1
Introduction

Key Points

- The purpose of data monitoring committees (DMCs) is to protect the safety of trial participants, the credibility of the study and the validity of study results.
- DMCs have a long history in trials sponsored by government agencies in the USA and Europe.
- Pharmaceutical companies are increasing their use of DMCs in trials of investigational drugs, biologics and medical devices.
- Statistical methods have been developed for interim monitoring of clinical trials.
- While not all trials need DMCs, trials that address major health outcomes and are designed to definitively address efficacy and safety issues should incorporate DMC oversight.

1.1 MOTIVATION

In randomized clinical trials designed to assess the efficacy and safety of medical interventions, evolving data are typically reviewed on a periodic basis during the conduct of the study. These interim reviews are especially important in trials conducted in the setting of diseases that are life-threatening or result in irreversible major morbidity. Such reviews have many purposes. They may identify unacceptably slow rates of accrual or high rates of ineligibility determined after randomization, protocol violations that suggest that clarification of or changes to the study protocol are needed, or unexpectedly high dropout rates that threaten the trial's ability to produce credible results. The most important purpose, however, is to ensure that the trial remains appropriate and safe for the individuals who have been or are still to be enrolled. Unacceptable levels of treatment toxicity may require adjustment of dosage or schedule of administration, or even abandonment of the study. Efficacy results, too, must be monitored to enable benefit-to-risk assessments to be made. Interim results may demonstrate

that one intervention group has such unfavorable outcomes with regard to survival or a major morbidity endpoint that its benefit-to-risk profile is clearly inferior to that of the comparator treatment. In such cases, it may be appropriate to terminate the inferior intervention or the entire trial early so that current study participants, as well as future patients, will no longer be provided the inferior treatment.

Relatively early in the development of modern clinical trial methodology, some investigators recognized that, despite the compelling ethical need to monitor the accumulating results, repeated review of interim data raised some problems. Repeated statistical testing was seen to increase the chance of a 'false positive' result unless nominal significance levels were somehow adjusted. In addition, it was recognized that awareness of the pattern of accumulating data on the part of investigators, sponsors or trial participants could affect the course of the trial and the validity of the results. For example, if investigators were aware that the interim trial results were favoring one of the treatment groups, they might be reluctant to continue to encourage adherence to all regimens in the trial, or to continue to enter patients on the trial, or they might limit the types of patients they would consider entering. Furthermore, influenced by financial or scientific conflicts of interest, investigators or the sponsor might take actions that could diminish the integrity or credibility of the trial. For example, a sponsor observing interim data showing that the new treatment had little if any effect on the prespecified primary endpoint but a much stronger effect on an important secondary endpoint might be tempted to switch the designation of these two endpoints.

A natural – and practical – approach to dealing with these problems is to assign sole responsibility for interim monitoring of data on safety and efficacy to a committee whose members have no involvement in the trial, no vested interest in the trial results, and sufficient understanding of trial design, conduct and data-analytical issues to interpret interim analyses with appropriate caution. These 'data monitoring committees' (DMCs) have become critical components of many clinical trials. The interim monitoring experience of an early AIDS clinical trial illustrates some of the inherent difficulties and challenges that are faced in reviewing the accumulating data from clinical trials.

Example 1.1: Treatment for HIV infection

Trial 002 of the Community Programs for Clinical Research in AIDS (CPCRA) was designed to compare the efficacy of two antiretroviral agents, zalcitabine (ddC) and didanosine (ddI), in HIV-infected patients who did not derive benefit from zidovudine (AZT), at that time the first-line treatment for HIV infection (Abrams *et al.*, 1994). When the trial was initiated, ddI was considered the first-line treatment in this patient population; the goal of the trial was to determine whether ddC was approximately equivalent to ddI by seeing whether as much as a 25% advantage for ddI in time to disease progression or death could be ruled out. A total of 467 patients were randomized to receive either ddI or ddC. To achieve

the desired level of statistical power, it was calculated that patient follow-up would be needed until 243 patients had been observed to reach the endpoint of disease progression or death.
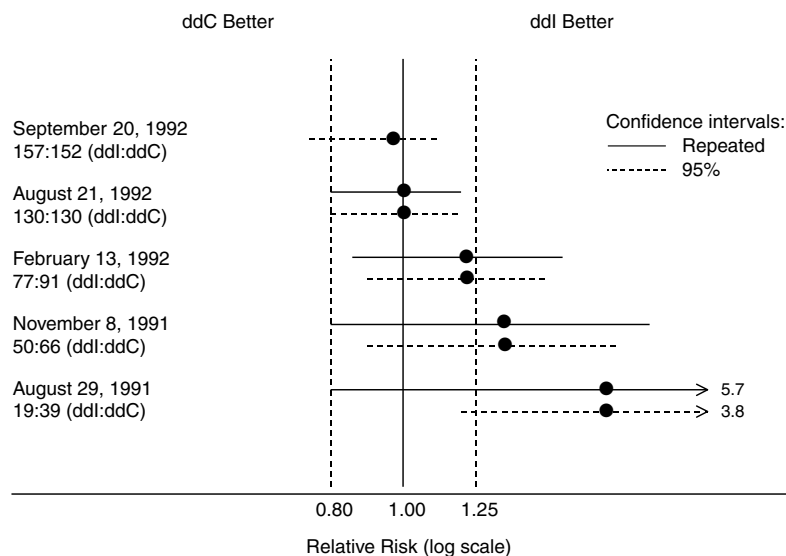
This trial was initiated in December 1990, at a time when little in the way of effective treatments for this population was available, when the numbers of new HIV infections and deaths were increasing, and when both the patient community and their physicians were increasingly desperate to identify treatments that could buy a little more time for those suffering from this disease. Patients entering such trials were generally young men who were facing a very premature death from a disease they may not have even known about at the time they contracted it. Further, more pharmaceutical companies were initiating drug development for treatment of HIV, but with a great deal of caution, as would be expected in a completely new disease area. While there are inherent tensions in all trials testing new agents for serious diseases, the atmosphere surrounding early trials of AIDS treatments, such as this one, was particularly 'high pressure'. Trial 002 was monitored by the DMC that had been established by the National Institute of Allergy and Infectious Diseases (NIAID) to oversee all of its extramural trials of treatment for HIV infection (DeMets *et al.*, 1995). The CPCRA was a clinical trials group funded by NIAID; therefore, access to interim data was limited to DMC members – none of whom were treating patients on this or any other NIAID-funded AIDS trial, or had any financial stake in the trial outcome – and to a limited number of NIAID staff.

