CASE 14

Data Monitoring in the Prospective Randomized Milrinone Survival Evaluation: Dealing With an Agonizing Trend

Susan Anderson
Robert Cody
Milton Packer
Richard Schwarz

ABSTRACT

The Prospective Randomized Milrinone Survival Evaluation (PROMISE) was conducted to clarify discordant findings in previous experimental and clinical studies with milrinone, a cyclic AMP-enhancing positive inotropic agent. Earlier studies had shown positive effects of milrinone on cardiac function and exercise performance in patients with chronic heart failure. To determine the effect of milrinone on mortality, patients with severe chronic heart failure who remained symptomatic despite conventional therapy were randomized to receive either active drug or a matching placebo. The trial was terminated after 20 months, before its scheduled completion, based on an observed adverse effect of milrinone on survival. This paper describes the experience of the Data Monitoring and Safety Monitoring Board (DSMB) in dealing with an emerging negative trend in survival when there were other known beneficial effects of the drug.

INTRODUCTION AND BACKGROUND

Chronic heart failure is an increasing problem with an aging population. Over 500,000 cases are diagnosed each year, and the mortality risk for these patients remains unacceptably high. In the early 1990s few effective treatments were available and the search for new effective agents was a high priority.

Cyclic AMP-enhancing agents are among the drugs developed to enhance the inotropic state of the failing heart. Because the production of cyclic AMP is deficient in patients with advanced heart failure, the use of cyclicAMP-enhancing positive inotropic agents had theoretical appeal.[1–3] Experimental studies of heart failure in rats with milrinone showed an encouraging attenuation in the progression of ventricular enlargement after acute myocardial injury and prolongation of survival.[4,5] Despite the theoretical appeal, clinical studies of positive inotropic agents were largely unfavorable, raising concern that cyclic AMP-enhancing agents may accelerate the progression of disease, ventricular arrhythmias, and possibly shorten survival of patients with chronic heart failure.[5–11]

PROTOCOL DESIGN

Because of three major limitations of the earlier clinical studies[1] (most had been carried out in patients with mild-to-moderate symptoms,[2] most were conducted in patients not taking angiotensin converting-enzyme inhibitors,[3] and all were too small to evaluate influence of therapy on survival), PROMISE was designed to evaluate the effect of milrinone on survival in patients with severe chronic heart failure who remained symptomatic despite conventional therapy, which included digoxin, diuretics, and a converting-enzyme inhibitor.[12] Patients had dyspnea or fatigue at rest or on exertion, left ejection fraction ≤35% and symptoms of NYHA functional class III or IV for at least three months (including symptoms at rest within two weeks). Treatment with vasodilator drugs was allowed but not mandated.

Patients who met these and other eligibility requirements in screening assessments were randomized to receive milrinone (10 mg orally four times daily) or matching placebo, in addition to digoxin, diuretics, and a converting-enzyme inhibitor (captopril or enalapril).

The primary endpoint was death due to all causes. Secondary endpoints included cardiovascular mortality, number of hospitalizations, and addition of vasodilators due to worsening heart failure, symptoms, and adverse reactions. In addition, the effect of milrinone on survival was to be assessed in pre-specified subgroups defined by important prognostic baseline variables. The trial was designed to have 90% power to detect a 25% difference in mortality at a 0.05 significance level using a two-tailed logrank test. This design was event-driven and the study was planned to continue until 190 deaths had been observed on the placebo arm.

In order to conduct PROMISE as a model parallel to that conventionally used by the National Institutes of Health (NIH), PROMISE investigators modified the NIH model for application to an industry sponsored trial.[13] PROMISE had an independent statistical analysis center reporting to a Data and Safety

Monitoring Board (DSMB) scheduled to meet every four to six months; a Committee of Investigators, who designed the study; a Steering Committee, and a Clinical Coordinating center responsible for day-today policy decisions, all of whom functioned independently of the sponsor (Sterling Research Group). The organizational structure is shown in Figure 1. PROMISE was one of the first industry-sponsored trials to adopt such a model.[13] The Principal Investigator for the study as well as a few representatives (clinical, regulatory, and statistical) of the sponsor were present throughout the DSMB deliberations but were not voting members of the DSMB.

