Editorial

The Evolution of Clinical Trials in Oncology: Defining Who Benefits from New Drugs Using Innovative Study Designs

RICHARD M. GOLDBERG,[a,b,c] LAI WEI,[d] SOLEDAD FERNANDEZ[d]

[a]Department of Medicine, [b]West Virginia University Cancer Institute, and [c]Mary Babb Randolph Cancer Center, West Virginia University, Morgantown, West Virginia, USA; [d]Department of Biomedical Informatics, Ohio State University Center for Biostatistics, Ohio State University, Columbus, Ohio, USA

INTRODUCTION

The dawn of the modern era of randomized drug treatment trials (RCTs) in clinical medicine was ushered in by the seminal study of streptomycin versus placebo for the treatment of tuberculous meningitis reported in 1948 [1]. As drugs with apparent promise for treating cancer in humans emerged from cell culture and animal studies, the oncology community experimented with and subsequently embraced the notion of the RCT. Courageous advocates of the RCT in oncology such as Bernard Fisher and the clinical trial organization that he led, as well as the U.S. government, which funded these studies through the National Cancer Institute, challenged and ultimately contributed to changing the paradigm for testing cancer treatments in the U.S. [2]. Similar commitments of intellectual power and resources across the world led to international trials and global efforts to generate evidence of promise in preclinical model systems and translate those findings into improved patient outcomes. Advances in cancer treatment and the patients who have benefited as a consequence owe an enormous debt to the participants in and proponents of RCTs.

The science of clinical trials in oncology evolved to include phase I dose-finding trials, phase II studies to establish efficacy in a single tumor type, phase III trials comparing standards of care with potential advances in care, and phase IV studies to extend safety and activity data in a post-marketing scenario. As a consequence, we can expect higher cure rates in locally confined cancers and in some metastatic cancers. Investigators have substantially extended median overall survival statistics for trial subjects and patients treated off of trials who have had advanced stages of malignant diseases. An example of this extension of median overall survival in the setting of advanced, incurable disease comes from patients treated on clinical trials for advanced colorectal cancer. In 50 years of clinical trials, median survivals have improved from 12 months in the 5-Fluorouracil era of the 1980s to >30 months in the current era where targeted therapies are combined with chemotherapy strategies across multiple lines of treatment [3].

Efforts to improve clinical outcomes have spawned a new discipline of clinical trial design and led to synergistic intellectual partnerships between clinicians, translational scientists, statisticians, pharmaceutical companies, and regulatory authorities. At its best, this "concordance of interest" across the academic, clinical, industry, and regulatory stakeholders has revolutionized cancer care and benefited mankind. It has saved many lives and extended others, built careers, led to enormous and highly profitable pharmaceutical and medical device corporations, and consumed billions of health care dollars.

As the statisticians and clinical trialists were designing and completing studies, the laboratory scientists were laboring in parallel to unravel the biology of cancer and understand how to exploit new discoveries for the benefit of those afflicted with these diseases. The Human Genome Project is one high-profile and expensive example of an investment whose huge implications for changing cancer treatment are just becoming clear. It has been said that genomics is making every type of cancer into a rare cancer. It is allowing us to change our thinking and complicating our practice of medicine.

While adenocarcinomas look virtually alike under the microscope, the factors that drive their growth and the genomic targets that determine their vulnerabilities divide them into categories that can be exploited by drug developers and treating clinicians. This gives us the promise that the rather crude tools of chemotherapy drugs, with their narrow marginal differences between effects on tumor cells and rapidly dividing healthy cells, can be succeeded by more precise interventions that switch off the cellular defects that lead to cancer with a minimum of collateral damage to healthy tissues. In recent years, the Food and Drug Administration (FDA) in the U.S. has begun to recognize that data acquired from small trials in rare tumors or in biologically defined rare subtypes of common tumors can lead to startling evidence of efficacy that can provide an evidence base to permit rapid approval of new agents. Examples include findings of dramatic responses in *BRAF* mutated melanoma patients treated with vemurafinib and in patients with *ALK 1* mutated non-small cell lung cancers treated with crizotinib. The most extreme example of economical use of resources is the "N-of-1" trial, where a single patient is the sole unit of observation, and outcomes in such endeavors are

no longer considered anecdotes but rather important clues to drug efficacy [4].

