The Futility of Futility Analyses in Adjuvant Trials in Hormone Receptor–Positive Breast Cancer

Ana Elisa Lohmann, MD, PhD (ID),[1,*] Marguerite Ennis, PhD (ID),[2] Wendy R. Parulekar, MD, FRCPC,[3] Bingshu E. Chen, PhD (ID),[3] George Tomlinson, PhD (ID),[4,5] Pamela J. Goodwin, MD, FRCPC, MSc (ID)[4,6]

[1]Department of Medical Oncology, University of Western Ontario, London, Ontario, Canada; [2]Applied Statistician, Markham, Ontario, Canada; [3]Canadian Cancer Trials Group, Queen's University, Kingston, Ontario, Canada; [4]Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada; [5]Department of Medicine, University Health Network and Mount Sinai Hospital, Toronto, Ontario, Canada; and [6]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada

*Correspondence to: Ana Elisa Lohmann, MD, PhD, Department of Medical Oncology, University of Western Ontario, 800 Commissioners Rd East, PO Box 5010, Stn B, Rm A3-941, London, Ontario N6A5W9, Canada (e-mail: ana.lohmann@lhcs.on.ca).

Abstract

An interim analysis is commonly used in phase III superiority trials to compare treatment arms, with the goal of terminating exposure of patients to ineffective or unsafe drugs or to identify highly effective therapies for earlier public disclosure. Traditionally, interim analyses have been designed to identify early evidence of extremely large benefit of the experimental approach, potentially leading to early dissemination of effective treatments. Increasingly, interim analysis has also involved analysis of futility, which may lead to early termination of a trial that will not yield additional useful information. This presents an important challenge in early stage hormone receptor–positive breast cancer, where recurrence often occurs late, with a steady annual event rate up to 20 years. Early analysis of events may miss late treatment effects that can be observed only with longer follow-up. We discuss approaches to futility analysis in adjuvant clinical trials in hormone receptor–positive breast cancer, the role of the Data Safety Monitoring Committee in such analyses, considerations of the potential harms vs benefits of treatment, and the risks of continuing vs early termination of a trial.

Drug development in cancer is expensive and time consuming. To speed up this process, interim analyses have been used to identify 1) treatment benefit at the earliest possible time, allowing early adoption of effective therapies, and 2) lack of treatment benefit, allowing termination of studies of ineffective agents. The role of interim analyses for early detection of treatment benefit has been extensively addressed in prior publications (1,2). Typically, efficacy analyses of superiority trials are conducted when a target number of events has occurred (usually associated with a projected timeline, often 5 years) and interim analyses conducted when a portion of those events (eg, half or two-thirds) has occurred, often within the first 3 years after randomization. Although there is usually a strong justification for the number of events and associated timing that triggers the primary analysis, the triggers for interim analyses are often less fully justified and may not consider the timing of the analysis.

Here, we focus on the use of interim analyses of futility in adjuvant superiority trials in hormone receptor (HR)–positive breast cancer (BC), particularly those trials with time-to-event

endpoints. Some of these trials test interventions that may result in late emergence of treatment effects. The issue of late emergence of treatment benefit arises from the demonstrated efficacy of existing therapies during the first 5-10 years after diagnosis that leads to better early outcomes in comparison arms as well as the recent recognition of an ongoing risk of recurrence that persists for at least 2 decades after diagnosis (see below). Whether a treatment benefit will emerge early or late reflects many factors, including the nature of the treatment being tested, the efficacy of other treatments received by intervention and control groups, and factors that remain poorly understood.

Several statistical approaches to futility analysis have been described (3-5); all incorporate assumptions that early treatment effects will reflect later effects. The major statistical approaches are 1) frequentist analyses involving group sequential futility boundaries and 2) conditional analyses including stochastic curtailment and Bayesian methods (3,6,7). In group sequential futility tests, the analysis plan specifies multiple "looks" at the data (typically after predetermined numbers of

events) with stopping criteria designed, for the whole procedure, to maintain high power, the probability of rejecting the hypothesis of no effect when the specified alternative is true. Stochastic curtailment relies on conditional power, that is, the probability that a trial will result in statistical significance, given observed events, if it continues to its target number of events assuming the trial design alternative hypothesis for all remaining events; if the probability of statistical significance falls below prespecified criteria, the trial is stopped for futility. Bayesian predictive power uses the observed data to produce the posterior distribution of the treatment effect. The probability that the final result will meet a prespecified level of statistical certainty, assuming consistent treatment effects over time (7), is calculated over all possible study trajectories, incorporating uncertainty about treatment effects; the trial is stopped for futility if this probability falls below a prespecified level.

