTROLS IN CANCER CLINICAL STUDIES

Cancer, one of the most feared, publicized, and politicized diseases in the United States, has entailed numerous clinical trials, especially in the past few decades. In cancer clinical trials, a new treatment, say A, frequently becomes available for patients with a certain type of cancer. If treatment B is the standard treatment generally accepted and used, though of very low effectiveness, how should one evaluate the new treatment A? Should a randomized controlled clinical trial be conducted in which half of the patients are treated with each treatment? Or should all the patients be given treatment A and a comparison made with patients previously treated with B? The latter is usually termed as unrandomized controlled or historical controlled clinical trial.
The two kinds of controls, randomized and historical, have drawn many discussions. Chalmers et al. (1972), Ingelfinger (1970), Hill (1971), Sacks et al. (1982), and Micciolo et al. (1985) have emphasized the importance and advantages of randomized controls. Gehan and Freireich (1974), Freireich and Gehan (1974, 1979), and Gehan (1982a, b) argue that there still exists a place for a careful choice of historical controls. Pocock (1976a,b) suggests a compromise in which the comparative evaluation of a new treatment uses both randomized and historical controls. In the following, we discuss briefly randomized controls and historical controls.

13.4.1 Randomized Controls

In the early 1950s, Daniels and Hill (1952) first employed randomization in their studies of streptomycin combined with para-aminosalicylic acid in the treatment of pulmonary tuberculosis in young adults. Since then, treatment randomization has been an important aspect of clinical trials. Methods of randomizations have been developed, four of which are discussed in Section 13.2. In general, the advantages of randomization (or randomized controls) are that it achieves comparability of patients among treatment groups, it avoids conscious or unconscious bias in the assignment of patients to treatment groups, and it provides a firm basis for the statistical evaluation of any apparent treatment effects.

Some disadvantages of randomization were mentioned earlier. Possible unbalanced assignment of treatments due to simple randomization can be eliminated by using restricted or block randomization. Some disadvantages of restricted randomization, in turn, can be avoided by stratified randomization. However, if importance prognostic factors are unknown or if there are too many prognostic factors, stratified randomization is not easy to apply. Although randomization does not ensure that patients in the different treatment groups are comparable for all important prognostic factors, the validity of statistical tests and significance levels based on randomization does not require this unachievable assumption. As mentioned earlier, the effects of important prognostic factors can be adjusted in the analysis in order to correct treatment comparisons for possible bias due to imbalances.

The most serious objections to randomized controls center around ethical responsibility, which requires a physician to administer the treatment that he or she believes is best for a patient. However, the relative merits of the treatments under study are yet to be determined. If it is known that one treatment is better than the other, then there is no need for comparative clinical trial. Therefore, researchers advocating randomized controlled clinical trials argue that a physician initiating a comparative trial makes the honest admission that the best treatment is unknown and therefore randomization is more ethical than a procedure in which the merits of a new therapy are determined from clinical impression and comparison with past experience.

CONTROLS IN CANCER CHEMOTHERAPY

Despite the disadvantages, useful clinical information can be obtained by expensive, and time consuming methods for comparing randomized controls with conventional, randomized controls by Gehan (1975). They indicated that the small sample size in the new treatment control group makes patients receiving conventional treatment (Gehan, 1974) disagree. The only objective of phase I clinical trials is to test the new drug and it is not meant to obscure the results of phase II trials. They propose that alternate controls should be used in certain circumstances.

13.4.2 Historical Controls

Historical controls improve our understanding of a disease process from a previously established treatment sequence of studies. The selection of prognostic factors is important to define a group or patient to which there is a firm basis for assigning the new drug. However, the difference existing between the characteristics of patients treated with the new drug and the characteristics of patients treated with the conventional drug is not a valid measure of the outcome of treatment. Thus, the characteristics of the patients with the new drug must be comparable to the characteristics of the patients with the conventional drug.

A. Literature Controls

Literature controls are compared in patients with the same disease who received the same treatment. B. Controls reported in the literature must define a control group to detect and define the current patients and treatment A which is valid provided the patients in group B are comparable in present treatment. See Carbone et al. (1979) and Donadio and O'Ferrall (1979).

