NEJM Evidence

Published June 28, 2022
NEJM Evid 2022; 1 (7)
DOI: 10.1056/EVIDctw2100020

CLINICAL TRIALS WORKSHOP | DSMB MINI-SERIES

Early Termination of Clinical Trials for Futility — Considerations for a Data and Safety Monitoring Board

Susan S. Ellenberg, Ph.D.,[1] and Pamela A. Shaw, Ph.D.[2]

Abstract

Clinical trials may be stopped for futility if there is little or no chance of demonstrating the hoped-for effect. Reasons include evidence of no treatment effect, substantial missing data that would unacceptably undermine trial conclusions, or event rates too low to support meaningful comparisons. An example of the last type of futility can be seen in the Covid-19 epidemic. ACTIV-4b (Accelerating COVID-19 Therapeutic Interventions and Vaccines) was a placebo-controlled trial testing antithrombotic agents given prophylactically to people with Covid-19 who had not yet been hospitalized. Antithrombotic agents prevent clots which were common in patients with Covid-19, but they also increase the risk of bleeding. As the trial progressed, the DSMB noted that the overall rate of thrombotic events was far too low to ever observe a treatment benefit and too low to justify the use of anticoagulant or antiplatelet therapy; only three events had been documented among the 558 participants. By making this decision, the DSMB not only protected the patients who had enrolled in the trial, they conserved community and financial resources. The authors use this and other examples to illustrate the thinking that members of the DSMB use when examining data from an ongoing trial.

Scott R. Evans, Ph.D., DSMB Mini-Series Editor

Jeffrey Drazen, M.D., Editor

Introduction

In the early 2000s, epidemiologic data indicated a strong association between a patient's "inflammatory status," as measured by the serum level of C-reactive protein, and the risk of a major adverse cardiovascular event such as myocardial infarction, stroke, and other vascular events. Was inflammation a root cause of this cardiovascular disease? To address this question, the National Institutes of Health sponsored the Cardiovascular Inflammation Reduction Trial (CIRT), assessing whether low-dose methotrexate, an anti-inflammatory agent long used in the treatment of rheumatoid arthritis, could prevent cardiovascular events in a population at elevated risk for such events.[1] The trial protocol included a statistical plan for early termination for "futility," should the emerging data

The author affiliations are listed at the end of the article.

Dr. Ellenberg can be contacted at sellenbe@pennmedicine.upenn.edu or at Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, 423 Guardian Dr., Room 611, Philadelphia, PA 19104.

indicate that the study was highly unlikely to demonstrate a benefit of methotrexate and therefore would not lead to changes in clinical practice. Before the trial reached its full enrollment, the data and safety monitoring board (DSMB)[2] overseeing the trial recommended early termination for futility because methotrexate was showing no evidence of an effect on the primary clinical outcome — or even on the inflammatory biomarkers.

This example illustrates an important issue in the design and conduct of randomized controlled trials (RCTs). An RCT is an arduous enterprise, often involving years to decades of work to plan, conduct, and analyze and requiring substantial human and financial resources; it also imparts burdens and risks on trial participants. Once a trial has started — funding obtained, protocol developed, databases prepared, supplies purchased — investigators and sponsors will be reluctant to abandon it. Nevertheless, it sometimes becomes apparent partway through the trial that there is little to no chance that the hoped-for effect will be demonstrated. Reasons could be persuasive evidence from the interim analysis of no treatment effect, as noted earlier for the CIRT trial, or logistical issues such as substantial missing data that would unacceptably undermine trial conclusions, an event rate too low to support a meaningful comparison of rates, or an inability to enroll a sufficient number of participants to provide a reliable comparison.

Let us consider another example, ACTIV-4b (Accelerating COVID-19 Therapeutic Interventions and Vaccines), a placebo-controlled trial testing antithrombotic agents given prophylactically to people with Covid-19 who had not yet been hospitalized, was terminated early for logistical reasons. The basis for this study was the recognition that thromboses were a known risk of Covid-19. The DSMB recommended that the study be terminated early for futility when it was clear that the overall rate of thrombotic events was far too low to ever observe a treatment benefit and too low to justify the use of anticoagulant or antiplatelet therapy; only three events had been documented among the 558 participants.[3] The rationale for this decision is clear, and few would disagree that the decision was appropriate.