The trial began recruiting in January of 1989 and was projected (based on total mortality event rates) to be completed in March 1991. In October of 1990, the DSMB recommended to the sponsor that PROMISE be terminated early due to an observed adverse effect of milrinone on survival, particularly among patients with NYHA functional class IV.[12] Selected baseline

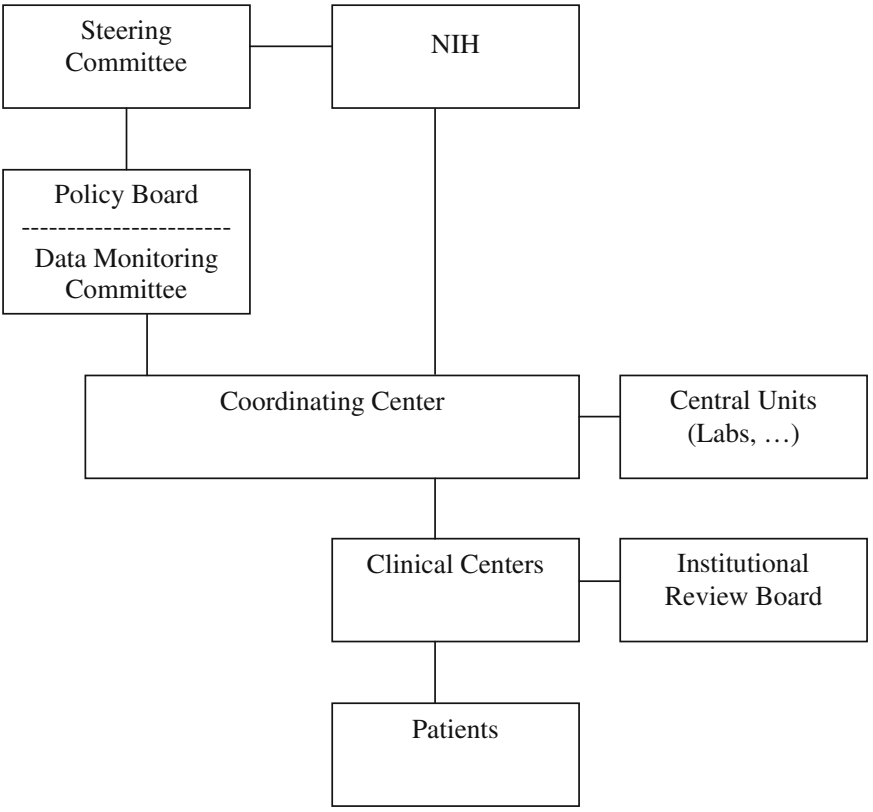

Figure 1 The NIH model.[13]

Table 1 Selected Baseline Characteristics by Treatment Group

Characteristic	Placebo	Milrinone
Number	527	561
Age	64.2	63.1
Gender (%Male)	80%	76%
Principal diagnosis: CHD	54%	54%
Functional class		
III	57%	58%
IV	43%	42%
Angina–previous mycardial infarction	27%	26%
Previous cardiac surgery	41%	39%

characteristics of the 1,088 randomized patients are shown in Table 1. The final results are shown in Figures 2 and 3 and in Table 2, indicating that the milrinone-treated patients had a higher mortality rate than those patients on placebo. The remainder of this discussion will be on the study history and decision process that leading to the recommendation.

DATA MONITORING EXPERIENCE

In November 1988, prior to the start of the study, the DSMB met to review the protocol and establish procedures for monitoring. At the request of the DSMB a written document, or charter, was prepared to specify the general guidelines for interim analysis and evaluation of interim data, including criteria for early stopping. Two-sided symmetric O'Brien–Fleming type boundaries as implemented by Lan and DeMets to allow flexibility in the number and timing of interim analyses while maintaining the total alpha at 0.05 were adopted.[14,15] The O'Brien–Fleming sequential boundaries were truncated at ±3.5 for the very early interim analyses. A number of considerations for the interpretation of the study data as an entirety were explicated and the guidelines stated that recommendation to modify or terminate the trial should not be based totally on statistical grounds. This document also specified procedures to be used to adjust the sample size in order to reach the target of 190 placebo deaths should the initial estimate of placebo mortality rate be incorrect.