Obviously, literature controls are valid and draw criticism if they are comparable or are not. Today, such comparisons are not sufficient data. Tends to the use of groups and computerized data, to retain the raw data collected in the clinical trials may use the data. This may be very tedious.
Despite the disadvantages of randomized controlled clinical trials, much useful clinical information has been derived using them. They are complex, expensive, and time-consuming but remain the most useful and acceptable methods for comparing treatments. Chalmers et al. (1970) even suggest that randomized controls be used in phase I and II studies (see also Chalmers 1975). They indicated that dosage can be randomized to give some insight into cumulative effects, and patients destined to receive multiple dosages of the new treatment could for ethical reasons be given a 50–50 chance of receiving conventional or placebo therapy. However, Gehan and Freireich (1974) disagree. They state that since comparing treatments is not an objective of phase I or II studies, randomized controls patients would only tend to obscure the real purpose of the studies and delay their completion. They propose that a selected rather than a randomized control group be used in certain circumstances.

13.4.2 Historical Controls

Historical controls include patients chosen from the literature, on a matching basis from a previous study, or from an immediately preceding trial in a sequence of studies. Gehan and Freireich (1974) state that knowledge of prognostic factors is the primary assumption in any selection of controls so that there is a firm basis for the selection of comparable groups and that the difference existing between the groups selected have little or no relation to the outcome of treatment. The investigator must hope that unknown prognostic factors are distributed equally to the two treatment groups.

A. Literature Controls. Suppose that treatments A and B are being compared in patients with a particular disease and that patients with the same disease who received one of the two treatments (say, A) have been reported in the literature. These patients can be selected as a historical control group to determine whether B is superior to A. In this case, all of the current patients are given treatment B. The comparison between A and B is valid provided that the literature control group and the current group are comparable in prognostic factors. Examples of literature controls can be found in Carbone et al. (1968), Frei et al. (1973), Reaman et al. (1987), and Donadio and Offord (1989).

Obviously, literature controls require careful checking for comparability and draw criticism when the patients either cannot be checked for comparability or are not comparable because the authors do not provide sufficient data. Today, many studies are conducted by large cooperative groups and computers are commonly used in analyzing data; it is possible to retain the raw data on magnetic tapes or diskettes so that other investigators may use the data. The detail work of checking for comparability could be very tedious.
B. Matched Controls. Matched controls are patients selected on a matching basis from a previous or concurrent study. Controls are matched with cases by using known prognostic factors. It is obvious that the use of this type of controls is feasible when important prognostic factors are known and there are not so many of them that pair matches are too difficult to find. Gehan and Freireich (1974) suggest selecting two control patients for each treated patient to test the selection process. These two control patients and the treated patient are as comparable as possible with respect to factors influencing prognosis. The randomness of the selection can be checked by comparing the two control groups of patients with respect to the endpoints in the analysis. If no differences are found, the validity of the selection process is confirmed. Examples of matched controls are given by Bodey et al. (1971), Hogan et al. (1987), and Ozer et al. (1989).

C. Controls Selected from a Sequence of Studies. Historical controls can also be selected from a sequence of studies, for example, Buzdar et al. (1978), Micciolo et al. (1985), Abdi et al. (1987), and Hersey et al. (1987). In many clinical research centers or cooperative groups for cancer research, a sequence of clinical trials is often conducted for a given disease. It is common to activate one trial for a given disease as soon as the previous one is terminated. Then the best treatment in the previous trial will usually be the standard control in the next trial. Again the patients selected must be comparable with respect to prognostic factors.

This type of historical control is reasonable, feasible, and probably the most acceptable. Since the controls have been treated in the recent past, criteria for diagnosis, types of patients, nature of the disease, means of treatment administration, supportive treatment, definition of response, and staff probably have not changed. Even if some changes have occurred, the nature of the changes would be known, and appropriate adjustments can be made. On the other hand, if a relatively long time has elapsed between studies or important changes have occurred in type of patients, supportive treatment, definition of response, or staff, controls from a previous study could lead to serious bias.

