The Role of the Data and Safety Monitoring Board in a Clinical Trial: The CRISIS Study

Richard Holubkov, PhD[1]; T. Charles Casper, PhD[1]; J. Michael Dean, MD, MBA[1]; K. J. S. Anand, MBBS, DPhil[2]; Jerry Zimmerman, MD, PhD[3]; Kathleen L. Meert, MD[4]; Christopher J.L. Newth, MD, FRCPC[5]; John Berger, MD[6]; Rick Harrison, MD[7]; Douglas F. Willson, MD[8]; Carol Nicholson, MD[9]; Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Collaborative Pediatric Critical Care Research Network (CPCCRN)

Objectives: Randomized clinical trials are commonly overseen by a Data and Safety Monitoring Board comprised of experts in medicine, ethics, and biostatistics. Data and Safety Monitoring Board responsibilities include protocol approval, interim review of study enrollment, protocol compliance, safety, and efficacy data. Data and Safety Monitoring Board decisions can affect study design and conduct, as well as reported findings. Researchers must incorporate Data and Safety Monitoring Board oversight into the design, monitoring, and reporting of randomized trials.

Design: Case study, narrative review.

Methods: The Data and Safety Monitoring Board's role during the comparative pediatric Critical Illness Stress-Induced Immune Suppression (CRISIS) Prevention Trial is described.

Findings: The National Institutes of Health-appointed CRISIS Data and Safety Monitoring Board was charged with monitoring sample size adequacy and feasibility, safety with respect to adverse events and 28-day mortality, and efficacy with respect to the primary nosocomial infection/sepsis outcome. The Federal Drug Administration also requested Data and Safety Monitoring Board interim review before opening CRISIS to children below 1 yr of age. The first interim analysis found higher 28-day mortality in one treatment arm. The Data and Safety Monitoring Board maintained trial closure to younger children and requested a second interim data review 6 months later. At this second meeting, mortality was no longer of concern, whereas a weak efficacy trend of lower infection/sepsis rates in one study arm emerged. As over 40% of total patients had been enrolled, the Data and Safety Monitoring Board elected to examine conditional power and unmask treatment arm identities. On finding somewhat greater efficacy in the placebo arm, the Data and Safety Monitoring Board recommended stopping CRISIS due to futility.

Conclusions: The design and operating procedures of a multicenter randomized trial must consider a pivotal Data and Safety Monitoring Board role. Maximum study design flexibility must be allowed, and investigators must be prepared for protocol modifications due to interim findings. The Data and Safety Monitoring Board must have sufficient clinical and statistical expertise to assess potential importance of interim treatment differences in the setting of multiple looks at accumulating data with numerous outcomes and subgroups. (*Pediatr Crit Care Med* 2013; 14:0–0)

Key Words: clinical trials; interim analysis; nosocomial infection; randomized; safety; sepsis

External oversight of interventional studies, including randomized clinical trials, is standard in contemporary clinical research. For example, the National Institutes of

[1] Division of Critical Care, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT.
[2] Departments of Pediatrics, Anesthesiology, and Neurobiology, Le Bonheur Children's Hospital, University of Tennessee Health Science Center, Memphis, TN.
[3] Department of Pediatrics, Seattle Children's Hospital, University of Washington School of Medicine, Seattle, WA.
[4] Department of Pediatrics, Children's Hospital of Michigan, Detroit, MI.
[5] Department of Anesthesiology and Critical Care Medicine, Children's Hospital Los Angeles, Los Angeles, CA.
[6] Department of Pediatrics, Children's National Medical Center, Washington, DC.
[7] Department of Pediatrics, University of California at Los Angeles, Los Angeles, CA.
[8] Department of Pediatrics, University of Virginia Children's Hospital, Charlottesville, VA.
[9] Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD.

Supported, in part, by cooperative agreements from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Department of Health and Human Services (DHHS) (U10HD050096, U10HD049981, U10HD500009, U10HD049945, U10HD049983, U10HD050012 and U01HD049934).

The authors have not disclosed any potential conflicts of interest.

For information regarding this article, E-mail: rich.holubkov@hsc.utah.edu

DOI: 10.1097/PCC.0b013e318274568c

Health (NIH) requires all agencies to establish a Data and Safety Monitoring Board (DSMB) for Phase III multicenter clinical trials involving potential risk to participants (1), and NIH agencies require DSMBs in earlier phase trials that involve vulnerable populations, including children (2). DSMBs typically review and approve the final protocol before enrollment occurs and meet periodically during the conduct of the trial to review all aspects of study progress, including patient enrollment, protocol compliance, data quality and completeness, reported adverse events, and other safety data. In many randomized trials, the DSMB additionally reviews interim efficacy of the proposed intervention and may recommend early termination due to evidence of efficacy (one treatment arm being superior to the other) and/or futility (the trial having little chance of demonstrating superiority of one treatment).

In this report, we discuss the role of the NIH-appointed DSMB during the planning and conduct of the randomized comparative pediatric Critical Illness Stress-Induced Immune Suppression (CRISIS) Prevention Trial, conducted within the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) Collaborative Pediatric Critical Care Research Network (CPCCRN). CRISIS compared the effect of daily supplementation with zinc, selenium, glutamine, and metoclopramide, vs. whey protein, on the occurrence of nosocomial infection/sepsis among long-stay intensive care patients aged from 1 to 17 yrs. The CRISIS study protocol and primary study results have been reported previously (3, 4).

MATERIALS AND METHODS

Study Design

This report is a narrative review of the role of the DSMB during the design and execution of CRISIS. Clinical and biostatistical issues addressed by the DSMB in the final development of CRISIS are discussed, along with the DSMB's decision process during two interim analysis meetings that culminated in the early stopping of CRISIS. We discuss general applications of our experience in CRISIS for future pediatric randomized trials.

The Institutional Review Boards of all CPCCRN centers approved the CRISIS protocol and informed consent documents. Parental permission was provided for each subject.

Key Definitions

Interim analysis is examination of available trial data (safety and possibly efficacy) at a timepoint before target recruitment has been reached, with the possibility of stopping or modifying the study based on the findings.