During the half-century that clinical trials have been on the rise in oncology, analytic technology has also evolved. We have gone from recording results in spiral-bound notebooks to worldwide, real-time, web-based data entry, and from being wedded to our pens and papers to having extensive daily relationships with our keyboards. We have also changed from experience-based medicine that made gray-haired clinicians the sages of the field to evidence-based medicine as the gold standard that is accessible to even the least-seasoned practitioner. That evidence base has led us to develop guidelines that standardize scientific inquiry as well as clinical practice. We have also learned how to design clinical trials to make them more readily comparable.

Along the way, incredible amounts of data on patients treated in clinical trials have been generated and archived. Some of those data were generated with public and some by corporate resources. Individuals such as Daniel Sargent, a statistician, and Aimery de Gramont, a clinical trialist, who are the custodians of clinical trial data from multiple studies whose results are in the public domain, realized the potential for pooling data in the 1990s and began efforts to do that [5, 6]. Pharmaceutical companies have begun to realize the power of pooling their data with the data from the public sector. An example of a global data pooling initiative is Project Data Sphere [7]. This bank of information is garnered from clinical trials conducted over the last few decades and is overseen by the CEO Roundtable on Cancer. The CEO Roundtable was convened by President George H.W. Bush to bring stakeholders from the public sector, patient advocacy community, governmental agencies, and industrial partners together in service of the public interest in advancing our knowledge about managing cancer.

As these various disciplines, common missions, and technological advances converge, it provides an opportunity to think about how the shared knowledge and advances in our understanding can permit us to evolve our evaluative paradigms if we are willing to consider that kind of adaptation. It is in this context that an international group of investigators led by H. G. Eichler thought carefully about how the scholarly work of the various sectors described above might change our models for drug testing in humans [8]. They used the work of scholars from around the globe and developed the treatise "'Threshold Crossing': A Useful Way to Establish the Counterfactual in Clinical Trials?," a paper that was recently published in Clinical Pharmacology and Therapeutics. The fact that these investigators punctuated their title with a final question mark illustrates their understanding that what they are proposing is a bold move, and in the paper, they invite the reader to consider diverging from classical thinking in the disciplines that they represent.

What they propose is to make use of the knowledge that we have generated in completed trials and deposit that information in data banks to enable us to modify clinical trial designs using the premise that what we have learned can translate into economic trial designs that use patient, drug, and financial resources with maximum efficiency to test new drugs and drug combinations in cancer clinical trials. These investigators have termed their approach "the counter-factual." How does the counter-factual approach work? What are the potential benefits? What are the pitfalls of this approach? How

different is what they propose from current trial methodology such as Simon 2-stage designs? What are the practical issues that need to be managed in order to test their proposition?

WHAT IS THE "COUNTER-FACTUAL?"

The assessment of the causal effects (benefits and harms) of any treatment revolves around the question of how the outcome of treatment (the factual) compares with "what would have happened if patients had not received the test treatment or if they had received a different treatment known to be effective" (the counterfactual). The authors contemplate the notion of opportunities taken (in an RCT by randomized assignment) and opportunities foregone by categorizing the latter as "the road not taken," just as Robert Frost did in his 1920 poem of the same name. The authors propose using data from patients accrued from trials that have been previously completed and that had their data internally and externally vetted as a control group that does not have to be simultaneously randomized to the intervention being tested. In a sense, the fact that others have been down Frost's more travelled road (the control arm in an RCT) before can be used to gain knowledge and predict outcomes for those taking the less travelled pathway (the experimental arm in an RCT).

Because there are now large numbers of patients treated with documented patient characteristics, the experience of these previous travelers can be matched to the study sample in the trial to be conducted. The principal drawback to this approach has been the potential existence of unapparent confounders that differentiate the prior travelers from the current ones. Use of simulation techniques where control groups from multiple studies can be employed as surrogate control groups when the trial was actually randomized have the potential to test the validity of this approach.

