Lessons From Adaptive Randomization: Spying the I-SPY2 Trial in Breast Cancer

Bishal Gyawali, MD, PhD, and Saroj Niraula, MD, MSc

Cancer drug development is tedious and costly, occurring conventionally in 3 sequential phases of clinical trials. Such rigor of evaluation maximizes the chances that the new drug is adequately safe and effective before being used in clinical practice. It is common therefore for a new drug to cost >\$600 million and take more than a decade to progress from preclinical stage to phase III clinical trial and subsequent registration for public use.^{1,2} The vast majority of drugs fail during the process, often at a later phase of development at the expense of considerable financial, temporal, and human resource.³ Therefore, it is prudent that drugs be rigorously scrutinized before being tested in a phase III trial so that only those with high chances of benefit are chosen to move forward.

Adaptive randomization is one of those innovative clinical trial designs intended for that purpose: to improve efficiency and save valuable resources.⁴ This trial design was recently in the spotlight after the results from another arm of the I-SPY2 trial, a phase II multiarm trial using adaptive randomization, was presented at the Plenary Session of the American Association for Cancer Research Annual Meeting 2020. The hallmark of adaptive randomization is the use of Bayesian theorem to estimate future probabilities based on previously known and evolving new information.⁵ This design offers several practical advantages over the conventional frequentist approach: it can test multiple drugs simultaneously, requires a smaller sample size, focuses resources on drugs that are most likely to succeed, and preferentially randomizes patients to treatments that are most likely to benefit. However, before widespread incorporation of this design into clinical trials, we need to ascertain that it does not compromise on the safety nets offered by conventional drug development. Drawing examples from the phase II I-SPY 2 trial,^{6,7} this commentary highlights some issues and challenges in the adaptive randomization approach to consider in interpretations of such trials.

I-SPY 2 is a 20-arm mega randomized controlled trial (RCT) in the neoadjuvant treatment of early breast cancer with the primary endpoint of pathologic complete response (pCR; ClinicalTrials.gov identifier: NCT01042379). The arms consist of one common control arm of standard neoadjuvant chemotherapy with paclitaxel (with an alternate control arm of paclitaxel + trastuzumab or pertuzumab + trastuzumab if HER2-positive) against which other experimental arms are tested. The other 18 experimental arms consist of a single agent or a combination therapy in addition to paclitaxel. Assignment of patients to one of the experimental arms depends on the results of 10 different biomarker signatures based on hormone receptor and HER2 status and a 70-gene genomic assay so that patients are randomized to the arm with highest likelihood of response. So far, 6 experimental arms have been considered successful ("graduated"),⁸⁻¹¹ 4 have failed, and the rest^{9,10} are either in progress or are halted.

Choice of Control Arm and Combination Experimental Treatments

Graduation in the I-SPY 2 trial has been defined as an estimated probability of success of >85% in a subsequent phase III trial of 300 patients. Of the experimental arms, combination veliparib/carboplatin was among the first tested; it was tested among 72 patients with HER2-negative breast cancer and graduated in the triple-negative breast cancer cohort. This was also the only combination for which we could find published results of a subsequent phase III RCT.¹² Of particular note, in the

BISHAL GYAWALI, MD, PhD

Bishal Gyawali, MD, PhD, is an Assistant Professor in the Department of Public Health Sciences, Scientist in the Division of Cancer Care and Epidemiology, a Clinical Fellow in the Department of Medical Oncology at Queen's University, Kingston, Canada, and an affiliated faculty member at the Program On Regulation, Therapeutics And Law (PORTAL) in the Department of Medicine at Brigham and Women's Hospital. He is also currently serving as a research consultant to the Nepal Health Research Council and has previously served as a medical consultant for the not-for-profit Anticancer Fund in Belgium.

Dr. Gyawali is also on the Editorial and Advisory Boards of multiple medicine and oncology journals, has authored or co-authored more than 100 peer-reviewed articles, and is the recipient of the 2020 ASCO Conquer Cancer Foundation Global Oncology Young Investigator Award in Implementation Science. He was also awarded by the Government of Nepal with Young Health Researcher Award in 2020. His areas of academic interest include cancer policy, global oncology, evidence-based oncology, financial toxicities of cancer treatment, clinical trial methods, and supportive care. He is an advocate of "cancer groundshot," a term he coined to imply that investment should be made in proven high-value interventions in cancer care that are easy to implement globally and are affordable. Dr. Gyawali tweets at @oncology_bg.

doi: 10.6004/jnccn.2020.7648

The ideas and viewpoints expressed in this commentary are those of the author and do not necessarily represent any policy, position, or program of NCCN.