In many trials, protection of trial participants and conservation of resources may justify early termination on grounds of futility. DSMBs for such trials assess futility to ensure that the burden on and risks to trial participants do not continue longer than necessary and for reasons of resource efficiency.

What Trials Should Incorporate Futility Assessment?

A formal assessment of futility with the possibility of early termination may be desirable in some settings but not in others. Fundamentally, trials for which a failure to show an advantage for a new treatment would not lead to changes in medical practice would be candidates for interim futility assessments.[4] Types of trials that would fall into this category are shown in Table 1. In addition, some experts have advocated futility analyses in publicly funded trials whenever possible for efficiency, lower cost, and fewer patients being recruited to failed trials.[5]

Futility assessments would not, however, be relevant for many other types of trials. These would include most trials of new treatments to relieve symptoms, because treatment effect estimates from such trials are notoriously imprecise and development programs for such drugs typically need to involve multiple trials.[6] Another type of trial that typically would not need to be considered for early termination for futility is one looking at two or more widely used therapies to see whether one has advantages over the other; some of these trials may be "noninferiority" trials, specifically addressing whether one such therapy is not too much worse than another. For these trials, the full data set would be important to support informed choices by clinicians and consumers. An example is the PRECISION (Prospective Randomized Evaluation of Celecoxib Integrated Safety versus Ibuprofen or Naproxen) trial.[7] This trial was implemented at the request of the Food and Drug Administration to determine whether one of several known effective nonsteroidal anti-inflammatory drugs was associated with greater cardiovascular risk than the others, motivated by epidemiologic studies suggesting that there might be important safety

Table 1. Trials That May Incorporate Interim Futility Analysis in Their Monitoring Plans.

Trial Type
Placebo-controlled trials of an investigational treatment for a serious medical condition
Trials comparing an investigational treatment with a standard treatment
Trials comparing a drug combination with one or more of its components
Trials comparing a higher dose with the standard dose of an available treatment
Trials involving considerable participant burden or cost

differences. Because the goal of this trial was not to establish an efficacy advantage for any of the products but rather to explore the relative safety profiles, early termination for futility would not have been a consideration; a DSMB would have considered early termination only if one of the treatments demonstrated a clear safety disadvantage.

Finally, an important type of trial for which early stopping for futility would not be advisable is a trial of a widely used product for which efficacy has never been definitively established. For such trials, a finding that benefit was unlikely might not be enough; more definitive evidence of lack of benefit might be needed. A prime example comes from the Women's Health Initiative, which studied the effect of hormone replacement therapy on cardiovascular outcomes in postmenopausal women.[8] Regulatory approval of these drugs was based on evidence from small randomized trials showing mitigation of menopausal symptoms, with most data supporting cardiovascular benefits arising from observational studies. Given that many millions of women were using this treatment on the basis of the strong belief of most gynecologists that it would prevent or delay adverse cardiovascular effects, an early finding of limited or no effect might well have provided insufficient data to dissuade providers and their patients from continued use of these products. As described by the Women's Health Initiative's DSMB, a lack of effect was clear long before the trial was ultimately terminated on the basis of harm.[9] Another example is the Testosterone Trial, which studied testosterone therapy in older men with documented subnormal testosterone levels.[10] Because this trial evaluated a widely used product that had not been well studied for many of the functional outcomes for which it was advertised, it was deemed important to collect as much data as possible; an early termination for futility would be less likely to persuade providers and consumers than would a larger database, and full safety information was particularly important given concerns about testosterone's effects on prostate and cardiovascular health. Therefore, the trial investigators did not provide a plan for stopping on the basis of futility to the trial's DSMB.