In adjuvant hormonal trials in HR–positive BC, there is a steady annual recurrence risk up to 20 years (Figure 1) that is dependent on initial stage (8). With current early treatment approaches, more relapses occur beyond 5 years than during the first 5 years after BC diagnosis. Thus, trials initiated at diagnosis may not show treatment effects until more than 5 years later (even if the treatment is effective); those initiated after 5 years may show benefits at variable (and unpredictable) times. As a result, the assumption that early treatment effects (ie, at the time of an interim analysis) will reflect later treatment effects may not be valid. As discussed below, the temporal pattern of emergence of treatment effects can vary considerably.

In a patient-level meta-analysis of randomized trials of 5 years of tamoxifen vs placebo, recurrence curves separated early on, between years 1 and 2 postrandomization, consistent with an early treatment effect [Figure 2 (9)]. Although the curves remained separate, the recurrence rate ratio reflected declining benefit over time: recurrence rate ratio of 0.53 (Standard Error [SE] 0.03 years 0-4), recurrence rate ratio 0.68 (SE 0.06) during years 5-9, (both 2-sided $P < .00001$), and recurrence rate ratio 0.97 (SE 0.10) during years 10-14. In contrast, in the Atlas trial that randomly assigned women who had received 5 years of tamoxifen to either placebo or an additional 5 years of tamoxifen, the benefit of tamoxifen was most pronounced after the extended treatment was completed [relative risk (RR) = 0.90, 95% confidence interval (CI) = 0.79 to 1.02; $P = .10$ years 5-9 postdiagnosis; RR = 0.75, 95% CI = 0.62 to 0.90; $P = .003$ beyond year 10; Figure 3 (10)]. This late emergence of benefit may reflect a carryover benefit of earlier tamoxifen during years 5-9 in placebo subjects.

In a patient-level meta-analysis of ovarian ablation or suppression vs no treatment, there was an improvement in recurrence-free survival (52.4% vs 46.1%; $P = .001$) and overall survival (OS = 45% vs 39%; $P = .0007$) (11); recurrence-free survival and OS curves separated between years 2 and 3 postrandomization (Figure 4) and remained separate through 15 years, consistent with an early treatment effect that was sustained over time. In contrast, in the SOFT trial, comparing tamoxifen vs tamoxifen with ovarian suppression vs exemestane with ovarian suppression for 5 years in premenopausal women, disease-free survival curves separated late, only after year 7, and the first evidence of a benefit favoring ovarian suppression was identified at 8 years (hazard ratio = 0.76, 95% CI = 0.62 to 0.93; Figure 5) (12,13).

