THE CHANGING FACE OF CLINICAL TRIALS
Jeffrey M. Drazen, M.D., David P. Harrington, Ph.D., John J.V. McMurray, M.D., James H. Ware, Ph.D., and
Janet Woodcock, M.D., *Editors*

Adaptive Designs for Clinical Trials

Deepak L. Bhatt, M.D., M.P.H., and Cyrus Mehta, Ph.D.

RANDOMIZED CLINICAL TRIALS SERVE AS THE STANDARD FOR CLINICAL research and have contributed immensely to advances in patient care. Nevertheless, several shortcomings of randomized clinical trials have been noted, including the need for a large sample size and long study duration, the lack of power to evaluate efficacy overall or in important subgroups, and cost. These and other limitations have been widely acknowledged as limiting medical innovation.[1] Adaptive trial design has been proposed as a means to increase the efficiency of randomized clinical trials, potentially benefiting trial participants and future patients while reducing costs and enhancing the likelihood of finding a true benefit, if one exists, of the therapy being studied.[2]

Adaptive designs are applicable to both exploratory and confirmatory clinical trials. Adaptive designs for exploratory clinical trials deal mainly with finding safe and effective doses or with dose–response modeling. The emphasis is on strategies that will assign a larger proportion of the participants to treatment groups that are performing well, reduce the number of participants in treatment groups that are performing poorly, and investigate a dose range that is larger than ranges in corresponding trials with nonadaptive designs, in order to select effective doses for the confirmatory stage of investigation. Control of the type I error rate is less of an issue. In Table 1, various types of adaptive designs for exploratory clinical trials are classified into categories that reflect the time sequence in which they would be performed in the drug-development process.

In confirmatory trials, the adaptive nomenclature refers to making prospectively planned changes to the future course of an ongoing trial on the basis of an analysis of accumulating data from the trial itself, in a fully blinded or unblinded manner, without undermining the statistical validity of the conclusions.[3] However, modifications of randomized clinical trials that are performed in an unblinded manner are subject to closer regulatory scrutiny than those performed in a blinded manner. They require careful attention to statistical techniques and operational procedures to ensure that the implementation is scientific, ethical, and free from bias. In Table 1, different types of adaptations for confirmatory trials are classified into four major categories — seamless phase 2–3 designs, sample-size reestimation, group sequential designs, and population-enrichment designs — and the strengths and weaknesses of each type are identified in relation to corresponding nonadaptive designs. There is some overlap among the different categories. For example, sample-size reestimation could be implemented on its own or incorporated into group sequential, dose-selection, or population-enrichment designs.

In this review, we focus on adaptive designs of confirmatory clinical trials. We discuss the benefits and limitations of such designs, using four case studies that highlight the statistical and operational considerations that are the prerequisites

From Brigham and Women's Hospital Heart and Vascular Center and Harvard Medical School (D.L.B.) and Harvard T.H. Chan School of Public Health (C.M.), Boston, and Cytel, Cambridge (C.M.) — all in Massachusetts. Address reprint requests to Dr. Bhatt at Brigham and Women's Hospital Heart and Vascular Center, 75 Francis St., Boston, MA 02115, or at dlbhattmd@post.harvard.edu.

Table 1. Types of Adaptive Designs.*

Stage of Development and Design Type	Strength	Weakness
First-in-human, phase 1 design with goal of establishing the MTD		
Single ascending dose or multiple ascending doses	Establishes MTD and biologic activity	Uses larger cohort sizes at potentially safe doses
Dose-escalation, 3+3, continual reassessment method, Bayesian logistic-regression method, or modified toxicity probability interval design	Provides more accurate estimate of MTD with smaller cohort size	May yield more cases of toxic events than design with single ascending dose or multiple ascending doses
Phase 2 design, with the goal of establishing efficacy and choosing doses for phase 3 trial		
Fixed-sample design, traditional proof-of-concept design (MTD vs. placebo), or dose-ranging design (MTD, placebo, and intermediate doses)	Is simple to implement and easy to design	Has less precision than adaptive design
Adaptive design with proof of concept, early stopping and sample-size reestimation, dose ranging with selection at interim analysis, or dose ranging with frequent Bayesian adaptation of randomization ratios	Yields more precise estimates for same sample size	Is more complex to implement; requires more lead time to set up; operating characteristics determined only by simulation
Seamless phase 2–3 design		
Operationally seamless	Eliminates time between phase 2 and phase 3; permits sponsor involvement for dose selection at the end of phase 2; uses conventional final analysis, parameter estimates, and confidence intervals	Has a final analysis based only on data from phase 3
Inferentially seamless	Combines data from both phases for final analysis	Has no sponsor involvement in dose selection at end of phase 2; has risk of inadequate dose–response modeling; has a complex final analysis involving closed testing; uses nonconventional parameter estimates and confidence intervals; data from the two phases may not be homogeneous
Sample-size reestimation		
With blinded data	Uses conventional final analysis; has fewer regulatory hurdles	Allows sample-size adjustments due to unknown variance only
With unblinded data	Allows sample-size adjustments due to unknown treatment effect or unknown variance; can determine sample size after review of actual data from the trial instead of from pilot studies	Interim estimate of treatment effect can be misleading; requires strict firewalls to prevent leakage of information about adaptive rules or decisions; potential for operational bias if investigator behavior changes; requires meticulous up-front planning; uses nonconventional final analysis with prespecified weighting of the cohorts before and after sample-size reestimation; may face regulatory hurdle