Efficacy *monitoring boundaries* are statistical guidelines for recommending whether a trial should be stopped at an interim analysis due to evidence that one treatment arm is superior with respect to efficacy. These prospectively determined boundaries are designed to limit the overall Type I error, or the chance that a trial reports a significant treatment effect when none truly exists, to an overall value such as 5%, considering multiple looks at the accumulating study data.

Conditional power of a trial is the chance that the (partially completed) trial will ultimately report a statistically significant treatment effect, given the treatment effect currently observed among patients for whom the outcome is already known. Conditional power can be evaluated under various scenarios (e.g., if the true treatment effect matches the magnitude that was initially expected, or the magnitude currently observed).

Futility is the state of a trial when interim analysis indicates it is unlikely that the trial will generate statistically significant findings if continued (e.g., the conditional power of the trial is judged unacceptably low).

Statistical Methods

The motivation for monitoring boundaries is that repeated analyses of accumulating data can increase the chances of false-positive claims if standard statistical methods are used for each interim analysis with no adjustments for the repetition. For example, assume we are testing the hypothesis of a significant difference between two treatments with a desired Type I error (also termed "α level") of 5%, declaring a significant treatment difference if we find $p < 0.05$. If there is truly no treatment difference, and we analyze our study data twice (once at the trial's halfway point, once at study end), the chance that at least one of the two analyses will show $p < 0.05$ is 8% rather than 5% (5). The chances of such a false-positive finding increase to 14% with five equally spaced interim data analyses and to 20% with 10 analyses.

Various statistical methods are available to control the study-wide Type I error accounting for multiple analyses. A very general, commonly used approach is the "alpha-spending function" (6), which prespecifies the Type I error to be used at each interim analysis, according to the proportion of the total study's statistical information available at each interim analysis. These functions (of which there are infinitely many, as Type I error spending over time can be varied per each trial's requirements) control the studywide α-level while allowing the number and exact timing of interim analyses to be flexible. Such flexibility is desirable because DSMB meetings are usually scheduled months in advance without knowledge of exact number of patients who will have outcome data and because DSMBs may schedule additional meetings in response to concerns either within or outside of the clinical trial.

RESULTS

CRISIS Study Design

The primary efficacy outcome in CRISIS was time to nosocomial infection or sepsis. Inclusion and exclusion criteria have been previously reported (3, 4). Nosocomial infection was clinically defined as a new microbiologically proven infection in a patient with fever, hypothermia, chills, or hypotension (7). Sepsis was defined as fever, hypotension, or oliguria, leading to initiation of new antibiotic therapy without microbiologic evidence of infection or other recognized cause of symptoms. Enrolled children were considered at risk for this outcome from 48 hrs after PICU admission until the earlier of hospital discharge or 3 days after PICU discharge. In the double-blind CRISIS setting, site investigators initially reported positive outcomes (dates of any infection and/or sepsis events) for study patients. For the final outcome, performance site investigator determinations

were reviewed and adjudicated by the (treatment-masked) CPCCRN Steering Committee during in-person final adjudication meetings, based on daily histories of relevant symptoms, cultures, and use of antibiotics for each patient.

CRISIS was designed to have sufficient power to detect a "hazard rate" for infection 1.5-fold higher in the whey arm compared with the active study arm. Assuming that the time-to-infection "event curves" follow an exponential distribution, this magnitude of relative hazard implies that the timepoint at which half of patients exhibit infection or sepsis would be 1.5 times higher in the active arm compared with the whey arm (e.g., a median time to infection of 6 days in the active arm vs. 4 days in the whey arm). **Table 1** shows numbers of patients needed to achieve 80% and 90% power under various assumptions regarding the infection/sepsis rate in the whey arm, number of days each patient is at risk for developing infection, and the hazard rate in the active arm relative to the whey arm. As the critical event rate and days-at-risk parameters were unknown, the sample size in CRISIS was initially specified as 600–800 patients.

CRISIS Analytic Plan

The primary analysis in CRISIS was specified as a time-to-event analysis of time until first infection/sepsis to be

TABLE 1. CRISIS Sample Size Requirements Under Various Assumptions About Time to Infection, Days at Risk, and Hazard Ratio

Median Time to Nosocomial Infection/Sepsis Among Patients at Risk	Median Days Patients Are at Risk for Outcome	Hazard for an Event in Placebo Arm, Relative to the Active Arm	Patients Required for 80% Power	Patients Required for 90% Power
6 days	3 days	1.5-fold	662	886
		2-fold	254	340
		2.5-fold	160	214
6 days	5 days	1.5-fold	476	636
		2-fold	180	240
		2.5-fold	112	150
6 days	7 days	1.5-fold	396	530
		2-fold	148	198
		2.5-fold	92	124
8 days	3 days	1.5-fold	820	1096
		2-fold	316	424
		2.5-fold	200	268
8 days	5 days	1.5-fold	570	762
		2-fold	218	290
		2.5-fold	136	182
8 days	7 days	1.5-fold	464	620
		2-fold	176	234
		2.5-fold	110	146
13 days	3 days	1.5-fold	1212	1622
		2-fold	474	634
		2.5-fold	302	404
13 days	5 days	1.5-fold	806	1078
		2-fold	312	416
		2.5-fold	196	264
13 days	7 days	1.5-fold	634	850
		2-fold	244	326
		2.5-fold	154	204

summarized by Kaplan–Meier "survival curves" and compared between treatment arms by the log-rank test, stratified by patient status as immunocompromised or immunocompetent at study entry. A secondary, supportive analysis would compare rates of events per study day (allowing counting of multiple events in the same patient) between study arms using Poisson count models. Additional secondary efficacy outcomes included study days free from antibiotic use and prolonged lymphopenia (absolute lymphocyte count $\leq 1000/mm^3$ for 7 or more consecutive study days).

As typically occurs in larger clinical trials, several patient subgroups were prespecified for analysis, including immunocompromised status at entry (vs. not), surgical procedure immediately preceding PICU admission (vs. not), gender, race/ethnicity, and clinical center.