These examples highlight the variable timing of emergence of treatment benefit across trials; benefits can even emerge after completion of study drug. They call attention to the potential pitfalls of using early absence of a treatment effect as a surrogate for later treatment effect. This may be particularly
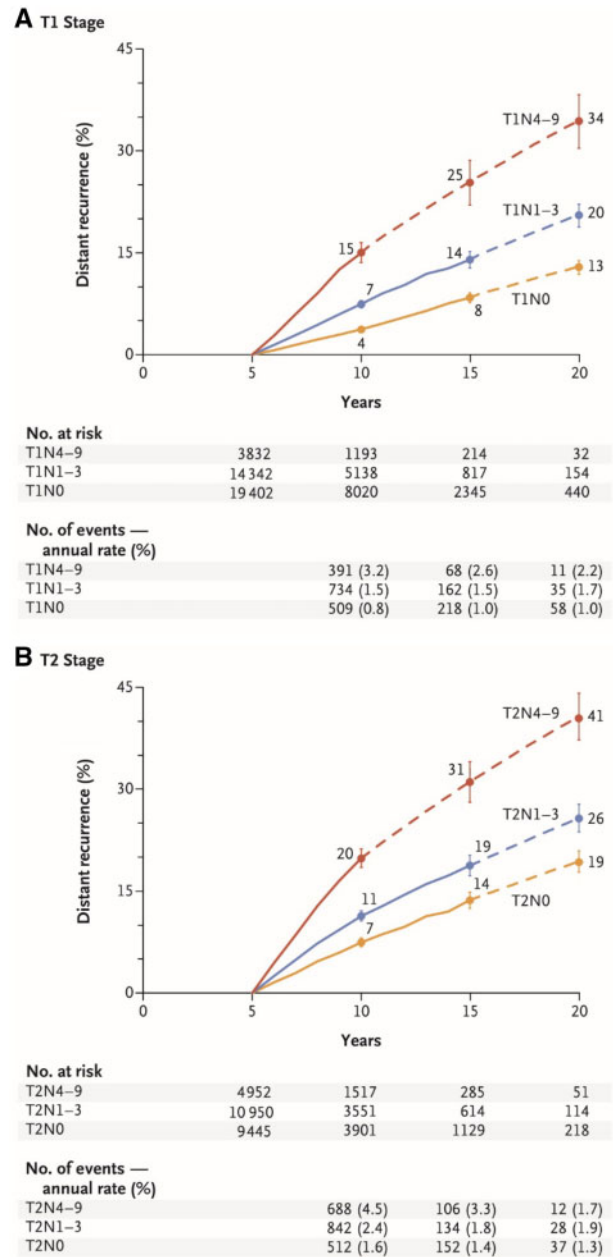

Figure 1. Recently reported data from the Oxford Overview, showing risk of distant recurrence (ignoring locoregional recurrence and contralateral BCs) out to 20 years postdiagnosis in women with ER positive BC who are free of recurrence at completion of their adjuvant hormone therapy at 5 years. Reproduced with permission from *New England Journal of Medicine* (8). BC = breast cancer; ER = estrogen receptor.

relevant in studies investigating the addition of a new treatment to an effective endocrine treatment. When treatment benefit emerges late, an early futility analysis may lead to an incorrect conclusion of futility for treatments that are ultimately shown to be effective and practice changing.

The increasingly common use of composite endpoints, such as invasive disease-free survival, that includes events with variable temporal patterns may also contribute to changing hazard ratios over time in HR–positive BC adjuvant trials. Hazards of some events that are not affected by the study treatment (potentially including non-BC deaths or new cancers in other sites) may be constant over time. When treatment effects emerge
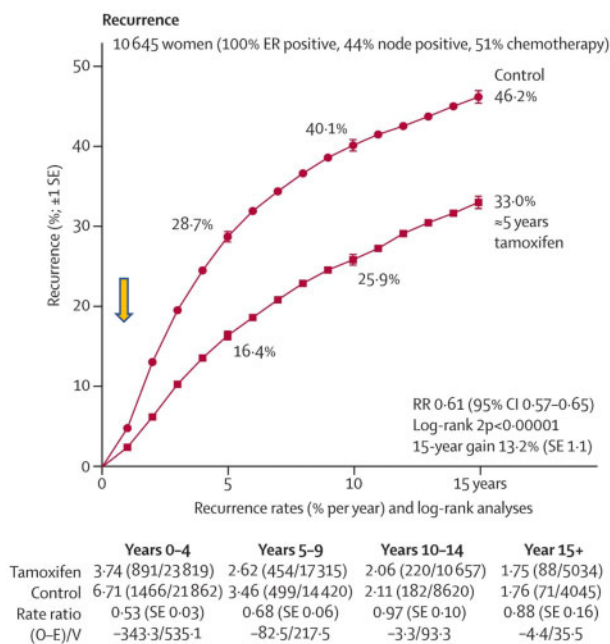
Figure 2. Results of randomized trials of endocrine therapies in the adjuvant breast cancer setting. As shown in Figures 2–5, in adjuvant endocrine trials in hormone receptor–positive breast cancer, the timing of emergence of treatment benefit has varied. Effects of approximately 5 years of tamoxifen on the 15-year probabilities of breast cancer recurrence, estrogen receptor–positive disease. The **arrow** shows the early separation of the curves. Reproduced with permission from *Lancet* (9). CI = confidence interval; ER = estrogen receptor; O–E = observed minus expected, with variance V; RR = relative risk.
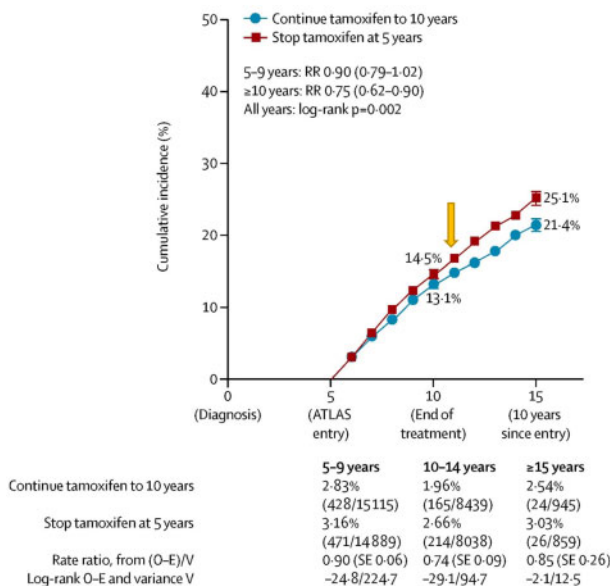
Figure 3. Breast cancer recurrence by treatment allocation for 6846 women with ER-positive disease in the ATLAS trial. Recurrence rates are percentage per year (events/patient-years of follow-up). The **arrow** shows the late separation of the curves. Reproduced with permission from *Lancet* (10). CI = confidence interval; ER = estrogen receptor; O–E = observed minus expected, with variance V; RR = relative risk.