Design	Advantages	Limitations
Group sequential design		
Classic	Enables early stopping for efficacy, futility, or harm; has flexible alpha spending functions; can alter maximum sample size in a blinded manner	Cannot alter maximum sample size or events in an unblinded manner; uses nonconventional parameter estimates and confidence intervals; if trial terminates early for efficacy, overruns pose risk of downturn from significance to nonsignificance; greater burden on data and safety monitoring committee to review totality of evidence before premature termination[†]
Adaptive	Includes all advantages of classic group sequential designs; can alter maximum sample size in an unblinded manner; can switch end point from noninferiority to superiority; can alter number and spacing of interim analyses, and alpha spending function, on the basis of unblinded interim analysis; overruns are not a problem since trial proceeds to completion with increased enrollment and resolution of responses in all patients instead of being terminated early with risk of downturn from unknown or unadjudicated responses	Includes all the disadvantages of sample-size reestimation with unblinded data; uses nonconventional parameter estimates and confidence intervals
Population-enrichment design	Can eliminate nonperforming subgroups at interim analysis if treatment is effective in selected subgroups only	Must prospectively identify which subgroups to target; may eliminate subgroups in which treatment is effective; loses power as number of targeted subgroups increases; loses power if there is low prevalence of effective subgroups; biomarker cutoff points for subgroup partitioning not known

* MTD denotes maximum tolerated dose.
† In an overrun situation, patients for whom the primary end points are unknown (because of unadjudicated data or delayed response) at the time of an early-termination decision will be included in the final analysis.

for a successful trial. The statistical methods for hypothesis testing and parameter estimation are provided in the Supplementary Appendix, available with the full text of this article at NEJM.org.

FOUR CASE STUDIES

SEAMLESS PHASE 2–3 DESIGN — THE INHANCE TRIAL

The Indacaterol to Help Achieve New COPD Treatment Excellence (INHANCE) trial was an adaptive two-stage (i.e., phase 2–3), confirmatory, randomized clinical trial of inhaled indacaterol, a once-daily long-acting beta$_2$-agonist bronchodilator for the treatment of chronic obstructive pulmonary disease (COPD); the trial featured multiple treatment groups, with dose selection at the end of stage 1.[4,5] In stage 1, patients with COPD were randomly assigned in a double-blind, double-dummy manner to one of seven groups to receive four doses of indacaterol, placebo, formoterol, or tiotropium; the last two regimens were considered to be standard-of-care comparators. Two of the four indacaterol doses were to be selected for further testing at stage 2 along with placebo and tiotropium. The final analysis would be based on the combined data from the two stages.

The primary efficacy objective was to show the superiority of at least one dose of indacaterol over placebo at week 12 with respect to the 24-hour postdose (trough) forced expiratory volume in 1 second (FEV$_1$). Although the final efficacy analysis was to use the FEV$_1$ data through week 12, the dose selection at the interim analysis was to be based on data from patients who had been treated through week 2 only, since indacaterol is known to reach pharmacodynamic steady state within 2 weeks.

The two most important statistical considerations for a design of this type are the dose-selection rule at the interim analysis and the statistical inference at the final analysis. The dose selection would have to be made by an external data and safety monitoring committee that had been equipped with clear, unambiguous decision rules for determining which doses to pick and also some flexibility to deviate from these rules in case of unexpected safety signals or a lack of dose response (see the Supplementary Appendix). Accordingly, a rather complex set of decision rules covering all anticipated contingencies was included in the charter for the data and safety monitoring

committee (Table S1 in the Supplementary Appendix).[6] The sections on Statistical Methodology in the Supplementary Appendix describe how the type I error is controlled when ineffective doses might be dropped at the end of stage 1 and multiple doses might be compared with a common control group in the final analysis.

In the INHANCE trial, the interim analysis was to be performed when 770 patients (110 patients per group) had completed at least 2 weeks of treatment (Fig. S1 in the Supplementary Appendix). On the basis of the detailed dose-selection guidelines that had been prespecified in the charter, the data and safety monitoring committee selected doses of 150 μg and 300 μg, whereupon the recruitment of patients was immediately resumed for the second stage of the trial. The final analysis was performed when 285 additional patients had been enrolled and evaluated. The difference between each indacaterol dose and either placebo or tiotropium was significant with respect to the primary and key secondary end points.[5]

This example shows several conditions that are essential for the successful implementation of an adaptive design. First, the highly quantitative, precise, and easily obtained early readout of end-point data made it possible to eliminate two of the trial groups quickly and thereby enroll many more patients in study groups that were receiving the doses and treatments of primary interest. Trials that require rapid recruitment or lengthy or complex patient follow-up (e.g., assessment of freedom from a heart attack over a period of a few years after treatment) may not be suitable for adaptive designs, since enrollment may be almost complete by the time the stage 1 cohort has met its follow-up requirements for decision making. Second, the preliminary planning for this trial was meticulous, with detailed dose-selection criteria, a communication plan for disseminating interim results that would not unblind the interim results, a hypothesis-testing strategy that controlled the type I error, and detailed simulations of the operating characteristics before the initiation of the trial.