SAROJ NIRAULA, MD, MSc

Saroj Niraula, MD, MSc, is an Associate Professor of Medicine at the Department of Internal Medicine, University of Manitoba; Medical Oncologist at CancerCare Manitoba; and Affiliate Scientist at the Research Institute of Oncology and Hematology, CancerCare Manitoba. He chairs the Medication Management and Safety Committee at CancerCare Manitoba, a body responsible for the review and approval of new cancer drugs in the Province, and also chairs multidisciplinary breast cancer group at CancerCare Manitoba.

Dr. Niraula's research interests include cancer drug regulation, cancer policy, clinical trials methodology, evidence synthesis, and global oncology, and he is well published in these areas. He has been the recipient of multiple awards from professional organizations such as ASCO and ESMO, as well as other national recognitions. He volunteers at ASCO in their international outreach program, and is actively involved in mentoring young oncologists from countries with limited resources. He is also involved in several global efforts to improve disparities in cancer outcome between high-income and low-and-middle-income countries and serves on the health committee of the Non-Resident Nepali Association, an organization spread across most countries around the world. Dr. Niraula tweets at @sarojniraula.

subsequent study, a third arm that was not evaluated in I-SPY 2 (carboplatin + paclitaxel) was added. The study showed significantly higher rates of pCR in the paclitaxel/carboplatin/veliparib group compared with patients receiving paclitaxel alone (168/316 [53%] vs 49/158 [31%] patients, respectively; $P < .0001$), but not compared with patients receiving paclitaxel + carboplatin (92/160 [58%]; $P = .36$).¹²

The veliparib/carboplatin/paclitaxel combination has not yet been approved for routine use, 4 years after the phase III trial was published. This is appropriate because the addition of veliparib to carboplatin + paclitaxel did not improve pCR rates. In fact, the paclitaxel + carboplatin arm resulted in numerically higher pCR rates than the arm with veliparib added to that combination. This makes us wonder about the appropriateness of paclitaxel as a common control arm in all subtypes of breast cancer, especially the absence of carboplatin in the triple-negative breast cancer subtype. Despite the veliparib/carboplatin combination graduating in I-SPY 2, the subsequent phase III trial showed that the benefits of combination veliparib/carboplatin was primarily derived from carboplatin. Thus, adaptive randomization may not be the best platform to test the efficacy of combination therapies upfront. Rather, this platform could be best used to test efficacy of monotherapies, and the drugs that graduate could be considered for testing in rational combinations in the subsequent phase III trials.

Premature Leakage of Information

Premature leakage of information is another challenge with adaptive design, particularly in unblinded trials in which investigators can observe the trends of higher assignments of patients to arms that match a particular tumor phenotype. Because Bayesian posterior probabilities depend on ongoing trial data, the ongoing results are usually known and conveyed to the new enrolling patient while obtaining informed consent. Whether a patient would want to be randomized to a control cohort or to an experimental agent that has not performed particularly well in the biomarker signature of that patient remains an ethical challenge. Additionally, the information leaking prematurely to the public and media can create unnecessary hype about an ongoing trial and drug, with numerous detrimental sociomedical consequences.

Ethical Issue in Confirmatory Trials

Conducting confirmatory phase III RCTs on drugs that have been successful in adaptive randomization trials poses a major ethical dilemma. Clinical trials are conducted based on the principle of clinical equipoise. With a chance of predicted success as high as $\geq 85\%$ clinical equipoise is clearly violated. Would it be ethical to randomize patients to a placebo arm when the experimental arm has such a high chance of success? And will the patients accept randomization with this information disclosed? Will this promote off-label use of these agents without confirmatory data available from phase III trials? These are a few fundamental questions about conducting phase III clinical trials on drugs that graduate based on adaptive randomization.