As this critically ill population of children was expected to develop substantial numbers of adverse events related to their underlying medical conditions, the initial trial protocol specified that unexpected adverse events would be collected and assessed according to severity and relationship to the study drug.

Initial DSMB Monitoring Plan

The original CRISIS analytic plan proposed that after an initial meeting to review the final protocol, the DSMB would meet twice for interim safety and efficacy analyses, after approximately 200 and 400 patients had completed the study. At the time of the first interim analysis, the DSMB would also be asked to approve a final study sample size. Because modifying a trial's sample size conditional on knowledge of observed treatment effect can also modify the study's chances of incorrectly declaring a significant treatment effect (8), the DSMB's sample size review would be performed without knowledge of the observed treatment effect at time of interim analysis. Parameters such as overall rates of infection/sepsis and distributions of PICU length of stay across both study arms combined would be examined and the final sample size determined (along the lines of the Table 1 calculations) before DSMB review of the interim efficacy data.

For formal interim review of efficacy data, it was proposed that the DSMB adhere to O'Brien-Fleming-type monitoring boundaries (9) to guide stopping recommendations. The CRISIS Data Coordinating Center biostatisticians proposed using these very conservative boundaries (which, with two interim looks, would recommend stopping only if the *p* value for significance of treatment effect were less than or equal to 0.0002 with one-third of the study data available, or less than or equal to 0.012 with two-thirds available) due to potential concerns about unequal study rollout across CPCCRN centers, possible "learning effects" in delivering treatments at the beginning of study implementation (10), criticism of studies stopped early for large effects for statistical as well as clinical reasons (11), and a lower sample size penalty for early looks when boundaries are conservative (12).

The DSMB was to be initially masked to the identity of treatment arms during interim analyses, with study arms labeled as "Arm A" and "Arm B" in all materials presented. The DSMB would have the option to unmask treatment assignments at any time.

Protocol Review by the DSMB

The NIH-appointed CRISIS DSMB, whose membership is listed in the section "Acknowledgements" at the end of this report, was comprised of four experts in the areas of pediatric critical care medicine and biostatistics, who were not affiliated with CRISIS and had no other potential conflicts of interest. The DSMB's operation was formalized in a written charter that specified the DSMB's composition and requirements for membership, along with the projected enrollment, initial meeting schedule, and early stopping guidelines as discussed above.

The initial, in-person CRISIS DSMB meeting occurred in November 2006 prior to initiation of patient enrollment. At this meeting, the DSMB approved the study design including the clinical protocol, frequency of meetings, and monitoring boundaries. However, the DSMB requested that the target sample size calculations at the first interim meeting be made more accurate by taking any treatment effect observed at that time into account. At the time of interim analysis, the Data Coordinating Center (DCC) biostatistician would assess any observed treatment effect, blinded to treatment arm identities, and in the presence of a substantial effect calculate sample size under a scenario assuming that the better-performing arm is the active arm. Specific technical logistics of this approach were to be developed by the DCC prior to the interim analysis.

Federal Drug Administration Review and Input

CRISIS was performed under an Food and Drug Administration (FDA) Investigational New Drug application, and the DCC and FDA interacted during 2006 and 2007 with conference calls and paper/electronic correspondence. Four requests by the FDA substantially affected the study design and conduct: 1) the study's age criteria, by design 40 weeks gestational age to 17 yrs, were to be limited to children aged 1–17 yrs pending safety review by the DSMB after enrollment of 33% of patients; 2) patients in the study were to have all adverse events, expected and unexpected, reviewed from study entry until 28 days after entry and be assessed for survival at 28 days; 3) DCC staff involved in the analysis were to be blinded to the identity of treatment arms in the study; and 4) the interim analyses of the efficacy data in the trial were to be based on numbers of observed events in the trial rather than on the numbers of patients recruited.

The final FDA request above, which was made after the initial DSMB meeting, was most helpful to the trial conduct, facilitating formal study monitoring (as will be illustrated below), as well as assessment of recruitment targets by the DSMB. The statistical power of a trial is determined by the total amount of statistical "information" collected, and for time-to-event trials, this information may be expressed as the total number of patients experiencing an event. In particular, when one study arm is assumed to have an event hazard rate 1.5-fold higher than the other, enrollment until 263 events are observed (increasing to 268 events, if early stopping is possible with conservative monitoring boundaries) yields 90% power to find a significant treatment effect under standard assumptions. Viewing the accumulating study data in this information-based fashion prevents the need to recalculate sample size mid-study, because recruitment simply

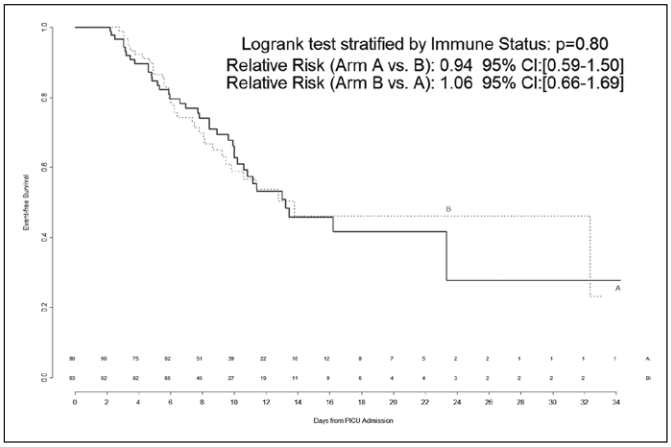

Figure 1. Freedom from nosocomial infection/sepsis according to assigned treatment for all randomized patients, using data available at time of first interim analysis. Numbers above the horizontal time axis denote number of patients remaining at risk at each timepoint. $p = 0.80$ for log-rank test comparing curves between study arms, stratified by immune competent status. CI = confidence interval.

continues (subject to funding and other resource availability) until the required number of events occurs.