Although a nonadaptive approach would have the advantage that the sponsor could be fully involved in the selection of the doses for follow-on phase 3 testing, the adaptive design combined the data from the two stages for the final analysis, which meant that the trial required fewer patients and had a shorter overall duration. This gain in efficiency, however, carried the risk that the totality of evidence at the end of the trial might not support a regulatory submission, possibly because of inadequate dose–response modeling or an inadequate safety profile. For this reason, extensive up-front planning and a thorough discussion by the trial team of all possible contingencies that might arise over the course of the two stages of the trial contributed to the success of the INHANCE trial.

SAMPLE-SIZE REESTIMATION — THE CHAMPION PHOENIX TRIAL

The Cangrelor versus Standard Therapy to Achieve Optimal Management of Platelet Inhibition (CHAMPION) PHOENIX trial was a double-blind, placebo-controlled trial in which patients who were undergoing urgent or elective percutaneous coronary intervention (PCI) for coronary insufficiency were randomly assigned to receive a bolus and infusion of the intravenous antiplatelet agent cangrelor or a loading dose of the oral antiplatelet agent clopidogrel.[7] The primary efficacy end point was a composite of death, myocardial infarction, ischemia-driven revascularization, or stent thrombosis within 48 hours after PCI. The initially planned enrollment of 10,900 patients, with possible early stopping for efficacy on the basis of a gamma (–5) alpha spending function (which generates group-sequential boundaries that resemble the O'Brien–Fleming boundaries) when 70% of the patients had been enrolled, provided the study with 86% power to detect a 24% lower relative risk, from an event rate of 5.1% in the control group to an event rate of 3.9% in the experimental-therapy group. However, small variations in the assumed magnitude of the difference in relative risk on the event rate in the control group could have led to a substantial reduction in power at the design stage (Table 2).

To mitigate this risk, the trial permitted a possible sample-size reestimation at the interim analysis when 70% of the patients had been enrolled. The sample space of possible outcomes at this interim analysis was partitioned into three zones on the basis of the observed percentage lowering in relative risk — unfavorable zone (observed difference, <13.6%), promising zone (≥13.6% to ≤21.2%), and favorable zone (>21.2%).[8] If the observed percentage lowering in relative risk fell in the promising zone, there would be an increase

Table 2. Comparison of Design-Stage and Interim Analysis–Stage Operating Characteristics of an Adaptive Trial.*

Design-Stage Operating Characteristics				Interim Analysis–Stage Operating Characteristics					
Event Rate in Control Group	Percentage Lowering of Event Rate with Experimental Therapy	Power	Average Sample Size	Zone†	Probability of Entering Zone	Conditional Power‡		Average Sample Size	
						adaptive	nonadaptive	adaptive	nonadaptive
	%		no.			%		no.	
5.1	24	86	9,231	Unfavorable	0.12	33	33	10,900	10,900
				Promising	0.25	97	81	16,726	10,900
				Favorable§	0.63	99	99	8,248	8,248
5.1	21	76	9,668	Unfavorable	0.21	24	24	10,900	10,900
				Promising	0.30	95	73	17,057	10,900
				Favorable	0.49	98	98	8,417	8,417
5.1	18	62	10,068	Unfavorable	0.32	17	17	10,900	10,900
				Promising	0.32	90	66	17,373	10,900
				Favorable	0.36	97	97	8,590	8,590
4.75	24	84	9,365	Unfavorable	0.14	31	31	10,900	10,900
				Promising	0.26	97	80	16,813	10,900
				Favorable	0.60	99	99	8,316	8,299
4.75	21	73	9,778	Unfavorable	0.23	23	23	10,900	10,900
				Promising	0.31	94	72	17,137	10,900
				Favorable	0.46	98	98	8,479	8,466
4.75	18	59	10,132	Unfavorable	0.34	16	16	10,900	10,900
				Promising	0.32	89	64	17,397	10,900
				Favorable	0.34	97	97	8,633	8,636

* The operating characteristics at the interim analysis stage of an adaptive trial such as the Cangrelor versus Standard Therapy to Achieve Optimal Management of Platelet Inhibition (CHAMPION) PHOENIX trial[7] are based on 100,000 simulated trials. For example, in a trial with an expected event rate in the control group of 5.1%, if the event rate in the experimental-therapy group is lower by 18%, the overall (or unconditional) power of the original design decreases to 62%. If, however, the interim results enter the promising zone, the conditional power can be boosted from 66% to 90% by an adaptive increase in the sample size. Outside the promising zone, the operating characteristics of the adaptive and nonadaptive designs are identical.