Appropriateness of Endpoint in Definitive Trials

Another important concern with adaptive randomization is the choice of endpoint. So far, 6 drugs have graduated from the I-SPY 2 trial for testing in phase III RCTs based on higher rates of pCR. Indeed, the phase III trials that have been planned based on the results of I-SPY 2 also have primary endpoints of pCR. Because Bayesian statistics involve continued updating of probabilities based on ongoing results, rapidly evaluable endpoints form the basis of adaptive randomization. However, improvement in pCR is controversial as an endpoint in breast cancer, especially in pivotal phase III registration trials, because it measures biologic activity of the drug rather than direct benefit to the patients. Although pCR is a prognostic marker at the individual patient

level, it remains a poor surrogate marker at the trial level. Pooled analyses have failed to validate pCR as a surrogate marker for improved event-free or overall survival in the overall breast cancer population.^{13–15} Furthermore, the only two HER2-targeted therapies aside from trastuzumab used in early-stage breast cancer, pertuzumab and lapatinib, both of which led to almost doubling of pCR rates compared with trastuzumab alone in phase II studies,^{16,17} unfortunately resulted in either marginal or no improvement in both disease-free and overall survival in phase III clinical trials.^{18,19}

Issue of Double Surrogacy

The pCR rates reported in the I-SPY 2 trial are not the true pCR rates but rather probable rates based on Bayesian theorem. These rates are continually estimated and updated based on the decrease in tumor volume as assessed by MRI scans. Decrease in tumor volume on MRI scans is yet to be validated as a surrogate for achieving pCR. Thus, in I-SPY 2, a correlate (pCR) is being predicted using another correlate (MRI), both of which lack validation as surrogates. Indeed, neratinib had previously attained an estimated probability of success of 85% in a phase III study based on MRI volume changes; however, when true pCR data were assessed later, estimated probability of success decreased to 79%.⁶ Similarly, although combination durvalumab + olaparib has graduated, the true probability of success in a subsequent phase III trial was 81%.¹¹

Absence of Patient-Reported Outcomes and Adequate Safety Data

Clinical benefit to patients does not imply efficacy rates alone. Some drugs are useful not because they are more effective than the controls but because they improve quality of life (QoL), and other effective drugs cannot be recommended because of their detrimental impact on QoL. Thus, health-related QoL and patient-reported outcomes (PROs) are integral to establishing therapeutic index (measured as the ratio of benefit and risk) for an individual patient. However, QoL and PROs do not have any biomarkers and thereby cannot be accounted for in “success of therapy” defined using the Bayesian probability.

The first Hippocratic principle is “primum non nocere.” Indeed, uncertainties regarding the efficacy of treatment are acceptable before definitive clinical trials, but substantial uncertainties regarding potential hazards of treatment are not. We and others have shown previously that new anticancer agents possess increased risk of serious adverse effects compared with conventional anticancer strategies.²⁰ Toxicity

that is not evident within the short trial period may become apparent later on, thus compromising patient safety. As these trials focus primarily on predicting efficacy, toxicity reporting may remain incomplete and even ignored. For example, although neratinib graduated in the I-SPY 2 trial based on probability of success in a phase III trial, it is associated with severe diarrhea, leading to detrimental patient QoL, at least in the short-term.²¹

Furthermore, the statistics used in conducting and assessing adaptive trials are more complex than those used in the usual trials, making it very difficult, if not impossible, for a clinician to read between the lines of trial publications and understand the potential pitfalls with the trial design and results. Hence, oncologists would need to trust the industry and believe the results they are presented with, which leaves more room for inappropriate manipulations. It is thus imperative that oncologists and other clinicians be acquainted with relevant statistics and be able to understand such complex trial designs.

With multiple drug arms tested simultaneously in adaptive designs, publication bias can also be a potential problem. Although we do not know whether the remaining drug arms from I-SPY 2 will be published, enthusiasm for publishing negative results will probably be low. So far, all published results from I-SPY 2 have been for the positive arms only.

Conclusions

The main purpose of adaptive trial designs is to expedite trials and bring drugs to market sooner. The main goal of oncology practice is to help patients live longer and/or better. Getting medications into the market sooner does not always contribute to this goal. This situation demands extra caution from all stakeholders to balance the desire for rapid access to drugs while ensuring that available drugs are both safe and effective. The point of discussing and acknowledging these challenges is not to neglect the potential benefits that this design has to offer—adaptive randomization is a highly efficient approach to get effective drugs quickly to patients—it is to fully understand the unique strengths and weaknesses of this design before incorporating it into routine oncology research.

Disclosures: The authors have disclosed that they have no financial interests, arrangements, or affiliations with the manufacturers of any products discussed in this article or their competitors.