Statistical Tools a DSMB May Use to Evaluate Futility

A variety of statistical methods have been developed to evaluate futility.[11-16] In settings in which a convincing positive result at the end of the trial is needed to change clinical practice, *conditional power* is a useful tool. Conditional power is the probability of obtaining a statistically significant result if the trial were to continue to its planned completion, given the data observed thus far and some assumptions about the pattern of the data to be observed in the remainder of the study. If the conditional power falls below a prespecified threshold (commonly 10 to 20%), termination for futility may be considered.[13,14]

Conditional power can be calculated under a variety of assumptions regarding the treatment effect in the future data, such as the observed treatment effect thus far or the original hypothesized treatment effect. One might also assume a larger treatment effect than is currently being observed (e.g., the upper 90% bound of the confidence interval around the effect observed in the interim data) to investigate whether, even under optimistic assumptions, the trial would have a reasonable chance of finding a significant treatment difference. A DSMB may wish to see several estimates of conditional power based on a range of assumptions about the true treatment effect.

The following example demonstrates the considerations that might be at play when conditional power analyses are presented to a DSMB. Consider a hypothetical RCT of treatments for a serious condition that was designed to detect an improvement in 28-day survival from 25% with drug A to 50% with drug B. The trial was designed to have 77 participants per arm to ensure 90% power. Suppose at the first interim analysis, drug A had a survival rate of 5 in 20 (25%) and drug B had a survival rate of 4 in 19 (21%). Although this trend in the wrong direction is clearly consistent with random chance (Fisher's exact $P=0.80$), a DSMB may want to consider the conditional power to reach a positive result at the end of the study. To calculate conditional power, we need to hypothesize a trend for the data not yet observed. Table 2 shows how the conditional power varies depending on the assumed trend for the remaining data and the proportion of the planned sample size already collected. The second column shows the conditional power with one quarter of the data collected, under varying assumptions for the remaining data. Assuming the future data follow the assumptions used to design the trial (row 1, column 1), the conditional power to obtain a positive result in favor of drug B is still 63%. This relatively high conditional power, in the absence of other concerns, would almost surely lead a DSMB to recommend continuing the trial. Had a similar trend been observed halfway through the trial (row 1, column 2), the conditional power under the original hypothesized effect would only be 17%, sufficiently low that a DSMB might consider

Table 2. Conditional Power at Various Interim Analysis Times as a Function of the Proportion of Data Yet To Be Observed and the Assumption about the Trend in the Remaining Data.*

Treatment Difference in Remaining Data	Potential Interim Analysis (%)		
	One Quarter through, 5/20 vs. 4/19 (n=39)†	Halfway, 10/40 vs. 8/38 (n=78)	Two Thirds through, 13/52 vs. 10/50 (n=102)
Original hypothesized effect: 25% vs. 50%	63	17	<1
No difference (null effect): 25% vs. 25%	<1	<1	<1
Current trend (in wrong direction): approximately 25% vs. 20%	<1	<1	<1
Optimistic trend (stronger than hypothesized): 25% vs. 60%	91	45	5

* The table shows example data from a hypothetical randomized controlled trial of treatments for a serious condition that was designed to detect an improvement in 28-day survival from 25% for drug A to 50% for drug B. The trial was designed to have 77 participants per arm to ensure 90% power.

† The fractions given represent the proportion of information at the time of interim analysis (e.g., in the first column, the observed survival rates in group A vs. group B are 5/20 and 4/19, representing one quarter of the total planned enrollment.)

recommending termination for futility. The last row of the table shows that there would still be reasonable power for an effect larger than originally hypothesized, at least with only 25% or 50% of the data available. If there was any basis for this larger effect — for example, if another study had been done that showed a very large effect — a DSMB would likely recommend continuing the study. Had a similar negative trend been observed two thirds of the way through the trial, we see that even assuming the future data follow a stronger trend in favor of B (row 4, column 3), there is a very low probability that the trial would ultimately reach a statistically significant result in favor of drug B. In this case, it would be reasonable for a DSMB to recommend stopping the trial for futility in the absence of any other reason to continue (e.g., need for more safety data on the drug, existence of other positive studies of the drug in this population, or need for a more definitive null conclusion).