First Interim Analysis

CRISIS began recruitment in April 2007, and 204 patients had been enrolled by the end of 2008. The DSMB met in February 2009 to review data for these patients, 183 of whom had infection/sepsis outcomes available. Events occurred in 40% of these 183 patients, leading to an estimated total recruitment of 670 patients to observe the required 268 study events. Based on to-date recruitment of approximately 10 patients per month, an estimated 40 additional months would be required to achieve the required sample size. However, based on screening data about numbers of children excluded from CRISIS solely because they were under 1 yr old, it was estimated that allowing such children into CRISIS might increase enrollment by up to 100%.

The interim analysis of efficacy found approximately equal freedom-from-event curves for the primary nosocomial infection/sepsis outcome between the two treatment arms (**Fig. 1**). Stopping the trial due to efficacy would have been recommended by the monitoring boundaries only if the *p* value for the log-rank test comparing the curves had been less than 0.00004, which was clearly not the case (observed $p = 0.8$). Outcomes were also examined for the prespecified study subgroups; immunocompromised status and gender were subgroup factors for which this interim analysis showed trends toward a differential treatment effect, although no subgroup effects were significant (**Fig. 2**). In light of multiple comparisons and per their clinical expertise, the DSMB was not excessively concerned about these subgroup trends. Analysis of event rates per 100 days was consistent with the above analyses.

Secondary analyses of antibiotic-free days (not shown) found no treatment differences. However, analyses of 28-day mortality and lymphopenia (**Table 2**) found some trends of potential concern, with Treatment A showing a trend toward lower mortality, whereas the Treatment B arm had lower rates of prolonged lymphopenia. The higher Arm B mortality was uniformly observed within patient subgroups. Review of causes of death, and of adverse event rates (comparable between study arms) and types, did not provide definitive information regarding possible cause for the differences noted, which could also have occurred due to chance. The DSMB elected not to unmask treatment arm identities during this interim analysis.

Based on their interim data review, the DSMB recommended that the trial continue; however, based on concerns about the mortality trend, they did not recommend expanding the trial to children under 1 yr of age. In addition, the DSMB elected to add a previously unscheduled meeting and reconvene after approximately six additional months of enrollment, to again review safety and efficacy data, and to reconsider the issue of expanding CRISIS to younger children. No technical or other design modifications were necessary due to this additional meeting, because of the use of flexible monitoring boundaries as discussed above. The DSMB also requested that mortality and efficacy be examined according to patient infection/sepsis status at study entry (presented with existing infection, existing sepsis, or neither).

Second Interim Analysis

The DSMB met again in November 2009 to review data for the 288 patients randomized by the end of October 2009, 273 of whom had infection/sepsis outcomes available. Events had occurred in 41% of patients, leading to a revised estimated requirement of 654 patients to observe 268 with events.

At the time of this second interim analysis, 53/133 (40%) of Arm A and 60/140 (43%) of Arm B patients had experienced events. The corresponding primary event curves, shown in **Figure 3**, indicated a weak trend ($p = 0.16$ by log-rank test) of shorter time to event in Arm B. Updated subgroup curves found continued reversal of treatment effect among immunocompetent vs. immunocompromised patients (**Fig. 4**, top panels), although the subgroup effect was still not significant. When counting multiple events per patient in a secondary Poisson analysis, this subgroup effect was significant ($p = 0.006$ unadjusted for multiplicity), with a significant Treatment B benefit among the 34 immunocompromised patients. The trend of gender-specific differences in time to infection/sepsis observed during the first interim analysis was no longer prominent in the second interim analysis (Fig. 4, bottom panels).

Antibiotic-free days in the PICU (not shown) were again comparable by treatment arm. The updated analysis of mortality and lymphopenia (**Table 3**) found that since the first interim analysis, the difference in 28-day mortality between treatment arms had diminished somewhat in terms of magnitude and statistical significance. Of patients randomized since the first interim analysis, 4/43 in Arm A and 5/50 in Arm B had died at 28 days. Rates of prolonged lymphopenia were now (unadjusted for the many comparisons in the DSMB reports) statistically significantly higher in Arm A, and rates of (at least) moderate lymphopenia were also somewhat higher in this arm.

In summary, at the time of the second interim analysis, 28-day mortality was somewhat higher in Arm B, but there was not