The interim results from this trial, shown in Figures 1.1 and 1.2, illustrate how substantially relative risk estimates can change over time. At the first interim analysis in August 1991, the early trial results strongly favored ddI. At that time, the ddI group had experienced many fewer disease progressions (19 vs. 39) and fewer deaths (6 vs. 12) than the ddC group. The effects on laboratory markers were also more favorable in the ddI group. While the nominal *p*-value for the treatment difference in progressions at this analysis was an impressive 0.009, this value did not approach the protocol-specified early termination criterion at this early stage in the trial. The DMC considered these data as well as available information on toxicities and other relevant outcomes and recommended that the trial continue as designed.
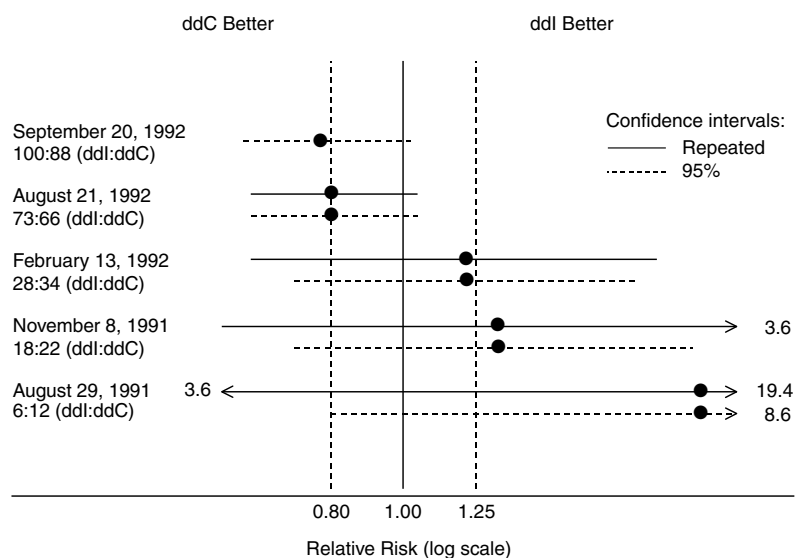
As the figures show, the differences favoring ddI steadily disappeared over successive meetings of the DMC. At the final review, in August 1992, the DMC recommended that the study end as originally planned since the required number of events had been observed. The results at the end of the trial had shifted from strongly favoring ddI to showing a small advantage for ddC in this population. These data did provide strong statistical evidence that ddC was not inferior to ddI in the sense noted earlier.

Had the results from the initial interim analysis of the CPCRA 002 trial been broadly disseminated, it is most unlikely that the trial would have continued, given the urgent desire to identify optimal therapeutic approaches and

ddC Better   ddI Better

September 20, 1992
157:152 (ddI:ddC)

August 21, 1992
130:130 (ddI:ddC)

February 13, 1992
77:91 (ddI:ddC)

November 8, 1991
50:66 (ddI:ddC)

August 29, 1991
19:39 (ddI:ddC)

Confidence intervals:
——— Repeated
- - - 95%

5.7
3.8

0.80   1.00   1.25

Relative Risk (log scale)

Figure 1.1   Relative risk of progression of disease (including death) by date of DMC review. Numbers to the right of the arrows are upper confidence limits. From Fleming *et al.*, Insights from monitoring the CPCRA ddI/ddC trial (1995), *Journal of Acquired Deficiency Syndromes and Human Retrovirology* **10** (Suppl. 2) Reproduced by permission of Lippincott, Williams & Wilkins.

ddC Better   ddI Better

September 20, 1992
100:88 (ddI:ddC)

August 21, 1992
73:66 (ddI:ddC)

February 13, 1992
28:34 (ddI:ddC)

November 8, 1991
18:22 (ddI:ddC)

August 29, 1991
6:12 (ddI:ddC)

Confidence intervals:
——— Repeated
- - - 95%

3.6

3.6
19.4
8.6

0.80   1.00   1.25

Relative Risk (log scale)

Figure 1.2   Relative risk of death by date of DMC review. Numbers to the right of the arrows are upper confidence limits. From Fleming *et al.*, Insights from monitoring the CPCRA ddI/ddC trial (1995), *Journal of Acquired Deficiency Syndromes and Human Retrovirology* **10** (Suppl. 2) Reproduced by permission of Lippincott, Williams & Wilkins.

the emerging positive data from other trials about the efficacy of ddI. Even without broad dissemination, if the data had been available to trial investigators and/or the participating pharmaceutical companies, it might have been difficult or impossible to continue the trial, given the intense pressures of the time. The investigators might have been unwilling to continue treating patients with an apparently inferior therapy; the pharmaceutical company whose product appeared superior might have chosen to end its participation and submit the available data to the Food and Drug Administration (FDA). Stopping the study early on, with a conclusion of an apparently large benefit of ddI, would clearly have been unfortunate; it would have misled patients regarding the relative efficacy of these two agents, and it would have precluded the obtaining of additional information that would ultimately contribute to the optimal continuing development of both agents as components of AIDS treatment programs.