late, the composite outcome early in a trial could primarily be made up of events not affected by study treatment, and an early futility analysis could find little treatment benefit, potentially leading to an incorrect decision to terminate the trial before treatment benefits emerged. It has recently been suggested that nonbreast primary cancers should not be included in composite outcomes in adjuvant BC trials because of concerns that

treatment effects on BC outcomes (both early and late) may be obscured (14).

Additionally, treatment effects seen early in a RCT may fluctuate and be unreliable because they are based on a small number of events. This may be exacerbated if there are random baseline imbalances in the study population or if endpoints are ascertained using nonstandardized approaches (15).
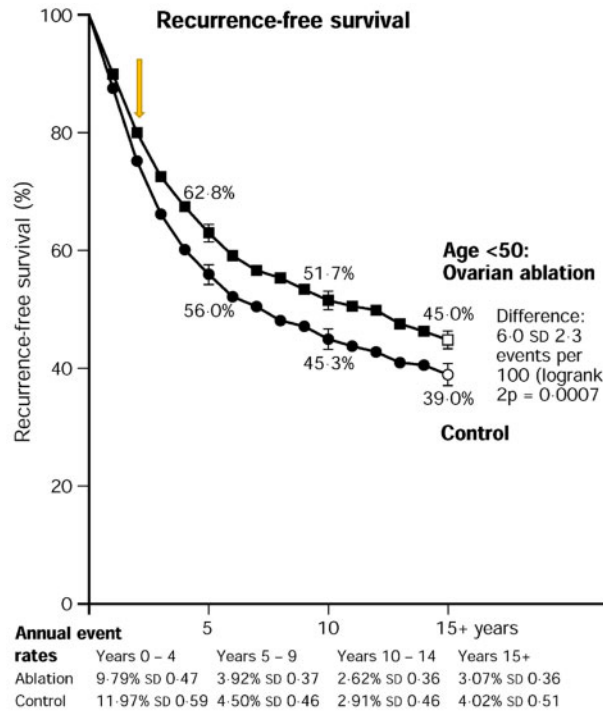
Figure 4. Recurrence-free survival for 2102 women aged younger than 50 years randomly assigned to ovarian ablation or control. The **arrow** shows early treatment benefit of ovarian suppression. Reproduced with permission from *Lancet* (11). CI = confidence interval; ER = estrogen receptor; O–E = observed minus expected, with variance V; RR = relative risk.
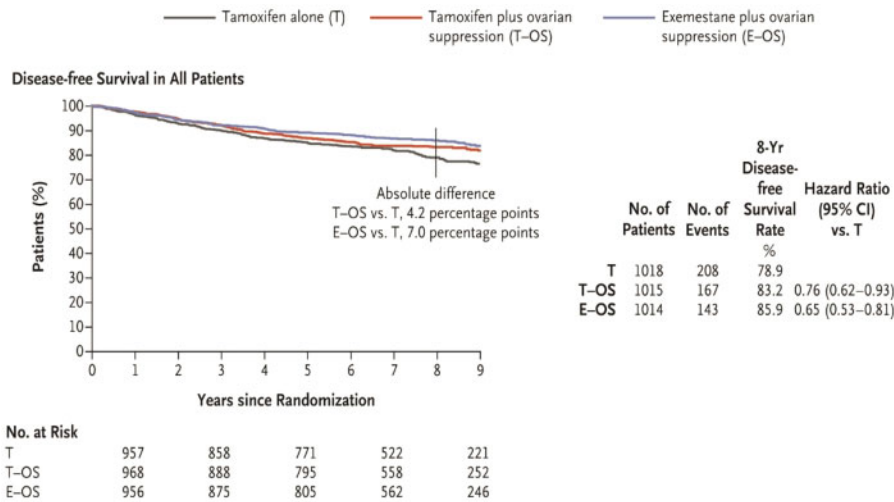