In July of 1989, the DSMB met to review study data for the first time. At that time data was available on 233 patients enrolled in the study, and 19 patients had died, 6 on arm A and 13 on arm B. While the DSMB reviewed initially the monitoring report by code (Treatment A and Treatment B), they elected to be informed of treatment identity at this first meeting. Treatment

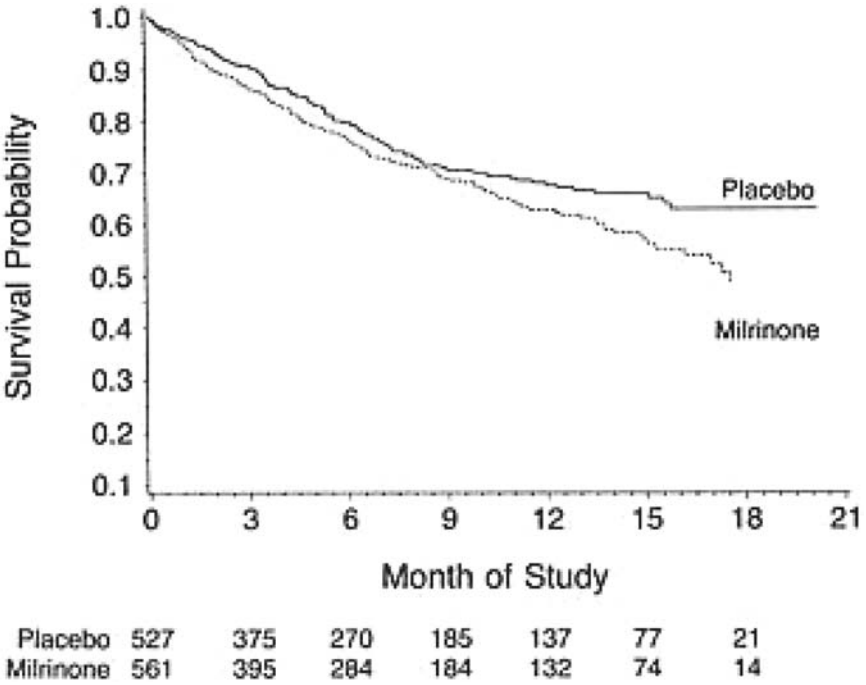

Figure 2 Kaplan–Meier analysis showing cumulative rates of survival in patients with chronic heart failure treated with milrinone or placebo. Mortality was 28% higher in the milrinone group than in the placebo group ($p = 0.038$). The numbers of patients at risk are shown at the bottom of the figure. Reproduced from Packer, et al. (1991) with permission of the *N Engl J Med.*

arm B was identified as milrinone. Of the 13 deaths in the milrinone arm, most of the adverse effect was in the most severe patients as determined by NYHA class (NYHA III 3 placebo versus 1 milrinone, NYHA IV 3 placebo vs. 12 milrinone). The logrank for the survival comparison at this meeting was −1.14, as shown in Figure 4. The DSMB elected to keep treatments coded in the reports but to maintain the same coding of treatments throughout a given monitoring report and across interim analyses. While the mortality trend was in the wrong or negative direction, the evidence was not judged as being convincing of harm and the DSMB recommended continuation of the trial.

At the second interim analysis in December 1989, a total of 450 patients had been enrolled. The mortality difference (logrank $Z = -1.50$) remained unfavorable to the treatment arm but was well below the monitoring guideline. The difference in observed deaths remained primarily among the patients who had NYHA class IV symptoms at baseline (NYHA III 11 placebo