1.2   HISTORY OF DATA MONITORING COMMITTEES IN GOVERNMENT-SPONSORED TRIALS

The concept of DMCs arose soon after the era of the modern randomized clinical trial began in the 1950s. Perhaps the first step in formalizing the concept of committees who would be charged with regular assessment of a trial's accumulating results was taken by the US National Institutes of Health (NIH). In the mid-1960s, the NIH was beginning to sponsor large, multicenter trials of new treatment interventions for serious diseases. At this time, a task force under the leadership of Dr. Bernard Greenberg of the University of North Carolina was constituted by the then National Heart Institute to develop an advisory document concerning the organization and conduct of such trials. This report, issued in 1967 (but not formally published until 1988), included among its recommendations the need for an advisory group of experts not directly involved in the conduct of the trial to review the study protocol and advise the Institute about the conduct of the trial (Heart Special Project Committee, 1988). In addition, the report addressed the need for a mechanism for terminating a trial early if it became evident that it could not meet its objectives or new information rendered it superfluous.

The influence of the 'Greenberg Report', as it came to be called, can be seen in an early trial sponsored by the NIH, the Coronary Drug Project (CDP); (CDP Research Group, 1973). This trial was initiated in the mid-1960s, and had an external committee charged with reviewing the trial conduct and the interim results on an ongoing basis. The experience in this trial reflected both the complexity of the data monitoring process and the value of an independent committee, and stimulated methodological development of new monitoring approaches.

Example 1.2: The Coronary Drug Project

The CDP was among the first, possibly the first, multicenter trial to be based on the operational model put forward by the Greenberg Report. The CDP was a multicenter, multiarm placebo-controlled trial designed to evaluate the effectiveness of five lipid-lowering treatments in patients who had experienced a cardiovascular event. More than 8000 patients were enrolled, with a planned minimum follow-up time of 5 years. A Policy Advisory Board (PAB) was initially established to review overall progress and conduct of the trial. Later, but still early in the trial, a subgroup of the PAB was formed to monitor efficacy and patient safety. This committee, which would today be called a DMC, would make recommendations to both the independent PAB as well as to the National Heart Institute. During the course of the trial, the committee recommended early termination of three of the five active treatment arms (high- and low-dose estrogen and dextrothyroxine).

The CDP Research Group (1981) published details of the consideration of interim results and the resultant decision-making. Two important themes emerged from their description of the process. First, early data trends can be very unstable, with much waxing and waning of risk ratios due to the small number of events at the early stages of the study. Thus, great caution is needed when interpreting results from early analyses. The CDP applied statistical procedures to take account of the repeated testing problem noted earlier, and they appear to be the first to describe such use in the context of a real clinical trial. These analytical approaches helped them resist the emotional pull of the early results.

The second theme was the complexity of the decision-making process, requiring multiple factors (many of which may not be readily quantifiable) to be taken into account. They noted: 'Although a number of rather sophisticated statistical tools are available in the decision making process, these are at best red flags that warn of possible treatment problems and can never be used by themselves as hard and fast decision rules' (CDP Research Group, 1981). The types of factors that need to be considered are listed in Table 1.1 and have been addressed by many DMCs since then. The continued development of statistical techniques for monitoring, far beyond what was available at the time of the CDP (see Chapter 8), has not altered the fact that statistical assessment is but one part of a highly complex decision process.

The value of DMCs to the clinical trials process was evident in the CDP, and such committees came to be a standard component of large multicenter trials sponsored by federal agencies such as the NIH and the Veterans Administration (VA). Soon after the CDP began, the National Heart Institute implemented several other trials, all with the same basic clinical trial organizational structure as the CDP, including use of a DMC. In 1968, the Urokinase Pulmonary Embolism Trial (UPET) was initiated to test the effectiveness of a thrombolytic therapy, urokinase, in resolving blood clots in the lung (UPET Study Group, 1970). This trial was followed immediately by the Urokinase-Streptokinase Pulmonary Embolism Trial

Table 1.1   Relevant factors in interim decision-making

1. Recruitment rate and completion schedule
2. Baseline characteristics and risk profile of participants
3. Baseline comparability across treatment arms
4. Compliance to intervention
5. Data completeness and follow-up
6. Internal consistency
   (a) Primary and secondary outcomes
   (b) Subgroups
   (c) Safety profile
7. External consistency
8. Statistical issues for interim analyses
9. Ethical issues
10. Impact of early termination

(USPET); see USPET Study Group (1974). Each of these trials used a DMC. In the area of heart disease, the Hypertension Detection and Follow-up Program (HDFP) was initiated in 1972 to evaluate the impact of blood pressure reduction on 5-year mortality in individuals with mild hypertension (HDFP Cooperative Group, 1979). The Coronary Artery Surgery Study (CASS) began in 1973 to assess the effect of coronary artery bypass graft surgery compared to best medical treatment, again with 5-year mortality as the primary outcome of interest (CASS Principal Investigators, 1983). Trials in lung diseases also followed a similar model. In 1973, the Extracorporeal Membrane Oxygenator (ECMO) trial was implemented to compare a mechanical blood oxygenation device to best standard care in patients who had suffered severe lung trauma (Zapol *et al.*, 1979). The Nocturnal Oxygen Therapy Trial (1980), the Intermittent Positive Pressure Breathing Trial (1983) and the Respiratory Distress Syndrome Trial (Collaborative Group on Antenatal Steroid Therapy, 1981) were all begun in 1975. All of these early lung trials included a DMC in their organizational structures. Thus, by the mid-1970s, the Institute (by then renamed the National Heart, Lung and Blood Institute (NHLBI)) was routinely establishing DMCs to monitor randomized clinical trials in all clinical areas.