Figure 5. Disease-free survival for 3037 premenopausal women randomly assigned to tamoxifen alone, tamoxifen (T) and ovarian function suppression (OFS), or exemestane (E) and OFS in the Soft trial. Late treatment effect is observed at 8 years. Reproduced with permission from *New England Journal of Medicine* (12). CI = confidence interval; ER = estrogen receptor; O–E = observed minus expected, with variance V; RR = relative risk.

The potential risks of stopping a trial early for presumed "futility" have been explored by Jitlal et al. (16), who retrospectively conducted futility analyses, using a frequentist approach, after 25%, 50%, and 75% of events occurred in 10 randomized clinical trials in breast, lung, head and neck, and gastrointestinal cancers conducted by Cancer Research UK. Five trials had no final treatment benefit, whereas 4 had moderate and 1 had large beneficial treatment effects. Using a rule that stopped a trial when the predicted probability of success was less than 15%, 3 of 5 studies with no final benefit would have been stopped early, leading to shorter trial duration. However, 2 of the 4 trials with moderate beneficial effects seen with longer follow-up would have been erroneously stopped early thereby missing an effective treatment. Reassuringly, the trial with large treatment effect would not have been stopped early for futility at any point.

Table 1. Factors that may favor terminating or continuation of study

Factors favoring early termination of study	Factors favoring continuation of study
Severe and/or unexpected toxicity of the new therapy (eg, chemotherapy)	Low toxicity profile of the new drug (eg, hormone therapy)
Study has a long follow-up period and/or expected early treatment benefit	Study with short follow-up period and/or expected later treatment benefit to occur
Accrual is slow	Accrual complete or near complete
Single or composite endpoint in which all events likely to be impacted by the study treatment	Complex endpoint that includes events that may not be impacted by the study treatment (eg, invasive disease-free survival); these events may disproportionately contribute to a negative early futility analysis
Repeated futility analyses suggest lack of benefit	Single futility analysis
All endpoints favor control arm	Inconsistent impact of the intervention on different endpoints
Possible impact of missing data, potential bias or unusual experience in 1 or 2 study sites has been excluded	Missing information on stratification factors or early imbalances; potential impact on interim analysis has not been excluded

Thus, even when futility analyses are well designed and properly conducted from a methodological perspective, there should be thoughtful interpretation, with reference to the specific clinical context, when deciding whether their results should change trial conduct. This is particularly salient in HR–positive BC where there is increasing evidence that treatment effects may appear only after prolonged follow-up.

