The articles by Rugo et al. (pages 23–34) and Park et al. (pages 11–22) in this issue of the *Journal* report results from the I-SPY (Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis) 2 platform, a promising adaptive strategy for matching targeted therapies for breast cancer with the patients most likely to benefit from them. I-SPY 2 identified two therapies (veliparib with carboplatin in triple-negative breast cancer and neratinib in human epidermal growth factor receptor type 2–positive, hormone receptor–negative [HER2+/HR−] cancer) that met prespecified criteria for testing in larger, phase 3 trials. The value of I-SPY 2, however, may well go beyond the clinical results described in the current articles. Adaptive multigroup trials such as I-SPY 2 have the potential to answer several questions simultaneously and more efficiently than traditionally designed trials. Which of several promising therapies appear best suited for larger, confirmatory trials? Which patients should be asked to participate in those trials? Is the chance of success in subsequent larger trials sufficient to justify the expense and time needed?

Therapies designed to target molecular subtypes of cancer may increase the chances of good responses and, equally important, may be useful in allowing patients to avoid treatments when meaningful benefit is unlikely. The challenges, however, in identifying successful targeted therapies in cancer are substantial. Targeted therapies may fail to hit their target, they may not have the predicted effect when they do, and they may also have a positive effect in the absence of a recognized target. Traditionally designed phase 2 trials that test treatments one at a time in heterogeneous groups of patients have created a traffic jam: there are too many new drugs, and the signal of a treatment effect can be diluted in these heterogeneous...
groups. The U.S. Food and Drug Administration and the European Medicines Agency have acknowledged that the commonly used designs need a makeover.5,6

The efficiency of multigroup early-phase trials has long been recognized,4 but I-SPY 2 differs in important ways from traditional early-phase trials. The I-SPY 2 platform will be used to compare up to 12 experimental therapies with a common control in subgroups of breast cancer with 10 distinct biomarker signatures. The randomization is stratified (eight strata defined by HER2 status, hormone-receptor status, and a commercially available classification based on 70 gene signatures), and adaptive randomization is used within strata to increase the likelihood of assignment to a given therapy as evidence accrues that it is more efficacious than the control in inducing pathologically confirmed complete responses in patients with locally advanced cancers. New drugs can enter the platform as they emerge from phase 1 testing and exit the platform with an estimate of the chances of future success in a phase 3 trial of prespecified size. The platform may be an appealing setting for cooperation among pharmaceutical companies and academic investigators. The entire process, including design and analysis, is carried out dynamically using Bayesian methods (see Tables 1 and 2).

Oncology has been slow to adopt Bayesian designs even though they are often well suited to settings in which inference and decisions benefit from adaptation based on accruing information. Some of the reluctance stems from a natural discomfort with replacing a familiar approach that has had some success in the past. There are other, more substantive reasons to be cautious about this new path. I-SPY 2 was designed in 2009. In the world of trial design, this new platform is still in its adolescence. There is much to be learned about the statistical models used to adaptively adjust randomization fractions and to predict the chances of success in a future trial.

How robust are the adaptive randomization probabilities and the predictive probabilities of success in a phase 3 trial to misspecifications of the model? What are effective ways to communicate
to our clinical colleagues the modeling assumptions used, the potential vulnerability of the model to errors, and the best ways to explain these designs to trial participants? What visual and numerical summaries provide insight into the trial data? Simple summary statistics such as odds ratios or relative risk can be misleading, and the usual CONSORT diagram does not reflect the dynamics of the I-SPY 2 randomization. Will the predicted chance of future success (85%) upset equipoise for trial investigators or influence the kinds of patients investigators choose to enroll or not enroll in a future trial? It is imperative to investigate these questions in depth.

Despite these unresolved issues, I-SPY 2 is an important addition to the inventory of trial designs. The second figures in the articles by Rugo et al. and Park et al., showing estimated distributions of rates of pathologically confirmed complete responses are appealing. The articles provide 95% probability intervals, but the graphs make it easy to identify, for instance, 90% or 99% intervals. Clinicians can interpret the results of the trial in a manner consistent with their own sense of acceptable uncertainty. Adaptively adjusting randomization probabilities makes much more sense than specifying an unbalanced but fixed randomization at the beginning of a trial. Perhaps most important, I-SPY 2 holistically integrates the ideas of Bayesian design and analysis in the important setting of phase 2 testing of new cancer drugs. The design of the platform acknowledges the complexity of phase 2 testing in cancer.

As George Box famously wrote, “Essentially, all [statistical] models are wrong, but some are useful.” The most useful models add much more than just statistical information; they also help bring clarity to cases in which noise and uncertainty threaten to overwhelm progress. The fundamental tenets behind the I-SPY 2 platform and model — the multi-group platform, options for “gradation” and addition of drugs, adaptive randomization, and prediction of success in confirmatory trials — are important first steps toward the efficient use of clinical resources. As more new targets and drugs are discovered, traditional statistical designs, at best cumbersome and inefficient today, will be wholly insufficient for matching patients with effective drugs. We applaud the use of I-SPY 2 described here and urge continued innovation in trial design, especially in both earlier phase 1 and later phase 3 settings.

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

From the Department of Biostatistics and Computational Biology, Dana–Farber Cancer Institute, and the Department of Biostatistics, Harvard T.H. Chan School of Public Health — both in Boston.


DOI: 10.1056/NEJMmp1602256
Copyright © 2016 Massachusetts Medical Society.