As noted, factors other than the conditional power may come into play. In the CIRT trial described earlier, the conditional power to detect the originally targeted effect size was 28%; however, the observed effect sizes in recently completed trials of anti-inflammatory agents in similar populations were much smaller than the targeted effect size in CIRT, increasing the DSMB's concern that CIRT was very unlikely to show a benefit. This information, together with the lack of effect on inflammatory markers, contributed to the DSMB's recommendation to terminate the trial for futility.

The calculation of *predicted intervals*, which predict the confidence interval that might be observed at trial's end under a given assumption about the future data,[17,18] can enhance the interpretation of the conditional power. The advantage of this approach is that it illustrates the uncertainty surrounding the magnitude of the projected effect. The comparisons of the width of the confidence interval based on observed interim data alone with the width of the predicted interval sheds light on the precision that could be gained with trial continuation, a potentially valuable tool for a DSMB. Figure 1 shows the predicted intervals for the hypothetical trial of A versus B just described, computed for the scenario in which the interim analysis was carried out halfway through the trial and assuming the hypothesized trend of a doubling of 28-day survival. In this graph, we can see that roughly 20% of the intervals are predicted to exclude the null hypothesis, and that there will be an appreciable narrowing of the confidence intervals if the trial is allowed to continue to the planned enrollment.

Risks of Stopping for Futility

Futility considerations do not inflate type I (false positive) error rates, but they do inflate type II (false negative) error, because they raise the overall possibility that a trial testing an effective intervention will erroneously come to a negative conclusion. A DSMB considering a recommendation to terminate for futility will be very conscious of this risk.

Another concern is the possibility that fully adjudicated data might present a different picture from the not fully quality-controlled interim data available to the DSMB at an interim analysis. LUME-Lung 2, a phase III trial of treatment of non–small cell lung cancer, illustrates this issue.[19,20] This trial was stopped early for futility on the basis of low conditional power (approximately 10%), but the final trial results showed a significant benefit for the novel treatment on the primary end point of progression-free survival. How did this happen? The interim analysis was based on the investigators' evaluation of disease progression, while the final analysis used the centrally adjudicated determination

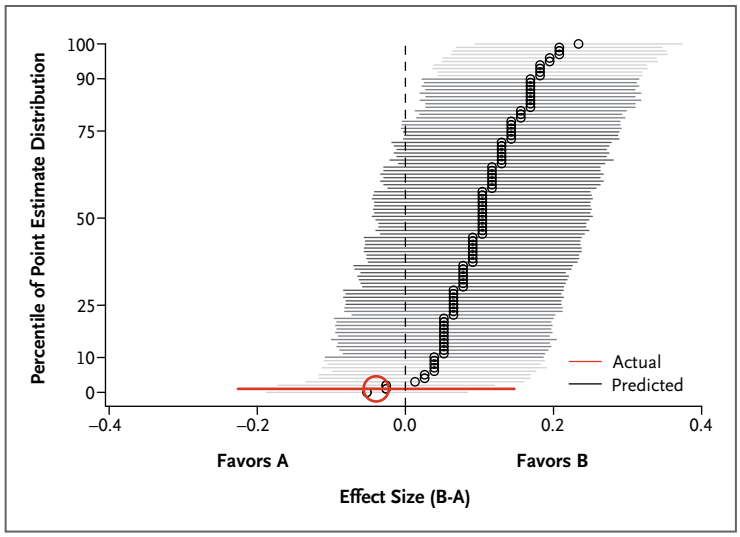

Figure 1. Trial B versus Trial A 95% Predicted Interval Plot.