† In the favorable and unfavorable zones, the sample size would not change; the probability of achieving statistical significance in the favorable zone would already be very high, making a sample-size increase unnecessary, whereas the probability of achieving statistical significance in the unfavorable zone would be too low to make an increase in sample size worthwhile. In the promising zone, there could be a substantial benefit from increasing the sample size.

‡ Conditional power is defined as the probability of achieving statistical significance conditional on the interim results and on the estimated reduction in the event rate in the experimental-therapy group.

§ The average sample sizes in the favorable zone are less than 10,900 because some of the simulated trials that landed in the favorable zone actually crossed the efficacy boundary for early termination of the trial.

in the sample size according to a prespecified formula. In the favorable or unfavorable zones, there would be no change in the sample size because the probability of achieving statistical significance under the current observed difference in relative risk would already be very high in the favorable zone, whereas in the unfavorable zone it would be too low to make an increase in sample size worthwhile. In the promising zone, however, there could be a substantial benefit from increasing the sample size.

For example, if the control group had an event rate of 5.1% and the experimental-therapy group had a relative risk that was lower by only 18%, the overall power would be reduced to 62% (Table 2). However, if an adaptive design were

implemented, then the power, which was conditional on falling inside the promising zone at the interim analysis, could be boosted from 66% to 90% by increasing the sample size from 10,900 to an average of 17,373.

The advantage of this approach is that the sample size is only increased after the interim results have been reviewed and observed to be promising (in this case, by the data and safety monitoring committee). This is the major innovation of the adaptive group sequential design as compared with the classic group sequential design, in which the maximum amount of statistical information (in this case, sample size) is fixed at the design stage and there is no flexibility to alter it on the basis of results observed at the interim analysis. Figure S2 in the Supplementary Appendix shows a detailed comparison between the operating characteristics of the adaptive design that was used in the CHAMPION PHOENIX trial and that of a competing group sequential strategy that used the same expected sample size over a range of clinically meaningful values for the difference in relative risk. In exchange for a small loss of overall power, the adaptive design provides a substantial gain in conditional power if the interim results are promising. Control of the type I error for this type of adaptive design is discussed in the sections on Statistical Methodology in the Supplementary Appendix, as well as in Figure 1 (and see the interactive graphic, available at NEJM.org).

An interactive graphic is available at NEJM.org

In the CHAMPION PHOENIX trial, the results fell in the favorable zone at the interim analysis, and the sample size was not increased. The final analysis showed statistical significance in favor of cangrelor. On the basis of the results of this trial, regulatory agencies in the United States and the European Union approved cangrelor for use in patients who undergo PCI.

CHANGING THE PRIMARY END POINT — THE EXAMINE TRIAL

Before any new antihyperglycemic agent can gain full regulatory approval in the United States, it must be shown to have no association with an unacceptable risk of major adverse cardiovascular events. The specific guidance is that the upper boundary of the two-sided repeated 95% confidence interval for the hazard ratio for major adverse cardiovascular events should not exceed 1.3 in the time-to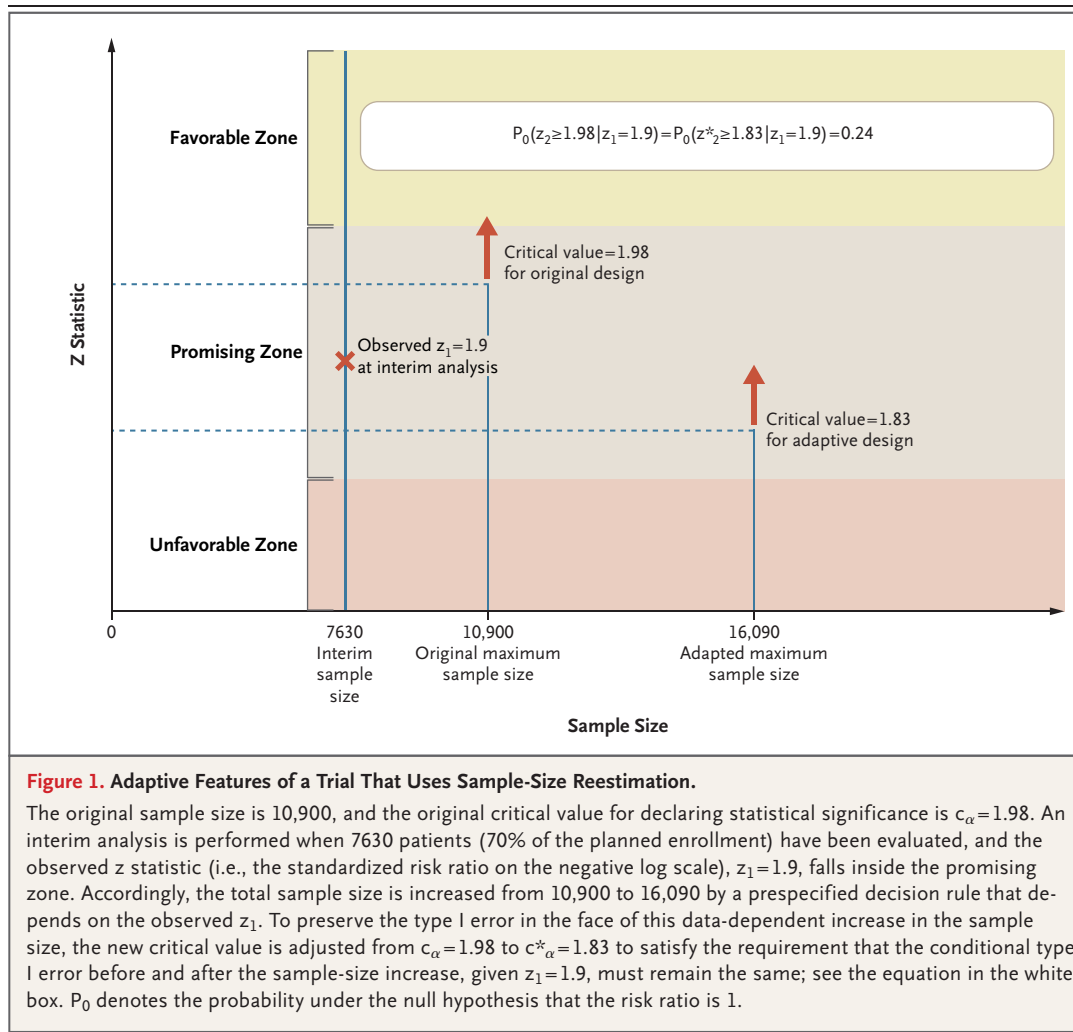-event analysis in a prospective phase 3 noninferiority trial of the new agent versus standard of care. The Examination of Cardiovascular Outcomes with Alogliptin versus Standard of Care (EXAMINE) trial was such a cardiovascular-outcome trial of alogliptin, a dipeptidyl peptidase 4 inhibitor.[9] The trial enrolled 5380 patients with a median follow-up of 18 months and showed noninferiority by obtaining an upper boundary of the confidence interval of 1.16.