Important reasons for considering historical control groups are the following:

1. When patients are randomized to a control or treatment group, the comparability achieved with respect to prognostic factors between groups is an average. When historical controls are selected, all patients in the two groups are guaranteed comparable with respect to the characteristic influencing prognosis.

2. The use of historical controls requires fewer patients and therefore a shorter time and less money. Suppose two treatments A and B are to be compared according to the proportion of patients responding to each treatment (response rate $\hat{p}_A$ and $\hat{p}_B$). Suppose further that enough patients will receive each treatment that a statistical test of the difference between level $\alpha$ and power 1-. $\beta$ number can then be expanded into a treatment group. All treatment combinations of this kind are used in the trial. To make $\hat{p}_A$ at the same significance level as $\hat{p}_B$ at the same time, the number of patients is needed. The number of patients needed is evident by comparison of two differences.

3. A comparative trial can benefit from evidence from initial studies. If there is evidence from initial studies, the trial has a good chance of being successful. Therefore, clinical responsibility if they are in the process of administering the medication. For example, there is no justification for introducing an old medication to concurrence patients when the new medication consists of the old medication and preliminary studies have shown it to be effective as the standard one. One must consider historical controls.

The main objection to this method is the protection against possible bias in the patient population, due to the pathologic agents. A subset of patients that require treatment, that is, the results obtained with the pathologic agents may be biased if the pathologic agents given an example in which the pathologic agents are breast cancer after chemotherapy. Pocock (1976a) gives an example of an acceptable historical control study:

1. The historical control group contains the same characteristics as the patients in the sample.

2. The group must have been treated with the same therapy that the patients in the sample would have received.
difference between treatments can be made at a given significance level \( \alpha \) and power \( 1 - \beta \). From Table 12.5 or 12.6, the required patient number can then be found. Assume that \( n \) patients are needed in each treatment group. Alternatively, suppose \( B \) is the control treatment and its response rate is well known, say \( P \). Then no patients need receive \( B \) in the trial. To make a statistical test of the difference between \( \hat{P}_A \) and \( P \) at the same significance level and power assumed above, only \( \frac{1}{4}n \) patients are needed on treatment \( A \), that is, only one-fourth the total number of patients needed for the randomized comparative trial. This is evident by comparing relevant tables in Natrella (1963) and Table 12.6. Readers interested in the derivation are referred to Sillitto (1949).

3. A comparative trial should be undertaken only when the best available evidence from initial clinical studies suggests that the new treatment has a good chance of being equal to or better than the old treatment. Therefore, clinical investigators are not fulfilling their ethical responsibility if they plan a randomized comparative trial instead of administering the better treatment to consecutive patients. For example, there is no justification for adding to the anguish of a cancer patient by introducing the irrational concept of “flipping a coin.” Evaluation of the treatment could proceed by comparison of results in concurrent patients with a historical control group. If the new treatment consists of the standard therapy plus an additional component and preliminary studies suggest that the new combination is at least as effective as the standard and possibly much more so, it is logical to consider historical controls.

The main objection to historical controls is that they provide limited protection against possible bias introduced by time changes in the nature of the patient population, diagnostic criteria, patient care, and exposure to pathologic agents. A subtle selection mechanism may require one to check a long list of prognostic factors to find a match. It is also very likely that prognostic data of historical controls are not available. Micciolo et al. (1985) give an example in which the results obtained by using historical control data in breast cancer after adjusting for differences in baseline characteristics given a biased conclusion in favor of the new treatment.

Pocock (1976a) gives the following conditions as requirements for an acceptable historical control group:

1. The historical control group must have a precisely defined standard treatment, that is, the same as the treatment for the randomized controls.
2. The group must have been part of a recent clinical study that contained the same requirements for patient eligibility.
3. The methods of treatment evaluation must be the same.
4. The distributions of important patient characteristics in the group should be comparable with those in the new trial.
5. The previous study must have been performed in the same organization with largely the same clinical investigators.
6. There must be no other indications leading one to expect different results between the randomized and historical controls.