In 1972, two senior statisticians left the NHLBI to establish a biometrics research group in the newly formed National Eye Institute (NEI). They brought with them the knowledge and experience of NHLBI's emerging clinical trials programs, including the role of the DMC. This influence can be seen in the Diabetic Retinopathy Study (DRS), one of the first NEI randomized trials, which evaluated a new photocoagulation treatment in diabetic patients experiencing proliferative retinopathy (DRS Research Group, 1976) by randomly assigning one eye of each

study participant to the new treatment, and the other eye to standard management. The DMC for this study made a significant protocol change early in the trial, based on an unexpected early large benefit of the photocoagulation treatment. Rather than recommend early termination, the DMC recommended that each 'control eye' should receive the photocoagulation treatment when it reached a specified level of retinopathy. This change permitted the evaluation of safety and duration of benefit based on longer follow-up. The NEI has routinely incorporated DMCs into the structure of the major Phase III comparative trials it sponsors.

In the mid-1970s, the VA first developed guidelines for their Cooperative Group network for conducting clinical trials for VA patients. These guidelines, which are regularly revised and updated, include the use of DMCs (Cooperative Studies Program, 2001).

Cancer trials sponsored by the National Cancer Institute (NCI) began to use DMCs in the early 1980s. The first group to establish DMCs was the North Central Cancer Treatment Group (NCCTG), headquartered at the Mayo Clinic. This group developed a DMC model that, instead of outside experts, involved study investigators and a statistician from the group's coordinating center. Although this group was clearly not independent of the trials, it established the key approach of not sharing interim data widely among all investigators, as was the common practice in cancer trials at that time; access to the interim data was limited to the DMC. Soon afterwards (and facilitated by a move of two NCCTG statisticians involved in that DMC to the statistical center for the Southwestern Oncology Group (SWOG)), the concept of the DMC was introduced to SWOG and adopted as part of SWOG operating procedures. Shortly thereafter an 'intergroup' study of adjuvant therapy of colon cancer (#0035) was initiated with clinical leadership from the NCCTG and statistical leadership from SWOG. The trial incorporated the DMC model developed by the NCCTG, and provided the first opportunity for the other cooperative groups to experience this approach. Other cancer cooperative groups established DMCs following issuance of data monitoring policies by the NCI in 1994 (Smith *et al.*, 1997). NCI-sponsored cancer prevention trials such as the Alpha-tocopherol, Beta-Carotene (ATBC) lung cancer prevention trial in Finland (ATBC Cancer Prevention Study Group, 1994) and the Beta-Carotene and Retinol Efficacy Trial (CARET) study (Omenn *et al.*, 1996) in the United States also had formal DMCs. It is interesting to note that despite the fact that cancer treatment trials (unlike most trials in cardiovascular disease) are generally unblinded because of the complex nature of administration of chemotherapy and the distinctive toxicities associated with different agents, DMCs have been found to be a valuable component of clinical cancer research.

The AIDS epidemic that emerged in the early 1980s led NIAID to form two clinical trial networks, the AIDS Clinical Trial Group (ACTG) and the CPCRA. These two NIAID-sponsored clinical trials groups were served by a single DMC that had to develop new operational approaches to deal with the many new challenges posed by HIV/AIDS trials (DeMets *et al.*, 1995; Ellenberg *et al.*, 1993b). These challenges included having to monitor multiple trials from two different organizational

groups, an unprecedented involvement of patient representatives and advocacy groups in the design of trials and interpretation of results, and scientific and political pressures to identify effective treatments as quickly as possible. In addition, the pharmaceutical industry was much more closely involved with these trials than had been traditional with NIH clinical trials programs, requiring careful attention to lines of communication with industry sponsors.

Clinical trials in Europe have also used DMCs for trial oversight. The International Studies of Infarct Survival (ISIS), initiated in the 1970s, have always incorporated DMCs in their trial structures. The Medical Research Council in the UK instituted a policy of establishing independent DMCs for 'high-profile' trials in the early 1990s (Parmar and Machin, 1993). In other countries, as in the USA, DMCs were first seen primarily in cardiovascular trials (19).

1.3   DATA MONITORING COMMITTEES IN TRIALS SPONSORED BY THE PHARMACEUTICAL INDUSTRY

While DMCs were widely used in government-sponsored trials, DMCs were only occasionally established for trials sponsored by the pharmaceutical industry until the early 1990s. Of the few industry-sponsored trials that did establish formal DMCs, many were in the area of cardiovascular disease with improvement of survival as the primary goal (Anturane Reinfarction Trial Research Group, 1980; Persantine-Aspirin Reinfarction Study Research Group, 1980; Swedberg *et al.*, 1992; APSAC Intervention Mortality Study Trial Group, 1988; European Myocardial Infarction Project, 1988), following the model established by the NHLBI. The move to greater use of DMCs in industry trials may have been influenced at least in part by three factors. First, concerns emerging in the late 1980s and early 1990s about the reliability of surrogate endpoints led to increased numbers of industry trials designed to directly assess effects on clinical endpoints such as mortality. Second, the increased collaborative efforts of industry and NIH in areas such as cardiovascular and AIDS research exposed pharmaceutical companies to clinical trial models that were new to them, particularly with regard to the assignment of responsibility for interim monitoring to an independent committee. These activities brought them into increased contact with researchers whose experience with DMCs led them to strongly advocate their use in industry trials. Third, although DMCs are not generally required for clinical trials performed by regulated industry (see Chapter 10), FDA staff increasingly recommend that companies establish DMCs for certain types of trials.