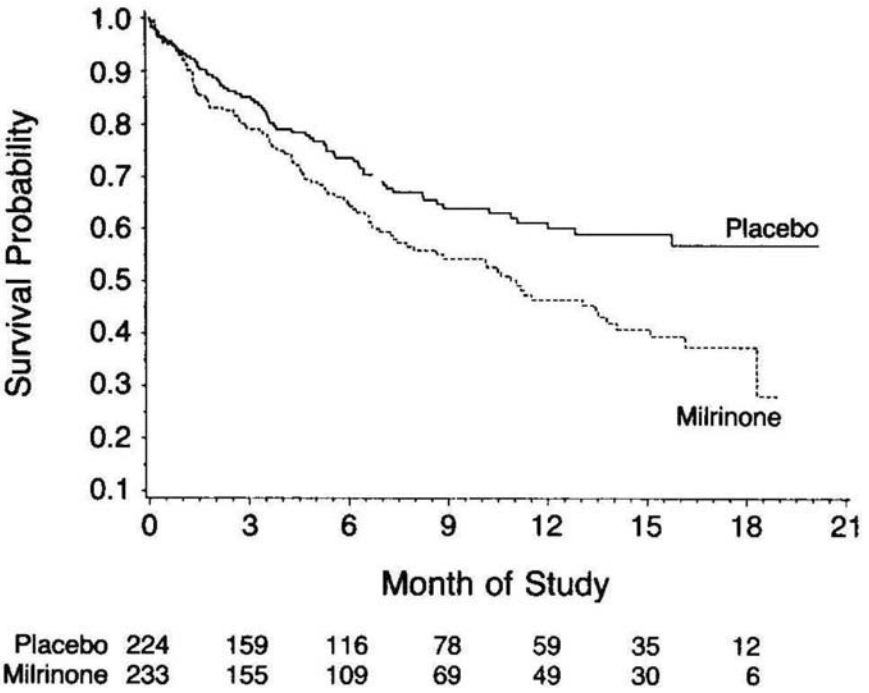

Figure 3 Kaplan–Meier analysis showing cumulative rates of survival in patients with class IV heart failure, According to Treatment Group. Mortality was 53% higher in the milrinone group (p = 0.006). Reproduced from Packer, et al. (1991) with permission of the *N Engl J Med.*

Table 2 Mortality Hazard Ratios by Prognostic Variables

Variable	Hazard ratio	p-value
Ejection fraction		
<0.21	1.26	0.115
>0.21	1.33	0.155
Principal diagnosis		
CHD	1.28	0.101
Other	1.26	0.214
Functional class		
III	1.03	0.859
IV	1.53	0.006
Age/yr		
<65	1.35	0.108
>65	1.34	0.051
Gender		
Male	1.26	0.082
Female	1.33	0.280

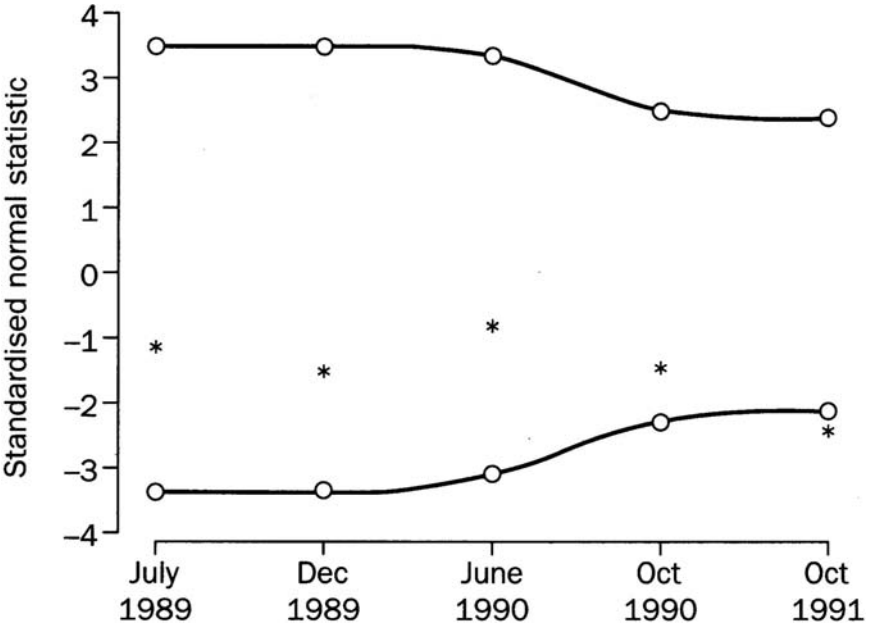

Figure 4 Group sequential boundries for the PROMISE trial. Horizontal axis = information fraction (observed fraction of total expected deaths). Group sequential boundaries set at two-sided 5% significance. Plotted points = logrank test. Crossing upper boundary = benefit, crossing lower boundary = harm. Reproduced from DeMets, et al. (1999) with permission of the *Lancet.*

vs. 11 milrinone, NYHA IV 17 placebo vs. 30 milrinone). Again, the DSMB recommended continuation of the trial.

By the third interim analysis in May 1990, a total of 683 patients of the initial estimated total sample size of 800 patients had been enrolled. The mortality difference was still unfavorable (logrank Z = −0.82) but less significant than at the previous monitoring meeting. Among patients with baseline NYHA class III symptoms, the mortality difference (logrank Z = 0.29) was very slightly in favor of the treatment arm but among patients with class IV symptoms the difference (logrank Z = −1.14) was still unfavorable though not nominally significant. At this analysis the DSMB considered two sets of projections. One set, the conditional power calculations, evaluated the likelihood of reaching a positive or negative conclusion (crossing of either the upper monitoring boundary indicating benefit or the lower boundary indicating harm) by the end of the trial given the current observed mortality and under a range of assumptions about the underlying mortality treatment difference. For example, the projected milrinone effect for the remainder of the

trial included the beneficial effect assumed in the design, half of that effect, a null effect, and the observed negative effect.