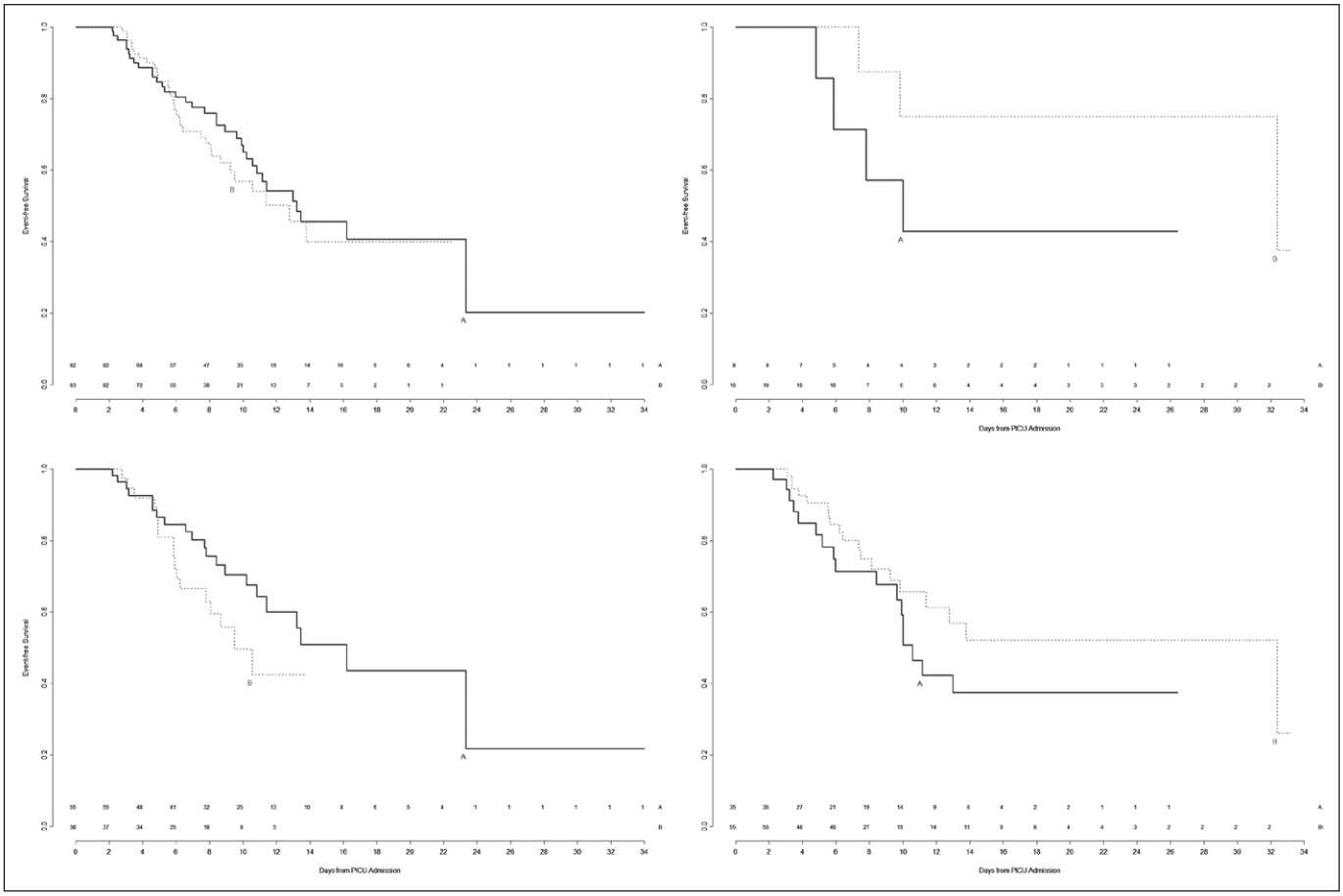

Figure 2. *Top left panel*: Freedom from nosocomial infection/sepsis according to assigned treatment for patients immunocompetent at study entry, using data available at time of first interim analysis. *Top right panel*: Freedom from nosocomial infection/sepsis according to assigned treatment for patients immunocompromised at study entry, using data available at time of first interim analysis. *Bottom left panel*: Freedom from nosocomial sepsis according to assigned treatment for female patients, using data available at time of first interim analysis. *Bottom right panel*: Freedom from nosocomial infection/sepsis according to assigned treatment for male patients, using data available at time of first interim analysis. In all panels, numbers above the horizontal time axis denote number of patients remaining at risk at each timepoint.

substantial concern about the rate differences. Arm A had somewhat lower rates of the primary efficacy outcome but trended toward higher rates of the secondary lymphopenia outcome.

At this point, the DSMB returned to the primary efficacy analysis and elected to consider study futility issues. At this point in the trial, approximately 42% of the total statistical information (113 of the 268 required patients with events) was available, and a trend was emerging of a higher event rate in Arm B. The DSMB wanted to know the estimated conditional power of the study to find a significant treatment effect if it were continued. The DCC biostatisticians, masked to treatment identity, calculated and presented conditional power under various scenarios (**Table 4**).

As the results were trending toward a lower event rate in Arm A with a substantial proportion of patients having completed the study, the power of the study to find a significant result would be substantial, an estimated 61%–86%, if Arm A were truly the superior treatment. On the other hand, if treatment B were truly superior, reducing risk of infection/sepsis 1.5-fold (and the CRISIS interim findings favoring Arm A were a coincidence due to random chance), the power of CRISIS to find a significant effect in favor of either treatment would be

TABLE 2. Observed Mortality and Lymphopenia at Time of First Interim Data and Safety Monitoring Board Analysis

Outcome	Treatment A, *n* = 90	Treatment B, *n* = 93	*p*
All-cause 28-day mortality	4/90 (4.4%)	11/90 (12.2%)	0.059
Prolonged lymphopenia	10/90 (11.1%)	3/93 (3.2%)	0.038
Moderate lymphopenia	21/90 (23.3%)	16/93 (17.2%)	0.30

Three patients assigned to Treatment B had unknown 28-day mortality status.

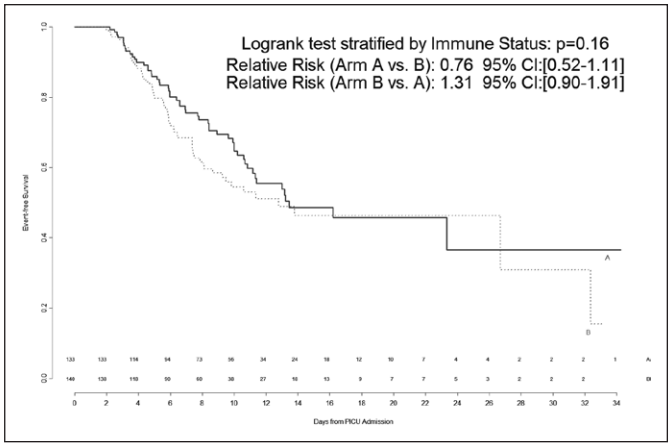

Figure 3. Freedom from nosocomial infection/sepsis according to assigned treatment for all randomized patients, using data available at time of second interim analysis. Numbers above the horizontal time axis denote number of patients remaining at risk at each timepoint. $p = 0.16$ for log-rank test comparing curves between study arms, stratified by immune competent status. CI = confidence interval.

only 10%, as insufficient new patients remained to reverse the trend in the interim data.

Based on the conditional power discussion and the observed differences in lymphopenia rates, the CRISIS DSMB elected to unmask themselves to the identities of the treatment arms. The DCC biostatisticians left the meeting room in order to maintain blinding if the trial were continued. The DCC's pharmacy monitor, necessarily aware of treatment arm identities due to on-site pharmacy visits, was called to the meeting room and opened a prepared sealed envelope with identities of treatment arms. Arm A, with the lower infection/sepsis rate, was the placebo arm. After additional discussion, the DSMB recommended that further recruitment in the CRISIS trial be stopped due to futility. Enrollment was immediately stopped at all centers, with patients still in the trial being followed per protocol although additional treatment with the study agent was halted.