The interpretation of futility analyses is typically the domain of the data safety monitoring committee (DSMC) that receives the unblinded results. Ideally, the DSMC will consider a broad range of information when assessing the results of a futility analysis, including information about accrual, duration of follow-up, potential imbalances between arms, missing information, adverse events, central (vs local) review of endpoints, the breakdown of events in a composite outcome, the magnitude of the expected treatment benefit, and the natural history of the disease under consideration (Table 1). If accrual is complete and the toxicity of an ongoing intervention is minimal or the intervention has been completed on everyone, there little may be gained by terminating a trial early. Cost savings may also be minimal in this situation. If the duration of follow-up is shorter than the potential emergence of treatment benefit, the futility analysis may not reflect final treatment benefit. It may be useful to consider whether the primary endpoint includes events not likely to be affected by the intervention, as well as evaluation of the consistency of treatment effects on other outcomes (such as OS and quality of life) or across key subgroups. Ideally, investigators will specify a priori the factors the DSMC should consider in interpreting futility analyses; however, not all scenarios can be anticipated, and it is essential that the DSMC include members with sufficient expertise to optimize decision making regarding termination of the trial (17,18). Although the sponsor will often accept a DSMC recommendation after a futility analysis, additional evaluation or discussion may be undertaken.

Increasing numbers of BC adjuvant trials include futility analyses; these analyses may be required by funders. The Food and Drug Administration (19) and the European Medicine Agency provide guidance on futility analyses (20). We recommend that investigators and sponsors carefully consider the factors discussed above in deciding whether to include futility analyses in these trials and to carefully consider the timing of any futility analyses that are planned. If there is potential for late emergence of treatment benefit, as in adjuvant trials in HR–positive BC, we argue against inclusion of early futility analyses, where there may be little to be gained and potentially much to be lost. Caution is advised in incorporating such analyses without full consideration of the risks of missing future beneficial effects when trials are terminated early.

The recent discontinuation of some adjuvant trials examining the potential benefits of adding CDK 4/6 inhibitors to endocrine therapy in HR–positive BC based on early futility analyses (21,22) raises the concern that treatment benefit that might have emerged with longer follow-up was missed. At 3 years of follow-up, the MonarchE trial showed early benefit for the addition of abemaciclib, with early separation of distant relapse-free Kaplan-Meier curves (23). In contrast, in the PALLAS trial, the addition of palbociclib to endocrine therapy did not improve invasive disease-free survival, with no separation of the curves up to 4 years postrandomization (24). It is not clear yet whether these differing results are related to drug scheduling (abemaciclib was given continuously, whereas palbociclib was given intermittently), different anticancer activity of the 3 drugs, or differences in study populations, or whether longer follow-up is needed to identify an impact of palbociclib on recurrence.

In adjuvant trials in HR–positive BC, when a futility analysis suggests a treatment benefit may not be present, the risks of missing a future beneficial effect are real and must be seriously considered, particularly as study outcome events will continue to occur. Unless there is confidence that a future benefit will not be seen, we recommend that these trials be continued, even when a futility analysis suggests no treatment benefit, particularly when continuation involves little or no risk to study participants. We do not recommend that futility analyses as currently designed be incorporated into all adjuvant trials in HR–positive BC; decisions about their incorporation should reflect the issues discussed above. New analytic designs that reflect the expected timing of emergence of treatment benefit should be developed to allow more thoughtful incorporation of futility analyses into future adjuvant trials in HR–positive BC.

Funding

Notes

Disclosures: AEL and PJG have received research funding in kind from Epic Sciences. AEL has received honorarium from La Roche Posay. ME has received personal fees for consulting, paid by Mount Sinai Hospital. The other authors have no disclosures.

Authors contributions: Conceptualization: AEL, ME, GT, and PJG. Methodology: AEL, ME, WRP, BEC, GT, and PJG. Writing- Original draft: AEL, PJG. Visualization, Investigation: AEL, GT, and PJG. Supervision: PJG. Writing- Review & Editing: AEL, ME, WRP, BEC, GT, and PJG.

Data Availability

No data were used or generated for the writing of this commentary.