Had the upper boundary of the confidence interval been less than 1, the trial would have shown superiority. That is, the trial would have shown that the new agent was protective instead of merely ruling out an unacceptable increase in cardiovascular risk.[10-12] Table S2 in the Supplementary Appendix shows the sample size that would be needed for a cardiovascular-outcome trial to have 90% power to show superiority over a range of hazard ratios. For example, even in the case of a drug with a favorable hazard ratio of 0.85 and an annualized event rate of 2.5%, a trial would require enrollment of almost 18,000 patients over a period of 2 years and an additional 3 years of follow-up. In this context, an adaptive design can generate the best possible estimate of the required sample size, since the actual interim results from the trial itself could be used to repower the trial for superiority. The EXAMINE trial had prespecified that the maximum number of adjudicated major adverse cardiovascular events would be 650, with a planned interim analysis after 550 events and an option to stop the trial and claim noninferiority if the P value for the between-group comparison was less than 0.001.

The trial design included one additional feature. The trial could proceed all the way to 650 events even though the early-stopping boundary for claiming noninferiority was crossed, provided that the conditional power or probability to show superiority by the end of the trial under the current trend exceeded 20%. This feature gave the sponsor a second chance to claim superiority. Since the primary analysis of the noninferiority hypothesis was prespecified to be performed in the intention-to-treat population, the change of goal from noninferiority to superiority would not entail a change of population. However, with only 20% conditional power and no option to increase the total number of adjudicated events beyond 650, the chances of actually claiming superiority were low. This design could have been improved by the inclusion of the adaptive option

Figure 1. Adaptive Features of a Trial That Uses Sample-Size Reestimation.

The original sample size is 10,900, and the original critical value for declaring statistical significance is $c_\alpha = 1.98$. An interim analysis is performed when 7630 patients (70% of the planned enrollment) have been evaluated, and the observed z statistic (i.e., the standardized risk ratio on the negative log scale), $z_1 = 1.9$, falls inside the promising zone. Accordingly, the total sample size is increased from 10,900 to 16,090 by a prespecified decision rule that depends on the observed z_1. To preserve the type I error in the face of this data-dependent increase in the sample size, the new critical value is adjusted from $c_\alpha = 1.98$ to $c^*_\alpha = 1.83$ to satisfy the requirement that the conditional type I error before and after the sample-size increase, given $z_1 = 1.9$, must remain the same; see the equation in the white box. P_0 denotes the probability under the null hypothesis that the risk ratio is 1.

to increase the required number of events for the final analysis if the noninferiority boundary were crossed at the interim analysis and the conditional power for claiming superiority were sufficiently high (Fig. 2).