While many companies had concerns about giving up access to the accumulating data as the trial progressed, the establishment of DMCs offered companies some clear advantages. For example, regulatory bodies traditionally have been uncomfortable about companies making changes to trials in progress, recognizing the potential biases that can arise in making such decisions because of the large financial stake a company has in the outcome of its trials. Even when the rationale

for such changes appears well founded, one cannot ignore the concern that a company might identify and recommend only those changes that are likely to be advantageous to the company. When changes are proposed by sponsors who do not have access to interim trial data, concerns about bias are substantially diminished. In addition, independent DMCs may protect companies against claims of misleading stockholders, as in the next example.

Example 1.3: Treatment of amyotrophic lateral sclerosis

A multicenter randomized double-blind placebo-controlled trial, the ALS CNTF Treatment Study (ACTS), evaluated a new ciliary neurotrophic factor (CNTF) in patients with amyotrophic lateral sclerosis (ALS) or Lou Gehrig's disease. In ALS, the patient's muscle strength, including respiratory function, rapidly deteriorates. Survival time from onset of disease is typically only a few years. The new nerve growth factor was believed to increase muscle strength or at least prevent further deterioration. ACTS was sponsored by a small biotech company that established an independent steering committee, an independent DMC and an independent statistical center for interim analyses. Thus, the sponsor was totally blinded during the conduct of the trial. The DMC terminated the trial early due to adverse effects, observing that measures of muscle strength were worse on the new therapy than on the placebo. Results of the trial have been published (ALS CNTF Treatment Study Group, 1996). Following the decision to terminate, the DMC immediately briefed the steering committee and the sponsor. The sponsor, within a day, alerted the financial community. Later, investors who had had great expectations for this new therapy brought legal action against the sponsor, arguing that the sponsor had misled them by not alerting them earlier about the impending negative results (Wall Street Journal, 1994). Since the sponsor was kept blinded to accumulating results during the course of the trial, they did not know the results until the day before the results were made public. The use of an independent DMC provided the sponsor with a strong defense against such claims of illegal activity.

As the use of DMCs in clinical trials increased, the existence of widely varying policies and practices for such committees by trial sponsors became evident. At an international conference held at the NIH in 1992, many differing views on the optimal approaches to data monitoring were presented (Ellenberg *et al.*, 1993a). In recent years, papers describing and/or advocating specific monitoring practices have appeared in greater numbers (Armitage, 1999a, 1999b; Armstrong and Furberg, 1995; Canner, 1983; DeMets *et al.*, 1982, 1984, 1995, 1999; Dixon and Lagakos, 2000; Fleming and DeMets, 1993; Freidlin *et al.*, 1999; Meinert, 1998a, 1998b; Pocock, 1993; Whitehead, 1999). Another recent development has been mention of DMCs for the first time in regulations and guidance documents of regulatory authorities such as the FDA (Code of Federal Regulations, Title 21, Part 50.24; US Food and Drug Administration, 1997, 1998, 2001).

1.4 STATISTICAL METHODS FOR INTERIM MONITORING

The practical experiences and challenges faced by DMC members in the 1970s led to the development of statistical methods that accounted for the multiplicity problem generated by repeated conduct of interim analyses at scheduled DMC meetings. These approaches, known as 'group sequential methods' to distinguish them from earlier approaches based on assumptions of continual interim analysis (Pocock, 1977; O'Brien and Fleming, 1979; Lan and DeMets, 1983), were rapidly adopted during the 1980s and provided a new structure to the data monitoring process.

The fundamental statistical problem is fairly simple. Under the null hypothesis in an intrinsically one-sided superiority or non-inferiority setting, one wishes to maintain an upper bound on the false positive error rate, at a nominal level that usually is 2.5%. (The standard for strength of evidence, corresponding to allowing a 2.5% false positive error rate, is achieved whether one is conducting a two-sided 0.05-level test or a one-sided 0.025-level test.) However, the performance of multiple tests of the null hypothesis over time will lead to a false positive error rate substantially higher than the nominal level at which the tests are performed. Thus, if we test our data frequently during the course of a trial, the probability that we will, at some point, observe a difference that is 'statistically significant' at the one-sided 0.025 level is in fact substantially greater than 2.5%. To preserve the desired level of false positive error, it is necessary to perform interim testing at more conservative levels.

Several approaches to this problem can be taken. Perhaps most simply, one can perform all interim tests at highly conservative levels (e.g., require a *p*-value of 0.001 or less to justify early termination with a conclusion of strong evidence against the null hypothesis), so that the impact on overall false positive error is minimal and final testing can be done at the conventional level without much worry about having inflated the false positive rate. This approach, first proposed by Haybittle (1971), is very conservative, even when the trial nears the time of completion. Another straightforward approach, described by Pocock (1977), uses the same significance level for all interim tests as well as the final test, with this level calculated to provide the desired overall false positive error. In order to calculate the testing levels to be used, the number of interim tests must be specified. At interim analyses, this approach is much less conservative than the first approach; unless a huge number of interim tests is specified, the required significance level for all tests will be substantially less stringent than 0.001. One difficulty with the Pocock approach is that, unlike the Haybittle approach, it requires the final test be performed at a lower than conventional level. For example, if four interim analyses are planned and an overall 0.025 false positive error rate is desired, the final (fifth) analysis will need to be done at a one-sided significance level of approximately 0.008. This allows for the uncomfortable situation of observing a final difference that produces a one-sided *p*-value of 0.01 but being unable to reject the null hypothesis at the one-sided 0.025 level.