References

1. Fossa SD, Skovlund E. Interim analyses in clinical trials: why do we plan them? *J Clin Oncol.* 2000;18(24):4007–4008.
2. Wayant C, Vassar M. A comparison of matched interim analysis publications and final analysis publications in oncology clinical trials. *Ann Oncol.* 2018; 29(12):2384–2390.
3. Dmitrienko A, Wang MD. Bayesian predictive approach to interim monitoring in clinical trials. *Stat Med.* 2006;25(13):2178–2195.
4. Gordon Lan KK, Demets DL. Discrete sequential boundaries for clinical trials. *Biometrika.* 1983;70(3):659–663.
5. Gordon Lan KK, Simon R, Halperin M. Stochastically curtailed tests in long-term clinical trials. *Commun Stat Part C: Sequential Analysis.* 1982;1(3):207–219.
6. Harrington D. *Designs for Clinical Trials: Perspectives on Current Issues.* 1st ed. New York: Springer; 2011.
7. Snapinn S, Chen M-G, Jiang Q, Koutsoukos T. Assessment of futility in clinical trials. *Pharm Stat.* 2006;5(4):273–281.
8. Pan H, Gray R, Braybrooke J, et al.; for the EBCTCG. 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. *N Engl J Med.* 2017;377(19):1836–1846.
9. Davies C, Godwin J, et al.; for the Early Breast Cancer Trialists' Collaborative Group. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet.* 2011;378(9793):771–784.
10. Davies C, Pan H, Godwin J, et al.; for the Adjuvant Tamoxifen: Longer Against Shorter (ATLAS) Collaborative Group. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet.* 2013;381(9869):805–816.
11. The Early breast Cancer Trialists' Collaborative Group. Ovarian ablation in early breast cancer: overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. *Lancet.* 1996;348(9036):1189–1196.
12. Francis PA, Pagani O, Fleming GF, et al.; for the SOFT and TEXT Investigators and the International Breast Cancer Study Group. Tailoring adjuvant endocrine therapy for premenopausal breast cancer. *N Engl J Med.* 2018;379(2): 122–137.
13. Francis PA, Regan MM, Fleming GF, et al.; for the International Breast Cancer Study Group. Adjuvant ovarian suppression in premenopausal breast cancer. *N Engl J Med.* 2015;372(5):436–446.
14. Tolaney SM, Garrett-Mayer E, White J, et al. Updated Standardized Definitions for Efficacy End Points (STEEP) in adjuvant breast cancer clinical trials: STEEP Version 2.0. *J Clin Oncol.* 2021;39(24):2720–1200.
15. Lesaffre E, Edelman MJ, Hanna NH, et al. Statistical controversies in clinical research: futility analyses in oncology-lessons from potential pitfalls from a randomized controlled trial. *Ann Oncol.* 2017;28(7):1419–1426.
16. Jitlal M, Khan I, Lee SM, et al. Stopping clinical trials early for futility: retrospective analysis of several randomised clinical studies. *Br J Cancer.* 2012; 107(6):910–917.
17. Slutsky AS, Lavery JV. Data safety and monitoring boards. *N Engl J Med.* 2004; 350(11):1143–1147.
18. Fleming TR, DeMets DL, Roe MT, et al. Data monitoring committees: promoting best practices to address emerging challenges. *Clin Trials.* 2017;14(2): 115–123.
19. Guidance for clinical trial sponsors: establishment and operation of clinical trial data monitoring committees [excerpts]. *Biotechnol Law Rep.* 2012; 31(3):3.
20. Committee for Medicinal Products for Human Use, Efficacy Working Party, Committee for Release for Consultation. Committee for Medicinal Products for Human Use (CHMP) guideline on the choice of the non-inferiority margin. *Stat Med.* 2006;25(10):1628–1638.
21. Mayer EL, Dueck AC, Martin M, et al. Palbociclib with adjuvant endocrine therapy in early breast cancer (PALLAS): interim analysis of a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol.* 2021;22(2):212–222.
22. Loibl S, Marme F, Martin M, et al. Palbociclib for residual high-risk invasive HR-positive and HER2-negative early breast cancer–the Penelope-B trial. *J Clin Oncol.* 2021;39(14):1518–1530.
23. Harbeck N, Rastogi P, Martin M, et al. Adjuvant abemaciclib combined with endocrine therapy for high-risk early breast cancer: updated efficacy and Ki-67 analysis from the MonarchE study. *Ann Oncol.* 2021;32(12):1571–1581.
24. Gnant M, Dueck AC, Frantal S, et al.; for the PALLAS groups and investigators. Adjuvant Palbociclib for Early Breast Cancer: the PALLAS trial results (ABCSG-42/AFT-05/BIG-14-03). *J Clin Oncol.* 2022;40(3):282–293.