13.4.3 The Combination of Randomized and Historical Controls

Pocock (1976a,b) suggests a compromise in which both randomized and historical controls are used. Let $T$, $R$, and $H$ denote the groups of patients on the new treatment, the randomized control, and the historical control, respectively. Let $N_T$, $N_R$, and $N_H$ be their respective sample sizes. The two questions discussed by Pocock using the Bayesian approach are:

1. How should one determine sample sizes $N_T$ and $N_H$ given the historical controls of size $N_H$?
2. How can one combine the data from the two sets of controls for comparison with data from the patients on the new treatment?

Suppose that the evaluation of treatment for a patient can be summarized by a single quantitative measurement $X$. For example, in a cancer clinical trial, $X$ might be the patient’s survival time, remission duration, or an ordered scale for the assessment of objective tumor response. Let $x$ be the observed value of $X$ for any patient in a particular treatment group. Suppose that the random variables associated with the three groups, $X_T$, $X_R$, $X_H$, follow the normal distribution with unknown means $\mu_T$, $\mu_R$, and $\mu_H$ and variances $\sigma_T^2$, $\sigma_R^2$, and $\sigma_H^2$. The sample means of $X_T$, $X_R$, and $X_H$ are $\bar{x}_T$, $\bar{x}_R$, and $\bar{x}_H$. The main objective of the comparative trial is to estimate the difference in effectiveness between the new treatment and the control treatment. In other words, the objective is to obtain an accurate estimate of $\mu_T - \mu_R$. In a randomized control trial without historical data, this is best estimated by $\bar{x}_T - \bar{x}_R$.

In the presence of historical data, $\bar{x}_H$ has to be incorporated. It should be noted that $\mu_T$ and $\mu_R$ are not necessarily equal. There exists a possible unknown bias $\delta = \mu_T - \mu_R$ in the historical control. We are unable to assess the exact value of $\delta$. A possibility is to assume that $\delta$ is a normal random variable with mean zero and variance equal to some fixed quantity $\sigma_\delta^2$. The choice of $\sigma_\delta$ is difficult in practice. On the side of caution a reasonably large value should be assigned. That is, it is better to place too little rather than too much confidence in historical data. Perhaps in practice one should consider several possible values of $\sigma_\delta$ in order to assess the effects on experimental design and analysis.

Sample Size Determination

Assume that the total sample size suggests that the “optimal” value of $\mu_T - \mu_R$ given the values

$$N_T, N_R, N_H$$

where $\sigma_T$, $\sigma_R$, and $\sigma_H$ are relatively small. However, prior to the trial, one may estimate $\sigma_T^2$. The three variances are assumed to be equal so that the variance of the historical treatment is $3\sigma_T^2$. Equation (13.1) reduces to

$$N_T + N_H = 3N_H$$

Having chosen $N$ and $N_H$ (e.g., $\frac{1}{3}$ or $\frac{2}{3}$) so that small trials are available; every three patients it is one for the old and two for the new treatments. The procedure is then:

**Example 13.2**

A trial is to be undertaken to compare the effectiveness of Procarbazine with Cytoxan. The main problem is to decide which of 200 patients be entered on Cytoxan compared to Cytoxan + CCNU. In this earlier work, Cytoxan + CCNU showed superior. The tumor response was at least 83 cases. The question is how many patients will the treatment Cytoxan + CCNU improve and already 83 patients in a year.

In this case, $N_H = 83$. The historical distribution, $\sigma_H$ is estimated that $\sigma_T = \sigma_H = 49.70$. The historical controls $\sigma_T = \sigma_H = 49.70$. The size

$$N_T = \frac{N_H}{3}$$

$N_T$ is close to $\frac{1}{3}N = 66.66$ patients. It would be to randomize patients between Procarbazine in a ratio...
Sample Size Determination