The final CRISIS results, reflecting findings after the few remaining patients had completed follow-up, closely reflected the DSMB-reviewed analysis. CRISIS was reported as a negative study without substantial safety issues, with the immunocompromised subgroup findings potentially worthy of further investigation.

DISCUSSION AND SUMMARY

Increased awareness of complications in clinical research studies, and an ethical imperative to ensure the safety of patients enrolled in a clinical trial, mandate the need for a DSMB. This report describes the role of the DSMB during the design and enrollment phases of the CRISIS trial, with the goal of identifying aspects of future trials that should be considered with the role of the DSMB in mind. A DSMB, which should consist of experts in relevant medical disciplines, biostatistics, and often ethics, should have very wide latitude in the recommendations that they can make. Key aspects of the protocol, such as patient entry criteria or follow-up schedules, may be modified during initial DSMB review (these were actually modified by our FDA reviewers) or after interim analysis. The DSMB may elect to meet more frequently than initially scheduled, necessitating extra clinical and biostatistical effort to gather and analyze data for an unscheduled interim analysis. Finally, the DSMB may recommend early stopping of a trial (or terminating only certain arms of a multi-arm trial, or stopping recruitment of a specific patient subgroup) after interim analysis. Trials and their infrastructure must be constructed with the flexibility to handle these possibilities and others.

From a biostatistical point of view, development of statistical monitoring boundaries for interim efficacy analysis is relatively straightforward for common settings. The study sample size must be adjusted upwards to maintain required statistical power when interim analyses occur; this adjustment is around 1%–3% when conservative monitoring boundaries are used.

Modification of the study sample size midtrial, to maintain desired power if initial estimates of study parameters were inaccurate, is more challenging. Trials where the primary outcome is time-to-event as in CRISIS, as well as studies with event count outcomes, are readily expressible in terms of numbers of events needed to achieve desired power under a specified relative hazard rate. This number of events remains constant regardless of the overall event rate and informs the DSMB precisely about the proportion of total statistical information available during an interim analysis, which is needed to apply monitoring boundaries and recalculate target sample size. We would certainly begin with such an information-based approach when designing future trials involving events. Studies with other types of outcomes, including continuous and binary, can similarly be treated as information-based, although derivations are more challenging (13). We did not encounter the issues of sample-size modification incorporating observed treatment effect. Conventional (14) and Bayesian (15) approaches exist for such midtrial sample size re-estimation; use of these approaches requires appropriate investigator and DSMB expertise, and explicit methodology details must be specified before trial launch.

During interim analysis, it is quite common to find strong, or even statistically significant, trends in the DSMB reports. Such findings are often due to the numerous safety and efficacy endpoints being examined, overall and often within a number of subgroups as well. The mortality trend in the first interim analysis was obviously of key concern as early mortality was the key safety endpoint. Because of the modest number of patients in CRISIS at that time, and limited available information on causes of death (a relatively frequent outcome in this long-term PICU stay population), the available information neither sufficiently assuaged DSMB concerns nor elucidated the potential mechanisms of increased mortality risk. This led to the decision for an additional, unscheduled analysis of study data in 6 months' time, with potential opening of CRISIS to children under 1 yr of age, if there were not continued safety concerns at that time.

During the second interim analysis, concerns about excess treatment-specific mortality were indeed moderated, although the DSMB did not deliberate formally on opening CRISIS to