Predicted confidence intervals for the final results of the hypothetical example trial as described in Table 2. This predicted interval plot derives from 100 simulations of a hypothetical trial where the data for survival were 10/40 in group A vs 8/38 in group B done at an interim analysis halfway through the trial, and the hypothesized trend of 25% mortality for group A vs 50 % mortality for group B for the yet-to-be-observed data. The observed confidence interval at the interim analysis is shown in red.

of progression. There were enough discrepancies between the investigator and central committee assessments, and additional follow-up on the enrolled participants, that the statistical comparison changed markedly for the final (primary end point) analysis. In discussing this trial, Lesaffre et al.[19] also showed the extent to which different futility monitoring criteria would have led to different decisions.

Distinguishing Futility from Harm

In settings in which negative trends emerge regarding the primary outcome, it may still be useful to continue the trial if the drug is associated with other clinical benefits, such as reduction of symptoms, so long as it is safe to do so. The Vesnarinone Trial (VEST) investigated the effect of two doses of vesnarinone versus placebo on mortality and morbidity in patients with severe heart failure.[21] Vesnarinone is an inotropic drug, a class known to improve cardiac function in patients with chronic heart failure and in some cases to improve quality of life. During the VEST trial, negative trends on mortality emerged; however, a prior trial had established benefit of vesnarinone on aspects of quality of life and mortality,[22] and the interim data in VEST showed an improved quality-of-life score. In light of this evidence of a possible benefit, the recognition of the importance of quality of life in this population, and the fact that the interim trend did not cross the Lan–DeMets boundary for harm, the DSMB allowed the trial to continue.[23]

DeMets[23] demonstrates the value of a "beta spending" (i.e., spending the probability of a false negative) approach to monitoring futility using the VEST data. This approach requires the calculation of a group-sequential monitoring boundary similar to the alpha-spending approach of DeMets and Lan[24] for monitoring efficacy or harm, but it instead allows for repeated monitoring for futility while controlling the trial's false negative rate, thereby retaining trial power despite the multiple testing. In the case of the VEST trial, the negative trend on the primary end point of death in the interim data crossed the beta-spending boundary, indicating there was less than a 2.5% chance of finding an effect the size of the prehypothesized benefit. The beta spending allows a DSMB to weigh this information, together with a monitoring boundary for harm and other trends in the data, when considering whether it is safe and informative for a trial to continue.

Additional Tools

Some DSMBs may wish to consider Bayesian approaches, which incorporate prior information about treatment effects along with the interim data, to assess the possibility of

ultimately having a positive finding.[16,25] For example, one can compute *predictive power*, which is the conditional power averaged over a range of assumptions about the treatment difference that will be observed in the future data.[16,25] In the case of the VEST trial, if one had based the prior probability distribution for treatment benefit on the relative risk and variance from the previous positive trial, the predictive power halfway through the trial would have been 28%; but a prior probability distribution that gave equal weight to a wide range of relative risks would have led to a much lower predictive power of less than 0.01%.[23] A DSMB cognizant of the uncertainties surrounding estimates from a single prior trial might be more likely to base decision-making on the latter approach to predictive power.

Revisiting Design Assumptions

An important aspect of futility analyses that a DSMB considers is whether the power calculations on which the sample size was based need to be reconsidered. Contrary to conditional power, a revised power calculation is not based on unblinded interim results, but simply revisits the assumptions used to power the trial. At the design stage of a trial, power calculations are based on assumptions regarding event rates, recruitment rates, or variability of outcome measurements that may or may not have been informed by sufficient prior information for the population and treatments under study or have been affected by environmental changes. If these assumptions vary substantially from that which is actually being observed in the trial, the actual power may be considerably less than that originally calculated. Even when a trial design has been rigorously supported by prior data, these parameters can be subject to change while the study is underway. For example, during the recent Covid-19 pandemic, many ongoing clinical trials experienced a much lower than expected recruitment rate, with many medical settings suspending or halting the screening of new patients.[26] A revised power calculation partway through a trial using updated information on important design considerations can be useful. The purpose of this calculation is to determine whether a null result at the end of the trial would be informative under the updated assumptions. If the revised power is low, a null result might not rule out the original hypothesized treatment effect, suggesting that continuing the trial may not yield a definitive result.