Assume that the total sample size \( N = N_1 + N_2 \) has been determined. Pocock suggests that the "optimal" value for \( N_r \) (which minimized the variance of \( \mu - \mu_r \)) given the values of \( N \) and \( N_h \) is

\[
N_r = \frac{\sigma_r}{\sigma_r + \sigma_i} \left( N - \frac{\sigma_r \sigma_i}{\sigma_h^2/N_h + \sigma_i^2} \right) \tag{13.1}
\]

where \( \sigma_r \), \( \sigma_h \), and \( \sigma_i \) are unknown and some estimates must be used. However, prior to the trial, we do not know the sample variance of \( \sigma_r^2 \) and \( \sigma_i^2 \). The three variances \( \sigma_r^2 \), \( \sigma_h^2 \), and \( \sigma_i^2 \) may all be set equal to the sample variance of the historical data \( s_h^2 \), where \( s_h^2 = (x - \bar{x}_h)^2/(N_h - 1) \). Thus (13.1) reduces to

\[
N_r = \frac{1}{2} \left( N - \frac{N_h}{1 + N_h \sigma_h^2/\sigma_i^2} \right) \tag{13.2}
\]

Having chosen \( N \) and \( N_r \), \( N_r/N \) may be rounded off to some simple fraction (e.g., \( 1/3 \) or \( 3/4 \)) so that small blocks can be formed (e.g., if \( N_r/N = 1/3 \), then for every three patients it is ensured that one will receive the control treatment and two the new treatment). Let us use Pocock's example to illustrate the procedure.

Example 13.2

A trial is to be undertaken for advanced small cell lung cancer patients to compare the effectiveness of Cytoxan + CCNU with Cytoxan + CCNU + Procarbazine. The main endpoint is tumor response. It is proposed that 200 patients be entered on the trial. In a preceding trial, Cytoxan alone was compared to Cytoxan + CCNU, and the latter treatment was found superior. In this earlier trial, 88 patients were treated with Cytoxan + CCNU. The tumor response rate was \( \frac{39}{88} = 45\% \), with five nonevaluable cases. The question is how many of the 200 patients should be entered on the treatment Cytoxan + CCNU as randomized controls, for which there are already 83 patients in a historical group.

In this case, \( N_h = 83 \), \( N = 200 \), \( \bar{x}_h = 0.45 \), and from the binomial distribution, \( \sigma_r \) is estimated to be \( \sqrt{0.45 \times (1 - 0.45)} = 0.497 \%. \) It is assumed that \( \sigma_r = \sigma = \sigma_h = 49.7\% \). Suppose the standard deviation of the bias in the historical controls \( \sigma_b = 3\% \); applying (13.2), we obtain

\[
N_r = \frac{1}{2} \left( 200 - \frac{83}{1 + 83(3/49.7)^2} \right) = 68.1
\]

\( N_r \) is close to \( 1/3 N = 66.6 \) or \( N_r/N \sim 3 \), so that a convenient practical solution would be to randomize patients to Cytoxan + CCNU and Cytoxan + CCNU + Procarbazine in a ratio of 1 : 2.
In many clinical trials, more than one response variable may be used to evaluate treatment effect. In the trial described in Example 13.2 the survival time was also used for comparing the two treatments. In such cases, the sample size \( N \) for the randomized controls can be determined separately for each response variable and the eventual choice be a weighted mean of these values, with weights determined by the investigator according to the relative importance of the variables.

**Combination of Controls in Analysis**

When both randomized and historical controls are used, we have two estimators of \( \mu \), namely \( \bar{x} \) and \( \bar{\bar{x}} \). The problem is then how to determine a combined estimator. Based on the assumptions that \( \mu = \mu_h + \delta \) and that all of the variables are normal, the distribution of \( \mu \) is normal with mean

\[
\bar{x}_c = \frac{(\sigma^2_x/N_x + \sigma^2_h/N_h)\bar{x}_x + (\sigma^2_x/N_x + \sigma^2_h/N_h + \sigma^2_\delta)\bar{x}_h}{\sigma^2_x/N_x + \sigma^2_h/N_h + \sigma^2_\delta}
\]

and variance

\[
V_c = \left( \frac{N_x}{\sigma^2_x} + \frac{1}{\sigma^2_h/N_h + \sigma^2_\delta} \right)^{-1}
\]

Equation (13.3) can be written as

\[
\bar{x}_c = \bar{x}_x + \frac{W_0\bar{x}_h}{1 + W}
\]

where \( W_0 = (\sigma^2_x/N_x)/(\sigma^2_x/N_x + \sigma^2_h/N_h + \sigma^2_\delta) \). Thus, the combined estimate \( \bar{x}_c \) for \( \mu \) is a weighted sum of \( \bar{x}_x \) and \( \bar{x}_h \).