To achieve the intuitively appealing property of the Haybittle approach (permitting final testing at nearly the conventional significance level) without its extreme conservatism in the latter stages of trial monitoring, O'Brien and Fleming (1979) developed an alternative to Pocock's approach that varied the significance level used for interim testing as the trial progressed. Their method provides for highly conservative criteria early in the trial, with progressively less stringent criteria at successive interim analyses, and a final analysis that can be performed at close to the nominal level.

The O'Brien–Fleming group sequential boundary is one member of a family of boundaries with similar characteristics that can be generated (Wang and Tsiatis, 1987). In this book, most of the examples presented use this statistical approach. The O'Brien–Fleming type boundary has become popular, probably because its properties reflect the thinking of many of those with experience in evaluating interim trial results. First, this boundary is very conservative early in the trial when the numbers of patients and events are small and any estimate of treatment effect is therefore unreliable, requiring great caution in interpretation. Second, as more patients are recruited and more events are observed, the information fraction increases and the O'Brien–Fleming criteria for statistical significance become correspondingly less stringent. A third reason is that at the completion of the trial (assuming termination did not occur earlier) the critical value for the test statistic is close to the nominal value (e.g., 0.05, two-sided), the same critical value that would be used for a trial with no sequential testing. Consequently, the power of the trial is maintained without having to increase the sample size. This is important because it permits the conduct of interim analyses while maintaining the false positive error rate at accepted levels without having to substantially increase trial size and cost. These and other methods will be discussed in more detail in Chapter 8.

1.5   WHEN ARE DATA MONITORING COMMITTEES NEEDED?

As noted at the beginning of this chapter, DMCs are most relevant to randomized clinical trials specifically focused on clinical efficacy and safety. They have been used primarily for trials that are expected to provide a definitive answer to a question about whether a drug is effective, or whether one drug regimen is more effective than another. Further, even in this setting they have been used mostly in trials that address major health outcomes such as mortality, progression of a serious disease, or occurrence of a life-threatening event such as heart attack and stroke. They have not been used as widely in the many randomized trials (nearly always short-term) that address symptom relief, nor in trials implemented early in drug development whose results will be examined in an exploratory fashion and whose successor trials will be looked to for definitive conclusions.

Although the monitoring of trials with regard to safeguarding the interests of study participants and to ensuring that the trial is being conducted properly is appropriate and necessary in every clinical study, a formal DMC is not routinely needed. We propose several general criteria that can help determine the need for and value of a DMC in a given trial:

1. Is the trial intended to provide definitive information about effectiveness and/or safety of a medical intervention?

2. Are there prior data to suggest that the intervention being studied has the potential to induce potentially unacceptable toxicity?

3. Is the trial evaluating mortality or another major endpoint, such that inferiority of one treatment arm has safety as well as effectiveness implications?

4. Would it be ethically important for the trial to stop early if the primary question addressed has been definitively answered, even if secondary questions or complete safety information were not yet fully addressed?

A DMC usually should be implemented if two or more of these criteria are met, and usually would not be considered if none are met. In some cases, when the treatments are novel and raise serious safety questions, DMCs are used even in early, non-randomized studies – such studies would meet only criterion 2. Considerations for use of DMCs will be further addressed in Chapter 9.

1.6   WHERE WE ARE TODAY

At present, DMCs appear to be a 'growth industry'. The incorporation of such committees into trial structures is increasing, not only in industry trials, but also in government-sponsored trials in disease areas (such as cancer) that did not initially use DMCs. With this rapid growth, a variety of approaches to DMC operations have been developed for trials in multiple medical areas. Although there is broad agreement on many principles and procedures relating to DMC functions, there are important aspects for which consensus currently does not exist. Further, as with any area where ethical issues arise, opinions tend to be strongly held among knowledgeable and experienced individuals regarding the optimal approach to interim monitoring and the operation of DMCs.

In this book, we will present principles and guidelines for constituting and implementing DMCs and for establishing the policies and procedures under which they operate. We will describe the variety of approaches that have been taken to address the operational aspects of the data monitoring of clinical trials, and will give special attention to some of the controversies that have arisen regarding optimal practices. While attempting to lay out the advantages and disadvantages of the various approaches, we will not hesitate to recommend, based on our own experience in a wide range of medical settings, what we believe to be the best way to proceed.

1.7   FUNDAMENTAL PRINCIPLES OF DATA MONITORING

We conclude this Introduction with a brief discussion of some fundamental principles that will be invoked throughout the book, and that will be addressed in more detail in later chapters.

Principle 1. The primary responsibilities of a DMC are to: (i) safeguard the interests of study patients; (ii) to preserve the integrity and credibility of the trial in order that future patients may be treated optimally; and (iii) to ensure that definitive and reliable results be available in a timely way to the medical community.

The DMC has responsibilities to the trial investigators and sponsor, and to the scientific community generally, to ensure the integrity and reliability of the scientific result that is obtained. As discussed more fully in Chapter 2, however, its primary responsibility is to ensure the safety of the study participants. It must provide assurance to patients and their treating physicians that patients' care will not be compromised because of participation in the study.

Principle 2. The DMC should have multidisciplinary representation, including physicians from relevant medical specialties and biostatisticians. In many cases, other experts such as bioethicists, epidemiologists and basic scientists should also be included.

Due to the complexity of clinical trials and the decision-making process (discussed in detail in Chapter 3), the DMC requires sufficiently broad membership to ensure that all relevant medical, ethical, safety and scientific issues can be adequately discussed and properly weighed in all recommendations concerning trial conduct and termination.

