STATISTICS IN MEDICINE

I-SPY 2 — A Glimpse of the Future of Phase 2 Drug Development?

David Harrington, Ph.D., and Giovanni Parmigiani, Ph.D.

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

Table 1. Bayesian versus Frequentist Approaches in Clinical Trials.

Variable	Bayes	Frequentist
Differences		
Main goal of inference	Predict outcomes of future trials and absolute risk for future patients.	Estimate population average effects.
Assumptions	Requires explicit specification of prior distributions of unknown population parameters. Incorporates a priori knowledge and clinical judgment formally. May be sensitive to specification of prior distributions.	Does not require explicit specification of prior distributions of unknown population parameters. Incorporates a priori knowledge and clinical judgment informally.
Interim monitoring	Only the data actually obtained are relevant for final conclusions (e.g., a credible interval or predictive probability). Whether or not a clinician examines accumulating evidence with the possibility of stopping the trial does not affect inference.	Both the data actually obtained and the probabilities of data not obtained are relevant for final conclusions (e.g., a P value). Whether or not a clinician examines accumulating evidence with the possibility of stopping the trial does affect inference.
Ease of use	Often computationally complex; careful modeling often requires simulation-based calculations.	Often computationally simple, though careful modeling may require simulation-based calculations.
Similarities		
Adaptation	Can incorporate adaptive designs, multistage trials, early stopping, and adaptive randomization.	
Role of statistical judgment	Options for data-driven analyses are available. Skill and substance-area knowledge of the data analyst are important in drawing correct conclusions.	
Compatibility	It is feasible to combine a Bayesian design with a frequentist analysis or a frequentist design with a Bayesian analysis.	
Prior knowledge	Both approaches rely on prior knowledge and clinical judgment (though they incorporate them in different fashions).	

Table 2. Bayesian vs. Hypothetical Standard Frequentist Design.

Variable	I-SPY 2 Design	Standard Frequentist Design for I-SPY 2
Main goal of inference	Posterior distributions of rates of pathological complete response for the investigational drug (neratinib or veliparib) and the control. Predicted probability of success in a subsequent phase 3 trial.	Odds ratio or relative risk of response, investigational drug vs. control, with confidence interval and P value
Assumptions	Specification of prior distributions of response rates for investigational drug and control; specification of model for adapting randomization fraction as information becomes available, including model for imputation of pathological complete response based on imaging in previous patients	Specification of anticipated rates of pathological complete response in the control group and of clinically relevant target differences; specification of prior stratification for randomization of subtypes of breast cancer; distributions of unknown population parameters
Randomization	Adaptive randomization increases likelihood of participant receiving treatment assignment that may be of benefit. Estimates of pathological complete response rates must be model-based because of lack of balance of patients' baseline characteristics across treatments.	Constant randomization probabilities do not preferentially target patients who may benefit from a treatment; heterogeneity of patient groups receiving a treatment may dilute estimates of treatment effects. Constant randomization probabilities ensure approximate balance of baseline characteristics across treatments and allow direct comparisons.
Interim monitoring	A treatment is declared potentially successful if predicted probability of success in phase 3 trial is at least 85%. Predicted probability of success is evaluated frequently during the trial. Experimental treatment is dropped for futility if predictive probability of success in a phase 3 trial is <10% in all 10 signatures.	Summary test statistics calculated a small number of times (typically 3 or 4) with P values checked against interim monitoring boundaries for futility and efficacy. Treatment effect estimates can be used for future trials, but groups are not selected on the basis of predicted success rates of future phase 3 trial.
Ease of use	Software for calculation of posterior distribution of pathological complete response or predictive probability of success not generally available. Accruing information must be updated frequently and accurately for adaptive randomization.	Summary and test statistics based on ratios or differences of proportions of pathological complete response. Open-source or other software for design and analysis widely available. Few software packages available for adjusting estimates for treatment effects after early stopping. Accurate data updates required for interim monitoring.

groups.[1] The U.S. Food and Drug Administration and the European Medicines Agency have acknowledged that the commonly used designs need a makeover.[2,3]

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."[5] 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 multigroup platform, options for "graduation" 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.

1. Betensky RA, Louis DN, Cairncross JG. Influence of unrecognized molecular heterogeneity on randomized clinical trials. J Clin Oncol 2002;20:2495-9.
2. Guidance for industry: adaptive design clinical trials for drugs and biologics. Rockville, MD: Food and Drug Administration, 2010 (http://www.fda.gov/downloads/DrugsGuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf).
3. Committee for Medicinal Products for Human Use (CHMP). Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. London: European Medicines Agency, 2007 (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf).
4. Freidlin B, Korn EL, Gray R, Martin A. Multi-arm clinical trials of new agents: some design considerations. Clin Cancer Res2008;14:4368-71.
5. Box GEP, Draper NR. Empirical model building and response surfaces. New York: John Wiley, 1987.

DOI: 10.1056/NEJMp1602256