To test the hypothesis that the new treatment is as effective as the standard treatment, or \( \mu = \mu_x \), we use the assumption that the distribution of \( \mu_x \) is normal with mean \( \bar{x} \) and variance \( \sigma^2_x/N_x \). Hence the distribution of \( \mu_x - \mu \) is normal with mean \( \bar{x} - \bar{x}_x \) and variance \( \sigma^2_x/N_x + V_c \).

In practice, the variances \( \sigma^2_x, \sigma^2_h \) and \( \sigma^2_\delta \) are unknown and either the sample variances \( s^2_x, s^2_h \), and \( s^2_\delta \) are substituted or the pooled variance

\[
s^2_p = \frac{(N_x - 1)s^2_x + (N_h - 1)s^2_h + (N_h - 1)s^2_\delta}{N_x + N_h + N_h - 3}
\]

is used.

When sample observations follow the exponential distribution with mean \( \mu \), the maximum likelihood estimate \( \hat{\mu} \) is the sum of all survival times, uncensored and censored, divided by the number of deaths, as given in (8.15). Let \( \mu_x, \mu_h, \) and \( \mu \) be the true mean survival times \( n_x, n_h, \) and \( n \), the observed deaths, and \( \bar{x}_x, \bar{x}_h, \) and \( \bar{x} \), the observed mean survival times of the \( R, H, \) and \( T \) groups, respectively. The historical bias \( \delta = \log_e(\mu_x/\mu_h) \). To determine a combined estimator, assume the distribution of \( (\log_e \mu_x, \log_e \mu_h, \log_e \mu) \)

\[
\log_e \bar{x}_c = \frac{(\sigma^2_x/N_x + \sigma^2_h/N_h + \sigma^2_\delta)\bar{x}_x + (\sigma^2_x/N_x + \sigma^2_h/N_h + \sigma^2_\delta)\bar{x}_h}{\sigma^2_x/N_x + \sigma^2_h/N_h + \sigma^2_\delta}
\]

and variance

For the new treatment this estimator is unbiased and variance \( 1/n \). The calculation is performed on the basis of \( \log_e(\mu_x/\mu_h) \) having standard normal distribution.

**Example 13.3**

This is a clinical trial to test the effectiveness of an advanced melanoma one of which was also a treatment in a historical group is considered acceptable. The group is summarized in Table 13.6 according to \( \delta = \log_e(\mu_x/\mu_h) \) having standard normal distribution.

The following example involves the use of historical controls.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Survival Time (Days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous</td>
<td>80</td>
</tr>
<tr>
<td>Current</td>
<td>90</td>
</tr>
</tbody>
</table>

**Table 13.4 Summary Survival Times**

Source: Pocock (1976a).
determine a combined estimator for $\mu$, and $\mu^*_h$, $\bar{x}_c$ say, it can be shown that the distribution of $(\log_e \mu_*)$ is asymptotically normal with mean

$$\log_e \bar{x}_c = \frac{(1/n_h + \sigma^2_h) \log_e \bar{x} + (1/n_r) \log_e \bar{x}_h}{1/n_r + 1/n_h + \sigma^2_h}$$ (13.6)

and variance

$$\left(n_r + \frac{n_h}{1 + n_h \sigma^2_h}\right)^{-1}$$ (13.7)

For the new treatment the distribution of $\log_e \mu_*$ is normal with mean $\log_e \bar{x}_c$ and variance $1/n_r$. The difference between $\mu_*$ and $\mu_*$ can then be tested on the basis of $\log_e \mu_* - \log_e \mu$.