ACTIV-4b, described earlier, is an example of a trial that was terminated because of a much lower than anticipated event rate. An example of a trial in which a low event rate led not to termination but to an increase in sample size is an international breast cancer trial that was launched in 2013 to study the effect on disease-free survival of surgical timing during the menstrual phase.[27,28] Prior studies had suggested that adjuvant oophorectomy surgery during the luteal phase of the menstrual cycle might improve disease-free and overall survival compared with surgery during the follicular phase. Partway through the trial, concern arose that the event rate originally assumed for the placebo arm was likely too high, meaning that the trial as designed was underpowered. In cooperation with the DSMB, investigators revised the power calculations, resulting in an increase in trial size from 340 to 510.[27,28] In the end, the hypothesized benefit of adjuvant luteal phase oophorectomy could not be demonstrated.[27]

Summary

When futility analyses are specified as part of the interim monitoring plan, a DSMB must consider the totality of the interim data on primary and secondary outcomes, as well as the emerging safety data, when making a recommendation to terminate a trial for futility. A promising trend on an important secondary outcome or the desire to accumulate further safety data may warrant continuing the trial irrespective of the futility analysis for the primary outcome, assuming the safety data are acceptable. The broader context of the goals of the trial, the risk-benefit balance for patients, and whether a specific futility boundary has been crossed are important factors that a DSMB should consider before a final recommendation regarding trial continuation is made. DSMBs should also consider whether the accumulating data are consistent with those used for the initial power calculations, because a lower than expected event rate or greater than expected variability of the primary outcome will have to be accounted for in any interim calculation of power or conditional power. Futility considerations are not appropriate for every clinical trial but when used properly, futility analysis implemented by a thoughtful DSMB can improve the efficiency of clinical research, as well as strengthen the protection of trial participants.

Disclosures

Author disclosures are available at evidence.nejm.org.

Author Affiliations

[1] Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia

[2] Biostatistics Unit, Kaiser Permanente Washington Health Research Institute, Seattle