Table S3 in the Supplementary Appendix shows the operating characteristics of the design. By doubling the required number of events in the promising zone, the chances of showing superiority increase from 64% to 96% if the true hazard ratio is 0.85. This dramatic increase in power would come at the cost of prolonging the trial by 1 year. In the EXAMINE trial, the early-stopping boundary for noninferiority was crossed after 550 events, but the conditional power for claiming superiority was less than 20%. Thus, the trial was stopped; this decision allowed the sponsor to file a claim of noninferiority without extend-

ing the trial for an additional year with a slim chance of being able to show superiority.

BIOMARKER-DRIVEN ADAPTIVE POPULATION-ENRICHMENT DESIGNS

It has become increasingly apparent that treatment effects can differ greatly among subgroups of patients with different genetic or biomarker characteristics. Table S4 in the Supplementary Appendix lists several targeted therapeutic agents that have been approved in the United States for specific subgroups of patients. These examples show the potential of predictive biomarkers to identify patients who are likely to benefit from targeted therapies and to thereby increase the success rate of confirmatory clinical trials. In these examples, we have focused on oncology trials, but the use of this approach will probably increase
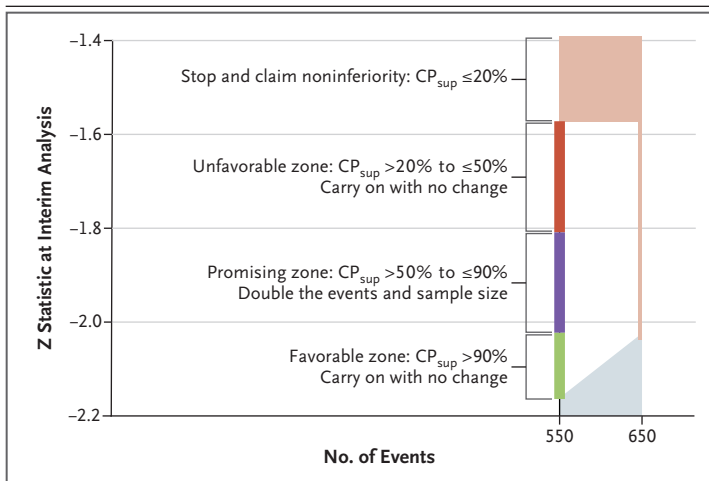
Figure 2. Adaptive Design of a Cardiovascular-Outcome Trial with Zones for Decision Making Regarding Superiority.

If the efficacy boundary for claiming noninferiority is crossed at the interim analysis, when 550 events have occurred, the region for continuing the trial is partitioned into four zones, on the basis of the conditional power for claiming superiority (CP_{sup}) in the final analysis with 650 events. Depending on the zone into which the interim result falls, the trial is either terminated immediately with a noninferiority claim or is continued — with or without an adaptive increase in the number of events — in the hopes of claiming superiority at the final analysis. The light blue area represents the efficacy zone, and the light red area the futility zone for claiming superiority. The light red bar extends down to the blue area because the efficacy and futility boundaries have to meet in the final analysis so that a decision can be made.

in other fields as validated biomarkers that predict response or lack of response to therapy emerge (see the Supplementary Appendix).[13]

However, most previous studies in which biomarkers have shown predictive capabilities were not designed for this purpose. Even in well-controlled phase 3 trials, the biomarker component of the analysis is often performed retrospectively or the trials restricted enrollment to the targeted subgroups from the start. However, the Food and Drug Administration guidance regarding enrichment strategies for clinical trials recommends that even in cases in which there is a strong biologic basis for a therapy to target a particular genetic marker, it is desirable to enroll patients in whom the marker is absent in order to show sensitivity in patients who have the marker and lack of sensitivity in patients who do not have the marker.[14]