The second set of projections addressed the necessity for extension of enrollment and/or total study length due to a smaller overall event rate than expected in order to obtain the target 190 placebo arm observed deaths. These calculations indicated that in order to reach the target by the expected completion date of March 1991 enrollment would need to be continued at least through October 1990, the time of the next scheduled DSMB review. The DSMB recommended continuation of enrollment and continuation of the study.

At the fourth interim monitoring meeting in October 1990, data were now available on 1,013 randomized patients. The negative trend in mortality continued, with 114 deaths on placebo and 143 deaths on milrinone. Although the overall mortality comparison (logrank $Z = -1.50$) was still within the monitoring guidelines, the comparison within the NYHA class IV subgroup (65 placebo versus 93 milrinone) was nominally statistically significant (logrank $Z = -2.55$, $p = 0.01$) and in fact was larger than the monitoring bound in place for the overall comparison, as shown in Figure 4. The DSMB recommended that the sponsor terminate the trial and initiate a complete surveillance of all patients for mortality status at the end of the trial. The sponsor and study chair were present at this DSMB meeting, participated in the discussion, but did not vote on the recommendation.

The trial was stopped promptly by the sponsor and Clinical Coordinating Center contacted the investigators to determine survival at the close of the study. In total 1,094 patients were enrolled in PROMISE. The final mortality experience was 127 deaths among 527 patients randomized to placebo and 168 deaths among 567 patients randomized to milrinone. The normalized logrank statistic ($Z = -2.08$) for the comparison of survival was just across the monitoring boundary for the final analysis. The final mortality difference in the NYHA class III subgroup slightly favored placebo (logrank $Z = -0.17$) but was not as striking as the difference in the NYHA class IV subgroup (logrank $Z = -2.77$).

LESSONS LEARNED

In the PROMISE trial the observed estimate of milrinone on survival was negative from the first monitoring meeting onward. As monitoring progressed and information (total deaths) accumulated. It seemed increasingly unlikely that PROMISE would show a mortality benefit for milrinone. However discouraging the mortality evidence, the drug milrinone was believed to improve other clinical measures of heart function, which could improve quality of life for patients with severe heart failure and thus perhaps

could be beneficial even without a survival benefit. A neutral mortality result might not be a reason to abandon use of milrinone. However, a truly harmful or negative effect of milrinone on mortality would be an important deterrent. Thus, the DSMB felt the need to purse this agonizing negative trend to distinguish between a neutral mortality effect from a truly harmful effect. This situation in general has been recently been discussed.[16] Throughout the PROMISE trial, the DSMB reviewed quality-of-life measures and changes from baseline in symptoms and measures of heart function. The DSMB recommended closing the trial when a significant negative effect on survival was apparent and outweighed any observed potential symptomatic benefit. In addition, the likelihood of acceptance of conclusion and disease background against which the trial is conducted were other factors seriously considered.

PROMISE was one of the first industry-sponsored trials to implement a fully independent data and safety monitoring board, supported by an independent statistical analysis center. The goal was to obtain the benefits of the clinical trial model pioneered by the NIH, especially with respect to credibility and acceptability by the cardiology community and to provide adequate monitoring for overall patient safety. While the results of milrinone were not expected, in fact the modified NIH clinical trial achieved the goal and performed well with the observed negative harmful treatment effect. This PROMISE model has been modified and adopted by several other trials.[13] Furthermore, this type of independent statistical center has also been suggested by the Food and Drug Administrtion (FDA) guideline on data monitoring committees.[17]

In PROMISE, the study chair and sponsor attended all parts of the DSMB meeting. The current practice of open, closed, and executive sessions was not yet widely practiced. Open sessions typically allow sponsor and investigator participation. In closed sessions, the DSMB and the statistical center independent statistician are in attendance. In the executive session, only the DSMB members attend and form their final recommendations. In PROMISE, study chair and sponsor attendance did not appear to interfere with any of the DSMB deliberations, but it would be hard to claim there was no influence at all. In hindsight, and with further experience using DSMBs for industry-sponsored trials, the open, closed, and executive session format would be the preferred or recommended practice.[13,17,18]

Analysis of subgroups is always a challenge due to the vulnerability of multiple comparisons and false claims. Monitoring overall results as well as selected subgroups is even more challenging. Not only are their typically several subgroups but these are now reviewed repeatedly, which further increases the chances for false claims. Subgroups also have smaller samples sizes; results are subject to the variability of a smaller number of events and possible imbalances in risk factors. Terminating a subgroup alone may also

have the effect of essentially terminating the entire trial. For PROMISE, the DSMB followed the high-risk (NYHA IV) heart failure subgroup but chose not to terminate this subgroup alone at earlier meetings. Rather, the DSMB sought to have a convincing overall result.