References

1. Ridker PM, Everett BM, Pradhan A, et al. Low-dose methotrexate for the prevention of atherosclerotic events. N Engl J Med 2019; 380:752-762. DOI: 10.1056/NEJMoa1809798.
2. Evans SR. Independent oversight of clinical trials through data and safety monitoring boards. NEJM Evid. 2022;1(1). DOI: 10.1056/EVIDctw2100005.
3. Connors JM, Brooks MM, Sciurba FC, et al. Effect of antithrombotic therapy on clinical outcomes in outpatients with clinically stable symptomatic COVID-19: the ACTIV-4B randomized clinical trial. JAMA 2021;326:1703-1712. DOI: 10.1001/jama.2021.17272.
4. Freidlin B, Korn EL. Monitoring for lack of benefit: a critical component of a randomized clinical trial. J Clin Oncol 2009;27:629-633. DOI: 10.1200/JCO.2008.17.8905.
5. Sully BG, Julious SA, Nicholl J. An investigation of the impact of futility analysis in publicly funded trials. Trials 2014;15:61. DOI: 10.1186/1745-6215-15-61.
6. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. Ann Intern Med 2000;133:455-463. DOI: 10.7326/0003-4819-133-6-200009190-00014.
7. Nissen SE, Yeomans ND, Solomon DH, et al. Cardiovascular safety of celecoxib, naproxen, or ibuprofen for arthritis. N Engl J Med 2016;375:2519-2529. DOI: 10.1056/NEJMoa1611593.
8. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 2002;288:321-333. DOI: 10.1001/jama.288.3.321.
9. Wittes J, Barrett-Connor E, Braunwald E, et al. Monitoring the randomized trials of the Women's Health Initiative: the experience of the data and safety monitoring board. Clin Trials 2007;4:218-234. DOI: 10.1177/1740774507079439.
10. Snyder PJ, Bhasin S, Cunningham GR, et al. Effects of testosterone treatment of older men. N Engl J Med 2016;374:611-624. DOI: 10.1056/NEJMoa1506119.
11. Herson J, Buyse M, Wittes JT. On stopping a randomized clinical trial for futility. In: Harrington, D, ed. Designs for clinical trials: perspectives on current issues. New York: Springer, 2012:109-37.
12. Lachin JM. A review of methods for futility stopping based on conditional power. Stat Med 2005;24:2747-2764. DOI: 10.1002/sim.2151.
13. Lan K, Simon R, Halperin M. Stochastically curtailed tests in long-term clinical trials. Commun Stat C 1982;1:207-219.
14. Freidlin B, Korn EL, Gray R. A general inefficacy interim monitoring rule for randomized clinical trials. Clin Trials 2010;7:197-208. DOI: 10.1177/1740774510369019.
15. Korn EL, Freidlin B. Interim monitoring for non-inferiority trials: minimizing patient exposure to inferior therapies. Ann Oncol 2018; 29:573-577. DOI: 10.1093/annonc/mdx788.
16. Snapinn S, Chen MG, Jiang Q, Koutsoukos T. Assessment of futility in clinical trials. Pharm Stat 2006;5:273-281. DOI: 10.1002/pst.216.
17. Evans SR, Li L, Wei LJ. Data monitoring in clinical trials using prediction. Drug Inf J 2007;41:733-742.
18. Li L, Evans SR, Uno H, Wei LJ. Predicted interval plots (PIPS): a graphical tool for data monitoring of clinical trials. Stat Biopharm Res 2009;1:348-355. DOI: 10.1198/sbr.2009.0041.
19. Lesaffre E, Edelman MJ, Hanna NH, et al. Statistical controversies in clinical research: futility analyses in oncology — lessons on potential pitfalls from a randomized controlled trial. Ann Oncol 2017;28:1419-1426. DOI: 10.1093/annonc/mdx042.
20. Hanna NH, Kaiser R, Sullivan RN, et al. Nintedanib plus pemetrexed versus placebo plus pemetrexed in patients with relapsed or refractory, advanced non-small cell lung cancer (LUME-Lung 2): a randomized, double-blind, phase III trial. Lung Cancer 2016;102: 65-73. DOI: 10.1016/j.lungcan.2016.10.011.
21. Cohn JN, Goldstein SO, Greenberg BH, et al. A dose-dependent increase in mortality with vesnarinone among patients with severe heart failure. N Engl J Med 1998;339:1810-1816. DOI: 10.1056/NEJM199812173392503.
22. Feldman AM, Bristow MR, Parmley WW, et al. Effects of vesnarinone on morbidity and mortality in patients with heart failure. N Engl J Med 1993;329:149-155. DOI: 10.1056/NEJM199307153290301.
23. Demets DL. Futility approaches to interim monitoring by data monitoring committees. Clin Trials 2006;3:522-529. DOI: 10.1177/1740774506073115.
24. DeMets DL, Lan KKG. Interim analysis: the alpha spending function approach. Stat Med 1994;13:1341-1352, discussion 1353–1356. DOI: 10.1002/sim.4780131308.
25. Dmitrienko A, Wang MD. Bayesian predictive approach to interim monitoring in clinical trials. Stat Med 2006;25:2178-2195. DOI: 10.1002/sim.2204.
26. Upadhaya S, Yu JX, Hodge J, Campbell J. COVID-19 impact on oncology clinical trials: a 1-year analysis. Nat Rev Drug Discov 2021;20:415. DOI: 10.1038/d41573-021-00086-8.
27. Love RR, Laudico AV, Van Dinh N, et al. Timing of adjuvant surgical oophorectomy in the menstrual cycle and disease-free and overall survival in premenopausal women with operable breast cancer. J Natl Cancer Inst 2015;107:djv064. DOI: 10.1093/jnci/djv064.
28. Hade EM, Young GS, Love RR. Follow up after sample size re-estimation in a breast cancer randomized trial for disease-free survival. Trials 2019;20:527. DOI: 10.1186/s13063-019-3632-9.