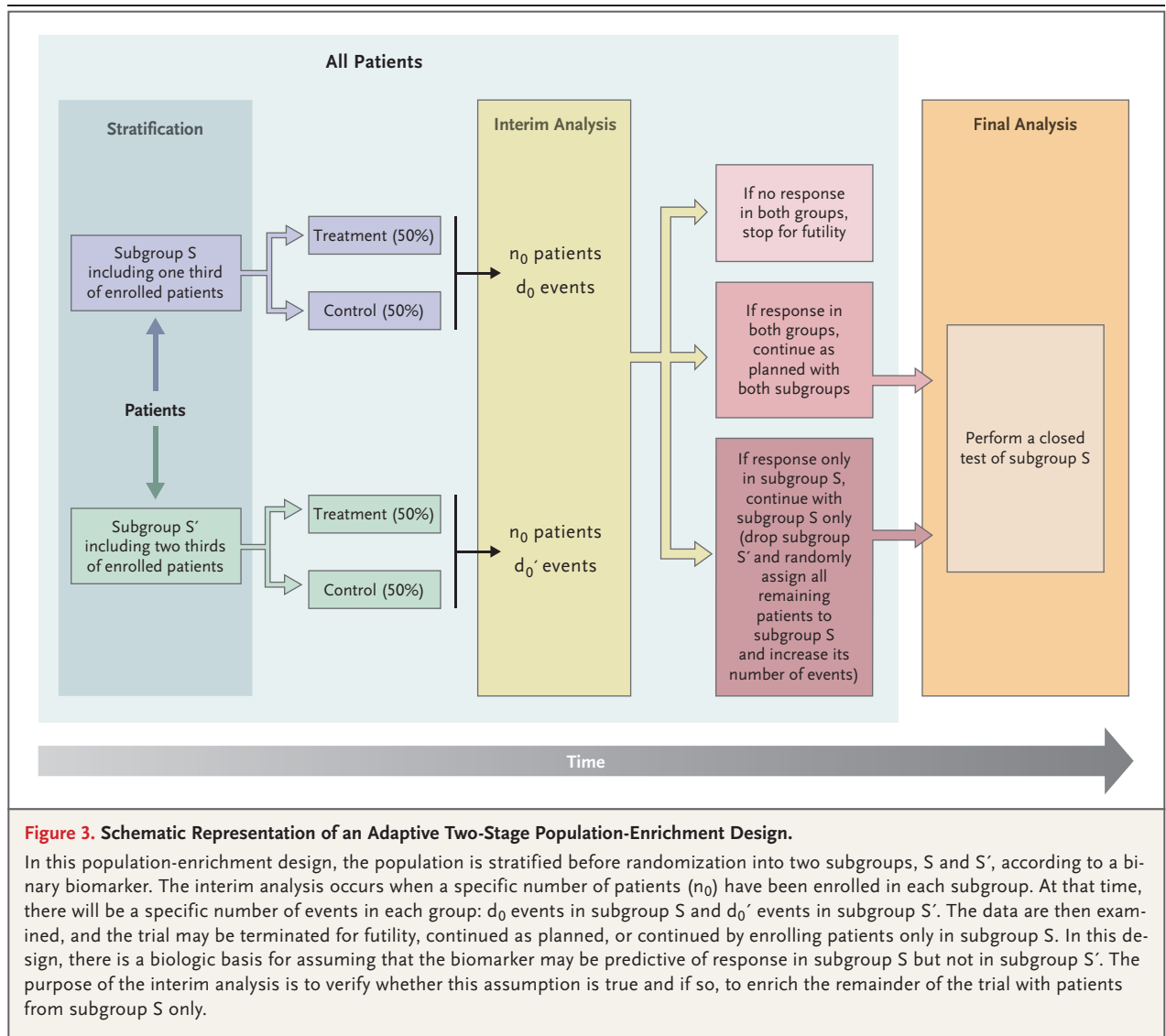Thus, the dilemma for the investigator planning a phase 3 confirmatory trial of a targeted therapy is whether to open the enrollment to all patients regardless of biomarker status or to restrict the enrollment to a targeted subgroup on the basis of a biologic understanding of the mechanism of action from early, possibly uncontrolled, clinical data. Restricting enrollment to the targeted subgroup without sufficient empirical evidence of a lack of efficacy in the nontargeted subgroup may deny a large segment of the population access to a potentially beneficial treatment. However, if a large trial is conducted in a heterogeneous population, the treatment effect may be diluted, thus resulting in an underpowered study.[15] An easily understood example is anemia due to vitamin B_{12} deficiency. In a randomized clinical trial involving patients with anemia, treating everyone in the experimental-therapy group with vitamin B_{12} would produce negative results, but the small subgroup of patients who truly have a deficiency would benefit.

An adaptive population-enrichment design is an efficient way to verify prospectively that a biomarker is predictive for a targeted therapy. The basic idea in such a design is for all participants to undergo randomization regardless of biomarker status but with the use of an interim analysis to identify whether the biomarker-positive patients benefit differentially from the targeted agent as compared with the biomarker-negative patients. If it appears that only the biomarker-positive patients are benefiting, then further enrollment in the biomarker-negative subgroup would be terminated. The final statistical analysis of the data would be based on data from the two stages with the use of closed testing and conditional error rate methods to prevent inflation of the type I error (see the sections on Statistical Methodology in the Supplementary Appendix).[16,17] Figure 3 is a schematic representation of such a design.

REGULATORY CONCERNS

At this time, regulatory agencies tend to review proposals for adaptive designs with greater scrutiny than they give to conventional designs. This situation is probably due to limited experience with such designs and serious concern that sponsors will submit poorly conceived designs that may not control the type I error and may actually be less efficient than conventional designs. As with any new approach, there must be clear design rationale, a demonstration of statistical validity,

Figure 3. Schematic Representation of an Adaptive Two-Stage Population-Enrichment Design.

In this population-enrichment design, the population is stratified before randomization into two subgroups, S and S′, according to a binary biomarker. The interim analysis occurs when a specific number of patients (n_0) have been enrolled in each subgroup. At that time, there will be a specific number of events in each group: d_0 events in subgroup S and $d_0′$ events in subgroup S′. The data are then examined, and the trial may be terminated for futility, continued as planned, or continued by enrolling patients only in subgroup S. In this design, there is a biologic basis for assuming that the biomarker may be predictive of response in subgroup S but not in subgroup S′. The purpose of the interim analysis is to verify whether this assumption is true and if so, to enrich the remainder of the trial with patients from subgroup S only.

simulation-based operating characteristics, and a comprehensive charter for the data and safety monitoring committee that addresses both the interim decision rules and the manner in which operational bias will be prevented.