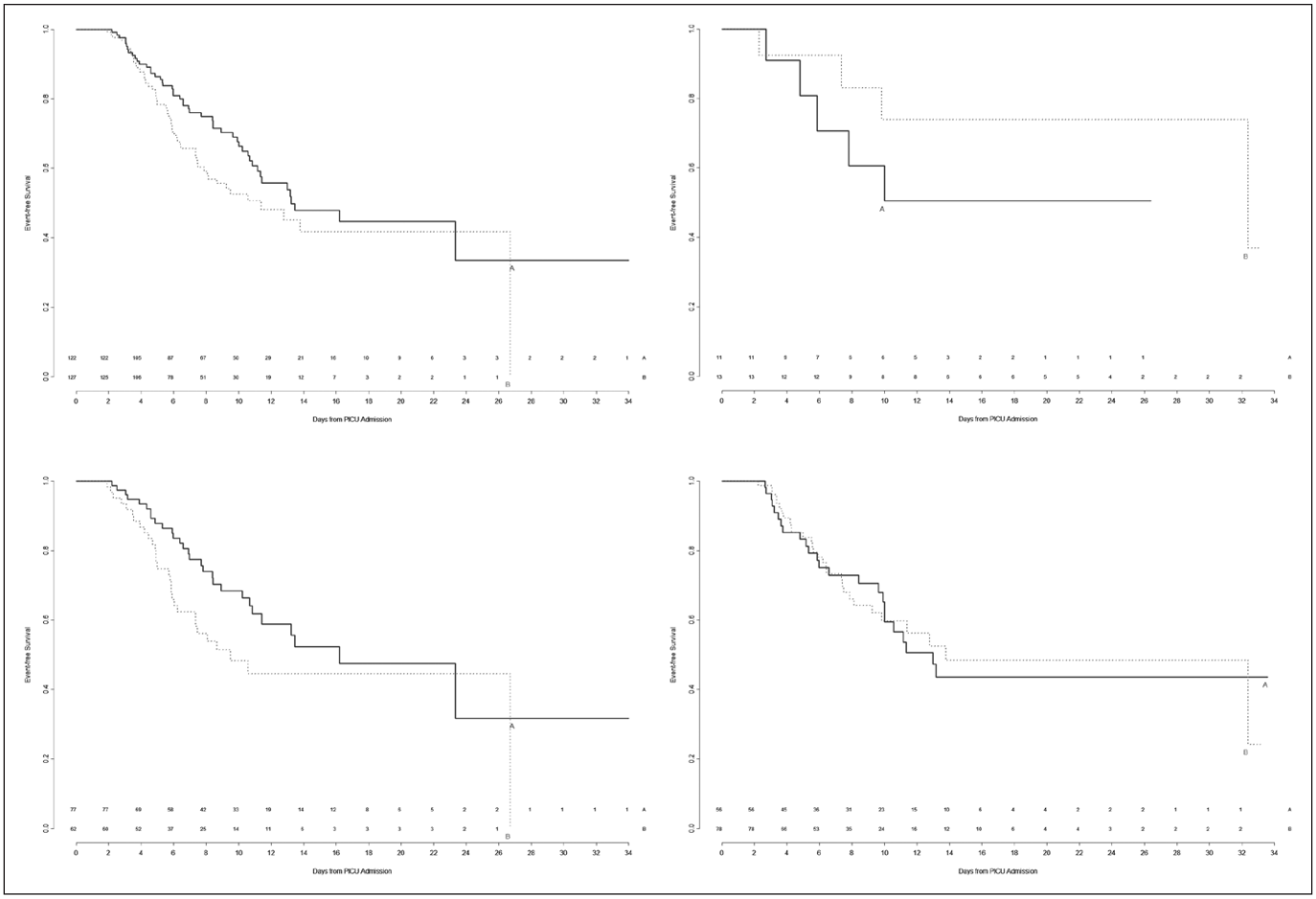

Figure 4. *Top left panel*: Freedom from nosocomial infection/sepsis according to assigned treatment for patients immunocompetent at study entry, using data available at time of second interim analysis. *Top right panel*: Freedom from nosocomial infection/sepsis according to assigned treatment for patients immunocompromised at study entry, using data available at time of second interim analysis. *Bottom left panel*: Freedom from nosocomial infection/sepsis according to assigned treatment for female patients, using data available at time of second interim analysis. *Bottom right panel*: Freedom from nosocomial infection/sepsis according to assigned treatment for male patients, using data available at time of second interim analysis. In all panels, numbers above the horizontal time axis denote number of patients remaining at risk at each timepoint.

younger children. Attention turned to a trend toward improved efficacy in one treatment arm, at a time when nearly one half of enrollment was completed. The DSMB elected to unmask treatment arms in response to this trend. This unmasking, and their subsequent decision to recommend stopping CRISIS enrollment, reflected a view that a potential definitive determination that the whey arm was superior to the active arm (if the trial were continued) was not justifiable, when weighed against potential risks to future study subjects. The observed early mortality differences in the trial may have played a part in this determination.

Some biostatisticians have opined that DSMBs should typically be unmasked to treatment identity beginning from initial data review (16), as the identity of treatment arms is key information potentially useful in decision making. Recent NIH guidance also encourages unmasked review of interim study data by DSMBs (17). It is difficult to conjecture whether the DSMB being unmasked at time of the first interim analysis (and the resulting knowledge that 28-day mortality was trending higher in the active arm, whereas lymphopenia rates were lower in that arm) would have led to different decisions

TABLE 3. Observed Mortality and Lymphopenia at Time of Second Interim Data Safety Monitoring Board Analysis

Outcome	Treatment A, *n* = 133	Treatment B, *n* = 140	*p*
All-cause 28-day mortality	8/133 (6.0%)	16/137 (11.7%)	0.102
Prolonged lymphopenia	13/133 (9.8%)	5/140 (3.6%)	0.039
Moderate lymphopenia	32/133 (24.1%)	21/140 (15.0%)	0.0585

Three patients assigned to Treatment B had unknown 28-day mortality status.

TABLE 4. Conditional Power of CRISIS at Time of Second Interim Analysis

Scenarios	Conditional Power of CRISIS to find a Statistically Significant Treatment Effect (In Either Direction) Under This Scenario
The true treatment effect is exactly as in the current analysis. The hazard of an event is 1.31 times higher in Arm B than in Arm A.	61%
The true treatment effect has the same direction as in the current analysis. However, the true hazard of an event is as hypothesized (a little higher than currently observed): 1.5 times higher in Arm B than in Arm A.	86%
There is truly no treatment difference between Arm A and Arm B.	8%
The true treatment effect is as originally hypothesized—1.5 times higher in one arm than the other. However, due to an unlucky coincidence, the trend in the interim data is reversed, and Arm A has the truly higher event rate.	10%

CRISIS = Critical Illness Stress-Induced Immune Suppression.

At the time of the second interim analysis, a 95% confidence interval for the true hazard ratio for Arm B versus Arm A ranged from 0.70 to 2.44.

regarding study continuation and timing of subsequent data reviews. Blinded review requires simultaneous consideration of different possible scenarios, and the CRISIS DSMB members were sufficiently comfortable with the two possibilities to maintain masking until the second data review.

The decisions to keep CRISIS closed to younger children, to unmask treatments, and to formally assess futility were DSMB specific, as no formal statistical criteria were in place for these decisions. All potential findings of an interim analysis cannot be predicted in advance, and in larger clinical trials formal prespecified guidelines are typically reserved for "alpha-level spending" and other efficacy-related decisions. Although formal futility stopping boundaries can also be constructed when planning a trial (12), many trialists believe that examining conditional power (as was carried out in our setting) is preferable when considering futility (18). When this approach is used, only general guidelines, such as recommending stopping if conditional power is below 20%–25% under most favorable realistic scenarios, are prespecified. Assessing conditional power under various scenarios engages the DSMB in active discussion of interim results, original study assumptions about treatment effect, and other pertinent issues. A criticism of this approach is potentially excessive focus on ultimate statistical significance of the data, rather than on more general usefulness of the trial's findings (19).

A more important controversy than technical futility monitoring details is whether such assessment should be considered at all during DSMB interim data review. If no safety concerns exist about either treatment, there is arguably not an ethical imperative to terminate a trial early solely because statistically significant findings are unlikely (18). When a study is stopped early, the resulting smaller dataset will limit ability to assess (and report to the clinical community) treatment-related complication rates and other safety outcomes. Precise assessment of treatment effect, overall and among important subgroups, will also be compromised, with respect to secondary and primary outcomes (20). The main argument in favor of stopping a likely futile trial is that the resources of the study sponsor and/or research network will be immediately made available for examining other potentially effective treatments or for studying other important research topics that are "in the pipeline" (21). In the CRISIS setting, the NICHD accepted as appropriate the DSMB's recommendation to stop the trial.