As described previously,[18–20] termination of a trial for benefit or harm is a complex decision process and depends not only on statistical analysis and monitoring boundaries but many other factors. These include internal consistency across various outcomes and subgroups, external consistency with other trials and preclinical data, and impact of the results on the practicing clinicians. In the case of PROMISE, the trial provided a definitive answer that has been accepted, providing important information on the use of the specific drug in the treatment of moderate to severe chronic heart failure.

REFERENCES

1. Packer M. 1988. Vasodilator and inotropic drugs for chronic heart failure: distinguishing hype from hope. *Am J Cardiol* 12:1299–1317.
2. Feldman MD, Copelas L, Gwanthmey JK, et al. 1987. Deficient production of cyclic AMP: pharmacologic evidence of an important cause of contractile dysfunction in patients with end-stage heart failure. *Circulation* 75:331–339.
3. Wilmshurst PT, Walker Jm, Fry CH, et al. 1984. Inotropic and vasodilator effects of amrinone on isolated human tissue. *Cardiovasc Res* 18:302–309.
4. Jain P, Brown EJ Jr, Langenback EG, et al. 1991. Effects of milrinone on left ventricular remodeling after acute myocardial infarction. *Circulation* 84:796–804.
5. Sweet CS, Ludden CT, Stabilito II, Emmert SE, Heyse JF. 1988. Beneficial effects of milrinone and enalapril on long-term survival of rats with healed myocardial infarction. *Eur J Pharmacol* 147:29–37.
6. Packer M, Leier CV. 1987. Survival in congestive heart failure during treatment with drugs with positive inotropic actions. *Circulation* 75:Suppl IV:IV-55–IV-63.
7. DiBianco R, Shabetai R, Kostuk W, Moran J, Schlant RC, Wright R. 1989. A comparison of oral milrinone, digoxin, and their combination in the treatment of patients wiht chronic heart failure. *N Engl J Med* 320:677–683.
8. Baim DS, McDowell AV, Cherniles J, et al. 1983. Evaluation of a new bipyridine inotropic agent—milrinone—in patients with severe congestive heart failure. *N Engl J Med* 309:748–756.
9. Simonton CA, Chatterjee K, Cody RJ, et al. 1985. Milrinone in congestive heart failure: acute and chronic hemodynamic and clinical evaluation. *J Am Coll Cardiol* 6:453–459.
10. Chesebro JH, Browne KF, Fenster PE, Garland WT, Konstam MA. 1988. The hemodynamic effects of chronic oral milrinone therapy: a multicenter controlled trial. *J Am Coll Cardiol* 11:144A. abstract
11. Packer M. 1989. Effect of phosphodiesterase inhibitors on the survival of patients with chronic congestive heart failure. *Am J Cardiol* 63:41A–45A.
12. Packer M, Carver JR, Rodeheffer RJ, Ivanhoe RJ, DiBianco R, Zeldis SM, et al. 1991. Effect of oral milrinone on mortality in severe chronic heart failure. *N Engl J Med* 325:1468–1475.
13. Fisher MM, Roecker EB, DeMets D. 2001. The role of an independent statistical analysis center in the industry-modified National Institutes of Health model. *Drug Information Journal* 35:115–129.
14. Lan KKG, DeMets DL. 1983. Discrete sequential boundaries for clinical trials. *Biometrika* 70:659–663.

15. O'Brien PC, Fleming TR. 1979. A multiple testing procedure for clinical trials. *Biometrics* 35:549–556.
16. DeMets DL, Pocock S, Julian DG: 1999. The agonising negative trend in monitoring clinical trials. *Lancet* 354:1983–1988.
17. US Food and Drug Administration (2001) Draft 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
18. Ellenberg S, Fleming T, DeMets D. 2002. *Data Monitoring Committees in Clinical Trials: A Practical Perspective*. John Wiley & Sons, Ltd., West Sussex, England.
19. Canner PL. 1983. Monitoring of the data for evidence of adverse or beneficial treatment effects. *Control Clin Trials* 4:467–483.
20. Friedman LM, Furberg CD, DeMets DL. 1998. Fundamentals of Clinical Trials. Third Edition, Springer-Verlag, New York.