The leakage of interim results could alter investigator behavior and lead to operational bias. Even if there is no leakage of interim results, the mere knowledge that there has been an adaptive change (e.g., sample-size reestimation) could cause investigators to speculate on the efficacy of the new compound, which could potentially change the enrollment and characteristics of the patients after the interim analysis. These risks can be mitigated by double-blind trials, appropriate communication with investigators, detailed and auditable standard operating procedures that document who saw what and when, and demonstration that the baseline characteristics of the patients who were enrolled before the adaptive change match those of the patients who were enrolled after the adaptive change.

Problems can arise with randomization, drug supply, and the recruitment of patients when there are adaptive changes due to dose selection, sample-size increases, or population enrichment. It is critical to ensure that the sample size at the interim analysis is adequate for making the adaptive decision. If patients are enrolled too rapidly relative to the time needed to observe the

primary end point, the planned enrollment might be completed before adequate information is available for an adaptive decision to be taken. To date, regulatory agencies have opined favorably about adaptive designs.[18,19]

FUTURE OF ADAPTIVE TRIALS

More widespread use of adaptive trial designs could accelerate the discovery process, especially if coupled with other evolving trial concepts, such as large, simple trials.[20,21] Advances in adaptive trial design will require further dissemination and acceptance of the sometimes complex statistical methods. There is an intuitive appeal of adaptive trial design and its attempt to identify the patients who are most likely to derive benefit from a therapy, and this feature will resonate well with most doctors and patients.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

REFERENCES

1. Fuster V, Bhatt DL, Califf RM, et al. Guided antithrombotic therapy: current status and future research direction: report on a National Heart, Lung and Blood Institute working group. Circulation 2012;126:1645-62.
2. Bauer P, Bretz F, Dragalin V, König F, Wassmer G. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. Stat Med 2016;35:325-47.
3. Food and Drug Administration, Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. Guidance for industry: adaptive design clinical trials for drugs and biologics. Silver Spring, MD: Food and Drug Administration, 2010.
4. Barnes PJ, Pocock SJ, Magnussen H, et al. Integrating indacaterol dose selection in a clinical study in COPD using an adaptive seamless design. Pulm Pharmacol Ther 2010;23:165-71.
5. Donohue JF, Fogarty C, Lötvall J, et al. Once-daily bronchodilators for chronic obstructive pulmonary disease: indacaterol versus tiotropium. Am J Respir Crit Care Med 2010;182:155-62.
6. Lawrence D, Bretz F, Pocock SJ. Indacaterol. In: Trifilieff A, ed. INHANCE: an adaptive confirmatory study with dose selection at interim. Basel, Switzerland: Springer Basel, 2014:77-92.
7. Bhatt DL, Stone GW, Mahaffey KW, et al. Effect of platelet inhibition with cangrelor during PCI on ischemic events. N Engl J Med 2013;368:1303-13.
8. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. Stat Med 2011;30:3267-84.
9. White WB, Cannon CP, Heller SR, et al. Alogliptin after acute coronary syndrome in patients with type 2 diabetes. N Engl J Med 2013;369:1327-35.
10. Scirica BM, Bhatt DL, Braunwald E, et al. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. N Engl J Med 2013;369:1317-26.
11. Green JB, Bethel MA, Armstrong PW, et al. Effect of sitagliptin on cardiovascular outcomes in type 2 diabetes. N Engl J Med 2015;373:232-42.
12. The ORIGIN Trial Investigators. Basal insulin and cardiovascular and other outcomes in dysglycemia. N Engl J Med 2012;367:319-28.
13. Everett BM, Brooks MM, Vlachos HE, et al. Troponin and cardiac events in stable ischemic heart disease and diabetes. N Engl J Med 2015;373:610-20.
14. Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Guidance for industry: enrichment strategies for clinical trials to support approval of human drugs and biological products. Silver Spring, MD: Food and Drug Administration, 2012.
15. Bhatt DL, Fox KAA, Hacke W, et al. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. N Engl J Med 2006;354:1706-17.
16. Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharm Stat 2011;10:347-56.
17. Mehta CR, Schäfer H, Daniel H, Irle S. Biomarker driven population enrichment for adaptive oncology trials with time to event endpoints. Stat Med 2014;33:4515-31.
18. Morgan C, Huyck S, Jenkins M, et al. Adaptive design: results of 2012 survey on perception and use. Ther Innov Regul Sci 2014;48:473-81.
19. Elsäßer A, Regnstrom J, Vetter T, et al. Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency. Trials 2014;15:383.
20. Calvo G, McMurray JJ, Granger CB, et al. Large streamlined trials in cardiovascular disease. Eur Heart J 2014;35:544-8.
21. Califf RM. Large simple trials: really, it can't be that simple! Eur Heart J 2014;35:549-51.

Copyright © 2016 Massachusetts Medical Society.