It is possible that in the CRISIS setting, other DSMB expert bodies may have differed in their decisions as to timing of a second interim analysis and assessment of efficacy and futility. However, it is extremely unlikely that any such variations would have led to a different conclusion regarding nonefficacy of the active CRISIS arm.

In summary, a combination of statistical rigor and maximal flexibility comes into play when an investigator designs a randomized trial with DSMB monitoring in mind.

This case study illustrates how a well-functioning DSMB provided guidance during the design phase of a clinical trial and made trial conduct decisions based on their real-time interpretation of the accumulating trial data at critical timepoints.

ACKNOWLEDGMENTS

The CRISIS DSMB members included Jeffrey R. Fineman, MD (Chair), Jeffrey Blumer, PhD, MD, Thomas P. Green, MD, and David Glidden, PhD. Additional members of the CPCCRN participating in CRISIS: Children's Hospital of Pittsburgh, Pittsburgh, PA: Joseph Carcillo, MD, Michael Bell, MD, Alan Abraham, BA, Annette Seelhorst, RN, Jennifer Jones, RN; University of Utah (Data Coordinating Center), Salt Lake City, UT: Jeri Burr, MS, RN-BC, CCRC, Amy Donaldson, MS, Angie Webster, MStat, Stephanie Bisping, RN, Teresa Liu, MPH, Brandon Jorgenson, BS, Rene Enriquez, BS, Jeff Yearley, BS; Children's National Medical Center, Washington DC: Angela Wratney, MD, Jean Reardon, BSN, RN; Children's Hospital of Michigan, Detroit, MI: Sabrina Heidemann, MD, Maureen Frey, PhD, RN; Arkansas Children's Hospital, Little Rock, AR: Parthak Prodhan, MD, Glenda Hefley, MNSc, RN; Seattle Children's Hospital, Seattle, WA: David Jardine, MD, Ruth Barker, RRT; Children's Hospital Los Angeles, Los Angeles, CA: J.

Francisco Fajardo, CLS (ASCP), RN, MD; Mattel Children's Hospital at University of California Los Angeles, Los Angeles, CA: National Institute of Child Health and Human Development, Bethesda, MD: Tammara Jenkins, MSN, RN.

REFERENCES

1. National Institutes of Health: NIH policy for data and safety monitoring of extramural clinical studies. Available at: http://grants.nih.gov/grants/guide/notice-files/not98-084.html. Accessed August 26, 2012
2. National Institutes of Health: Further guidance on a data and safety monitoring board for phase I and phase II trials. Available at: http://grants.nih.gov/grants/guide/notice-files/NOT-OD-00-038.html. Accessed August 26, 2012
3. Carcillo J, Holubkov R, Dean JM, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Collaborative Pediatric Critical Care Research Network: Rationale and design of the pediatric critical illness stress-induced immune suppression (CRISIS) prevention trial. *JPEN J Parenter Enteral Nutr* 2009; 33:368–374
4. Carcillo JA, Dean JM, Holubkov R, et al: Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Collaborative Pediatric Critical Care Research Network (CPCCRN). The randomized comparative pediatric critical illness stress-induced immune suppression (CRISIS) prevention trial. *Pediatr Crit Care Med* 2012; 13:163–173
5. Armitage P, McPherson CK, Rowe BC: Repeated significance tests on accumulating data. *J R Stat Soc Ser A* 1969; 132:235–244
6. Lan K, DeMets D. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; 70:659–663
7. Garner JS, Jarvis WR, Emori TG, et al. CDC definitions of nosocomial infections. *In*: APIC Infection Control and Applied Epidemiology: Principles and Practice. Olmsted RN (Ed). St. Louis, Mosby, 1996, pp A1–A20
8. Proschan MA, Hunsberger SA: Designed extension of studies based on conditional power. *Biometrics* 1995; 51:1315–1324
9. O'Brien PC, Fleming TR: A multiple testing procedure for clinical trials. *Biometrics* 1979; 35:549–556
10. Palestrant D, Frontera JA, Mayer SA: Treatment of massive cerebral infarction. *Curr Neurol Neurosci Rep* 2005; 5:494–502
11. Mueller PS, Montori VM, Bassler D, et al: Ethical issues in stopping randomized trials early because of apparent benefit. *Ann Intern Med* 2007; 146:878–881
12. Jennison C, Turnbull B: Group Sequential Methods With Applications to Clinical Trials. Boca Raton, Chapman and Hall, 2000
13. Mehta CR, Tsiatis AA: Flexible sample size considerations using information-based interim monitoring. *Drug Inf J* 2001; 35:1095–1112
14. Friede T, Kieser M: Sample size recalculation in internal pilot study designs: A review. *Biom J* 2006; 48:537–555
15. Wang MD: Sample size reestimation by Bayesian prediction. *Biom J* 2007; 49:365–377
16. Piantadosi S: Clinical Trials: A Methodologic Perspective. Second Edition. New York, John Wiley & Sons, 2005
17. National Heart, Lung, and Blood Institute, National Institutes of Health. NHLBI policy for data and safety monitoring of extramural clinical studies. October 2011. Available at: http://www.nhlbi.nih.gov/funding/policies/dsmpolicy.htm Accessed August 26, 2012
18. Snapinn S, Chen MG, Jiang Q, et al: Assessment of futility in clinical trials. *Pharm Stat* 2006; 5:273–281
19. Pocock SJ: Current controversies in data monitoring for clinical trials. *Clin Trials* 2006; 3:513–521
20. Meade MO: Pro/con clinical debate: It is acceptable to stop large multicentre randomized controlled trials at interim analysis for futility. Con: The hazards of stopping for futility. *Crit Care* 2005; 9:34–36
21. Schoenfeld DA: Pro/con clinical debate: It is acceptable to stop large multicentre randomized controlled trials at interim analysis for futility. Pro: Futility stopping can speed up the development of effective treatments. *Crit Care* 2005; 9:34–36