Principle 3. The DMC should have membership limited to individuals free of apparent significant conflicts of interest, whether they are financial, intellectual, professional or regulatory in nature.

There is an intrinsic need for judgment in the process of developing recommendations about important study conduct issues, including whether to terminate or continue a trial. Such recommendations must be perceived to have been made in a fair and unbiased manner. Study integrity and credibility are compromised if individuals with apparent conflicts of interest influence recommendations. This could occur, for example, if members of a DMC were in a position to realize financial or professional gain should the study produce a positive result. Because of this concern, discussed more fully in Chapter 4, sponsors or others having significant financial or professional interests that could be affected by the outcome of the trial should generally not be members of the DMC for that trial. The word 'significant' is key; it may be necessary in some cases to include individuals with some potential conflict of interest if there are no other individuals with the requisite

expertise available to serve on the DMC. In such cases, disclosure of the potential conflict – to other committee members, the sponsor, the regulatory agency, etc. – is necessary. The disclosure process permits independent assessments about whether the apparent conflict might have had a significant impact.

Principle 4. The DMC members should ideally be the only individuals to whom the data analysis center provides results on relative efficacy and safety of study treatments.

As we saw in the CPCRA 002 example, it is common for early results to be misleading by giving the inaccurate impression that treatment effects are markedly favorable or unfavorable. Thus, as will be discussed in detail in Chapter 5, widespread reporting of interim results greatly increases the risk of actions being taken based on unreliable information. Such actions could include inappropriate early abandonment of the trial, or even of other related trials. In the setting of oncology trials, Green *et al.* (1987) demonstrated that providing a DMC sole access to the interim data reduced the risk that trials would experience declining accrual rates over time, inappropriate early termination yielding equivocal results, or final results that were inconsistent with prematurely published early results (see also Armitage, 1999b). If it appears necessary in specific cases to make interim efficacy and safety results available to individuals outside the DMC, those individuals should agree to maintain the confidentiality of the information.

REFERENCES

Abrams D, Goldman A, Launer C *et al.* (1994) A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. *New England Journal of Medicine* **330**: 657–662.

Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study Group (1994) The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *New England Journal of Medicine* **330**: 1029–1035.

ALS CNTF Treatment Study Group (1996) A double-blind placebo-controlled clinical trial of subcutaneous recombinant human ciliary neurotrophic factor (rHCNTF) in amyotrophic lateral sclerosis. *Neurology* **46**: 1244–1249.

Anturane Reinfarction Trial Research Group (1980) Sulfinpyrazone in the prevention of sudden death after myocardial infarction. *New England Journal of Medicine* **302**: 250–256.

APSAC Intervention Mortality Study Trial Group (1988) Effect of intravenous APSAC on mortality after acute myocardial infarction: preliminary report of a placebo-controlled clinical trial. *Lancet* **1**: 545–549.

Armitage P on behalf of the Concorde and Alpha Data and Safety Monitoring Committee (1999a) Data and safety monitoring in the Concorde and Alpha Trials. *Controlled Clinical Trials* **20**: 207–228.

Armitage P on behalf of the Delta Data and Safety Monitoring Committee (1999b) Data and safety monitoring in the Delta Trial. *Controlled Clinical Trials* **20**: 229–241.

Armstrong PW, Furberg CD (1995) Clinical trials and safety monitoring boards: the search for a constitution. *Circulation* **91**: 901–904.

Buyse M (1993) Interim analyses, stopping rules and data monitoring in clinical trials in Europe. *Statistics in Medicine* **12**: 509–520.

Canner PL (1983) Monitoring of the data for evidence of adverse or beneficial treatment effects. *Controlled Clinical Trials* **4**: 467–483.

CASS Principal Investigators & Their Associates (1983) Coronary Artery Surgery Study (CASS): A randomized trial of coronary artery bypass surgery. *Circulation* **68**(5): 939–950.

Collaborative Group on Antenatal Steroid Therapy (1981) Effect of antenatal dexamethasone administration on the prevention of respiratory distress syndrome. *American Journal of Obstetrics and Gynecology* **141**: 276–287.

Cooperative Studies Program (2001) *Guidelines for the Planning and Conduct of Cooperative Studies.* Office of Research and Development, Department of Veterans Affairs. http://www.va.org/resdev.

Coronary Drug Project Research Group (1973) The Coronary Drug Project: design, methods, and baseline results. *Circulation* **47** (Suppl. 1), xlvii–xlviii.

Coronary Drug Project Research Group (1981) Practical aspects of decision-making in clinical trials: the Coronary Drug Project as a case study. *Controlled Clinical Trials* **1**: 363–376.

DeMets DL, Williams GW, Brown Jr. BW and the NOTT Research Group (1982) A case report of data monitoring experience: the Nocturnal Oxygen Therapy Trial. *Controlled Clinical Trials* **3**: 113–124.

DeMets DL, Hardy R, Friedman LM, Lan KKG (1984) Statistical aspects of early termination in the Beta-Blocker Heart Attack Trial. *Controlled Clinical Trials* **5**: 362–372.

DeMets DL, Fleming TR, Whitley RJ, Childress JF, Ellenberg SS, Foulkes M, Mayer KH, O'Fallon J, Pollard RB, Rahal JJ, Sande M, Straus S, Walters L, Whitley-Williams P (1995) The Data and Safety Monitoring Board and acquired immune deficiency syndrome (AIDS) clinical trials. *Controlled Clinical Trials* **16**: 408–421.

DeMets DL, Pocock SJ, Julian DG (1999) The agonising negative trend in monitoring of clinical trials. *Lancet* **354**: 1983–1988.