The following example, also from the Pocock, illustrates the procedure.

**Example 13.3**

This is a clinical trial to compare the survival experience of patients with advanced melanoma on two treatments, DTIC and TIC–Mustard. DTIC was also a treatment in a previous melanoma trial and this historical control group is considered acceptable. The survival experience for each treatment group is summarized in Table 13.4. Also, an exponential model seems reasonable according to survival plots of the data. The historical bias $\delta = \log_e (\mu_*/\mu_h)$ has been assigned a variance $\sigma^2 = 0.01$, which corresponds to $\mu_*/\mu_h$ having standard deviation $\sqrt{e^{0.01}(e^{0.01} - 1)} = 0.1$. That is, the potential historical bias in the mean survival times may be of the order of 10%.

Applying (13.6) and (13.7), the distribution of log mean survival for the randomized controls (DTIC) is approximately normal with mean $\log_e \bar{x} = 3.28$ and variance 0.0112. Thus, the combined point estimate of mean survival time on DTIC is $\bar{x} = 26.5$ weeks. For the new treatment the log mean survival has an approximate normal distribution with mean 3.30 and variance 0.0204. Thus, $\log_e (\mu_*/\mu_*)$, the log of the ratio of true mean survival times for DTIC and TIC–Mustard, has an approximately normal distribution with mean −0.02 (3.28 − 3.30) and standard deviation 0.18.

<table>
<thead>
<tr>
<th>Table 13.4 Summary Survival Data of Melanoma Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Previous trial</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Source: Pocock (1976a).
(\sqrt{0.0112} + 0.0204). This leads to a normal deviate for \(-0.02/0.18 = -0.1\), which is not significant. Thus, there is no evidence of a survival difference between the two treatments.

Although Pocock's procedure allows the use of both randomized controls and historical controls, it is not free of problems. The determination of the bias, \(\delta\) and \(\sigma^2\), in the historical data presents bias itself. To increase \(\sigma^2\), one places greater emphasis on the randomized controls and the historical data become of less value, and vice versa. In addition, the determination of \(\sigma^2\) is highly subjective.

Peto et al. (1976) suggest a much simpler allocation, that is, at least one-third of current patients should be randomized controls unless the new treatment is so superior to the standard that randomization is felt to be unethical. That is, one-third of the current patients are randomized controls and two-thirds are still available for comparison with whatever historical controls are chosen. For people who only believe randomized controls, this is as substantial a randomized trial as if the ordinary 1:1 randomization has been adopted. Other methods for combining historical and randomized controls include Tarone (1982) and Dempster et al. (1983).

**BIBLIOGRAPHICAL REMARKS**

In addition to the books cited in Chapter 12 and the medical and statistical journals that publish papers on practical and theoretical issues in clinical trials, a journal devoted to the subject, *Controlled Clinical Trials*, has been published since 1980. Readers who are interested in conducting clinical trials are referred to this journal for the most recent developments in this field.

**EXERCISES**

13.1 Use the random-number table in Appendix C (Table C-20) to assign two treatments A and B to 30 patients. Do you get equal numbers of patients in the two treatment groups by simple randomization? If not, use restricted randomization.

13.2 Suppose you are the coordinator of a study involving five hospitals. The purpose of the study is to compare the efficiency of treatments A and B. Set up a randomization scheme for 50 patients such that the treatment assignment over all hospitals and within a hospital will be balanced.

13.3 Use the random-number table (Table C-20) to assign three treatments A, B, and C to 45 patients in a way that the number of patients in each group is approximately equal.
13.4 Suppose a trial is to be conducted for advanced melanoma patients to compare chemotherapy only with chemotherapy plus immunotherapy. The main endpoint is response. In a previous trial, 60 patients were treated with the same chemotherapy alone. The response rate was 40%. For the present trial, it is determined to enter 120 patients. According to Pocock's formulas how many of these 120 patients should be entered on the randomized control for which there are already 60 patients in a historical control group? Let the standard deviation of the bias in the historical controls be (a) 1% and (b) 5%. 

...