Correspondence: Saroj Niraula, MD, MSc, University of Manitoba and CancerCare Manitoba, 675 McDermot Avenue, Winnipeg, MB, R3E 0V9, Canada. Email: sniraula@cancercare.mb.ca

References

- Jardim DL, Schwaederle M, Hong DS, et al. An appraisal of drug development timelines in the era of precision oncology. *Oncotarget* 2016;7:53037–53046.
- Prasad V, Mailankody S. Research and development spending to bring a single cancer drug to market and revenues after approval. *JAMA Intern Med* 2017;177:1569–1575.
- Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. *Nat Rev Drug Discov* 2019;18:495–496.
- Bhatt DL, Mehta C. Adaptive designs for clinical trials. *N Engl J Med* 2016;375:65–74.
- Harrington D, Parmigiani G. I-SPY 2—a glimpse of the future of phase 2 drug development? *N Engl J Med* 2016;375:7–9.

6. Park JW, Liu MC, Yee D, et al. Adaptive randomization of neratinib in early breast cancer. *N Engl J Med* 2016;375:11–22.
7. Rugo HS, Olopade OI, DeMichele A, et al. Adaptive randomization of veliparib-carboplatin treatment in breast cancer. *N Engl J Med* 2016;375:23–34.
8. Chien AJ, Tripathy D, Albain KS, et al. MK-2206 and standard neoadjuvant chemotherapy improves response in patients with human epidermal growth factor receptor 2-positive and/or hormone receptor-negative breast cancers in the I-SPY 2 trial. *J Clin Oncol* 2020;38:1059–1069.
9. DeMichele AM, Moulder S, Buxton M, et al. Efficacy of T-DM1 + pertuzumab over standard therapy for HER2+ breast cancer: results from the neoadjuvant I-SPY 2 trial. Presented at the AACR 107th Annual Meeting 2016; April 16–20, 2016; New Orleans, Louisiana.
10. Nanda R, Liu MC, Yau C, et al. Effect of pembrolizumab plus neoadjuvant chemotherapy on pathologic complete response in women with early-stage breast cancer: an analysis of the ongoing phase 2 adaptively randomized I-SPY2 trial. *JAMA Oncol* 2020;6:676.
11. Pusztai L, Han HS, Yau C, et al. Evaluation of durvalumab in combination with olaparib and paclitaxel in high-risk HER2 negative stage II/III breast cancer: results from the I-SPY 2 trial. Presented at the AACR Annual Meeting; April 27, 2020; virtual.
12. Loibl S, O'Shaughnessy J, Untch M, et al. Addition of the PARP inhibitor veliparib plus carboplatin or carboplatin alone to standard neoadjuvant chemotherapy in triple-negative breast cancer (BrighTNess): a randomised, phase 3 trial. *Lancet Oncol* 2018;19:497–509.
13. Cortazar P, Geyer CE Jr. Pathological complete response in neoadjuvant treatment of breast cancer. *Ann Surg Oncol* 2015;22:1441–1446.
14. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* 2014;384:164–172.
15. Gyawali B, Hey SP, Kesselheim AS. Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs [published online April 22, 2020]. *EClinicalMedicine*, doi: <https://doi.org/10.1016/j.eclinm.2020.100332>journal
16. Gianni L, Pienkowski T, Im YH, et al. Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. *Lancet Oncol* 2012;13:25–32.
17. Baselga J, Bradbury I, Eidtmann H, et al. Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): a randomised, open-label, multicentre, phase 3 trial. *Lancet* 2012;379:633–640.
18. von Minckwitz G, Procter M, de Azambuja E, et al. Adjuvant pertuzumab and trastuzumab in early HER2-positive breast cancer. *N Engl J Med* 2017;377:122–131.
19. Piccart-Gebhart M, Holmes E, Baselga J, et al. Adjuvant lapatinib and trastuzumab for early human epidermal growth factor receptor 2-positive breast cancer: results from the randomized phase III adjuvant lapatinib and/or trastuzumab treatment optimization trial. *J Clin Oncol* 2016;34:1034–1042.
20. Niraula S, Amir E, Vera-Badillo F, et al. Risk of incremental toxicities and associated costs of new anticancer drugs: a meta-analysis. *J Clin Oncol* 2014;32:3634–3642.
21. Delaloge S, Cella D, Ye Y, et al. Effects of neratinib on health-related quality of life in women with HER2-positive early-stage breast cancer: longitudinal analyses from the randomized phase III ExteNET trial. *Ann Oncol* 2019;30:567–574.