Diabetic Retinopathy Study Research Group (1976) Preliminary report on effects of photocoagulation therapy. *American Journal of Ophthalmology* **81**: 383.

Dixon DO, Lagakos SW (2000) Should data and safety monitoring boards share confidential interim data? *Controlled Clinical Trials* **21**: 1–6.

Ellenberg SS, Geller N, Simon R, Yusuf S (eds) (1993a) Proceedings of 'Practical issues in data monitoring of clinical trials', Bethesda, Maryland, USA, 27–28 January 1992. *Statistics in Medicine* **12**: 415–616.

Ellenberg SS, Myers MW, Blackwelder WC, Hoth DF (1993b) The use of external monitoring committees in clinical trials of the National Institute of Allergy and Infectious Diseases. *Statistics in Medicine* **12**: 461–467.

European Myocardial Infarction Project (1988) Potential time saving with pre-hospital intervention in acute myocardial infarction. *European Heart Journal* **9**, 118–124.

Fleming TR, DeMets DL (1993) Monitoring of clinical trials: issues and recommendations. *Controlled Clinical Trials* **14**:183–197.

Freidlin B, Korn EL, George SL (1999) Data monitoring committees and interim monitoring guidelines. *Controlled Clinical Trials* **20**: 395–407.

Green SJ, Fleming TR, O'Fallon JR (1987) Policies for study monitoring and interim reporting of results. *Journal of Clinical Oncology* **5**: 1477–1484.

Haybittle JL (1971) Repeated assessment of results in clinical trials of cancer treatments. *British Journal of Radiology* **44**: 793–797.

Heart Special Project Committee (1988) Organization, review and administration of cooperative studies: A report from the Heart Special Project Committee to the National Advisory Heart Council, May 1967. *Controlled Clinical Trials* **9**: 137–148.

Hypertension Detection and Follow-up Program Cooperative Group (1979) Five-year findings of the hypertension detection and follow-up program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension. *Journal of the American Medical Association* **242**(23): 2562–2571.

Intermittent Positive Pressure Breathing Trial Group (1983) Intermittent positive pressure breathing therapy of chronic obstructive pulmonary disease. *Annals of Internal Medicine* **99**(5): 612–620.

Lan KKG, DeMets DL (1983) Discrete sequential boundaries for clinical trials. *Biometrika* **70**: 659–663.

Meinert CL (1998a) Clinical trials and treatment effects monitoring. *Controlled Clinical Trials* **19**: 515–521.

Meinert CL (1998b) Masked monitoring in clinical trials – blind stupidity? *New England Journal of Medicine* **338**: 1381–1382.

Nocturnal Oxygen Therapy Trial Group (1980) Continuous or nocturnal oxygen therapy in hypoxemic chronic obstructive lung disease. A clinical trial. *Annals of Internal Medicine* **93**(3): 91–98.

O'Brien PC, Fleming TR (1979) A multiple testing procedure for clinical trials. *Biometrics* **35**: 549–556.

Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL Jr, Valanis B, Williams JH Jr, Barnhart S, Hammar S (1996) Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *New England Journal of Medicine* **334**: 1150–1155.

Parmar MKB, Machin D (1993) Monitoring clinical trials: experience of, and proposals under consideration by, the cancer therapy committee of the British Medical Research Council. *Statistics in Medicine* **12**: 497–504.

Persantine-Aspirin Reinfarction Study Research Group (1980) Persantine and aspirin in coronary heart disease. *Circulation* **62**: 449–461.

Pocock SJ (1977) Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64**: 191–199.

Pocock SJ (1993) Statistical and ethical issues in monitoring clinical trials. *Statistics in Medicine* **12**: 1459–1469.

Smith MA, Ungerleider RS, Korn EL, Rubinstein L, Simon R (1997) Role of independent data-monitoring committees in randomized clinical trials sponsored by the National Cancer Institute. *Journal of Clinical Oncology* **15**: 2736–2743.

Swedberg K, Held P, Kjekshus J, Rasmussen K, Ryden L, Wedel H (1992) Effects of early administration of enalapril on mortality in patients with acute myocardial infarction: results of the Cooperative North Scandinavian Enalapril Survival Study II. *New England Journal of Medicine* **327**: 678–684.

Urokinase Pulmonary Embolism Trial Study Group (1970) Urokinase pulmonary embolism trial: phase I results. A cooperative study. *Journal of the American Medical Association* **214**(12): 2163–2172.

Urokinase-Streptokinase Pulmonary Embolism Trial Study Group (1974) Urokinase-streptokinase pulmonary embolism trial: phase II results. A cooperative study. *Journal of the American Medical Association* **229**(12): 1606–1613.

US Food and Drug Administration (1997) Guidance for industry: Good clinical practices. *Federal Register* **62**: 25 691–25 709.

US Food and Drug Administration (1998) Guidance for industry: Statistical principles in clinical trials. *Federal Register* **63**: 49 583–49 598.

US Food and Drug and Administration (2001) *Guidance for Clinical Trial Sponsors on the Establishment and Operation of Clinical Trial Data Monitoring Committees*. Rockville, MD: FDA. http://www.fda.gov/cber/gdlns/clindatmon.htm.

Wall Street Journal (1994) Small fast-growth firms feel chill of shareholder suits. *Wall Street Journal*, April 5.

Wang SK, Tsiatis AA (1987) Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics* **39**: 193–199.

Whitehead J (1999) On being the statistician on a data and safety monitoring board. *Statistics in Medicine* **18**: 3425–3434.

Zapol WM, Snider MT, Hill JD *et al.* (1979) Extracorporeal membrane oxygenation in severe acute respiratory failure. A randomized prospective study. *Journal of the American Medical Association* **242**(20): 2193–2196.