Important considerations about the "threshold-crossing" design are listed and discussed below.

BORROWING INFORMATION FROM HISTORICAL DATA FOR DESIGNING TRIALS

There are currently two major types of approaches for clinical trial designs: (a) the traditional frequentist approach and (b) the Bayesian approach. As Berry explains in a 2006 review, under the frequentist approach, the parameters are fixed and not subject to probability distributions. Using the frequentist approach, a study is designed and its boundaries are set and adhered to at the outset [9]. Under the Bayesian framework, anything that is unknown (parameters) is assumed to have a probability distribution. Those distributions are dynamically updated as information is accumulated during the trial conduct. Such updating can be incorporated completely, explicitly, and prospectively (Fig. 1).

The frequentist approach has been widely used in the latter half of the 20th century. In the last 10 years, Bayesian designs have seen increasing use in medical research. For example, in 2006, the Center for Devices and Radiological Health issued a draft of "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials" [10]. Using these approaches depends upon real-time data entry, high-speed computers, and efficient computational algorithms. An example of the new approach is the I-SPY trial. Rugo and colleagues developed the I-SPY 2 study, a phase 2, multicenter, adaptively randomized trial to
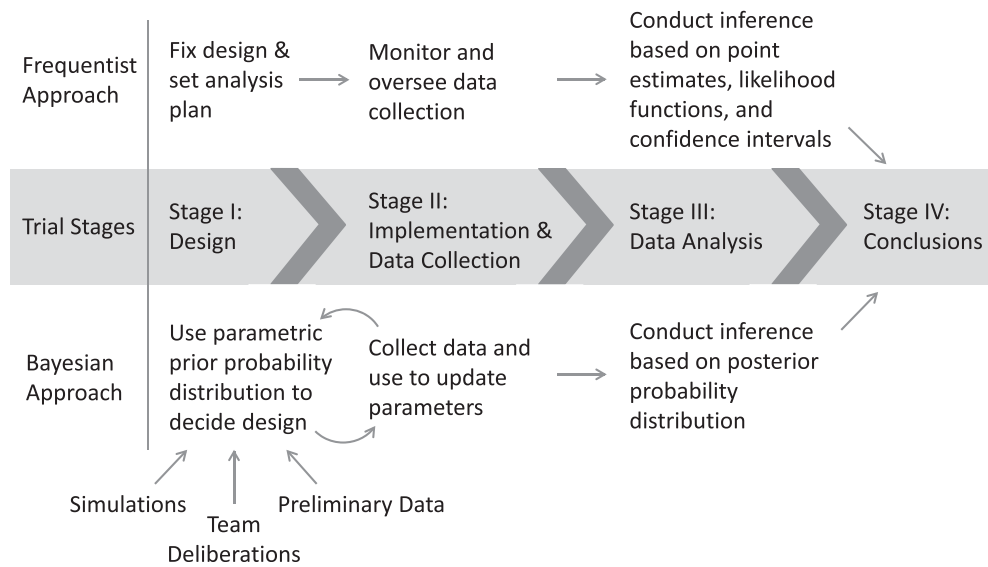
Figure 1. A comparison of the frequentist and Bayesian approaches to clinical trial design, conduct, and analysis.

screen multiple experimental regimens in combination with standard neoadjuvant chemotherapy for breast cancer [11]. Patients with eight biomarker subtypes underwent adaptive randomization to receive regimens that had the potential to achieve better performance than standard, non-biomarker-based therapy. Regimens were to be selected to move on from phase 2 if and when they were predicted to have a high probability of success in a subsequent phase 3 neoadjuvant trial within the relevant biomarker signature. The study showed that, in one of eight biomarker-defined subtypes of triple-negative breast cancer, a subset of patients had higher rates of pathological complete response when veliparib-carboplatin was added to standard therapy compared with standard therapy alone. The investigators predicted an 88% probability that the new regimen would outperform standard treatment using a parsimonious sample size of 72 randomized patients and 44 patients concurrently assigned to receive control therapy. This dynamic approach is an efficient method of taking genomically defined subsets of patients with a common tumor, narrowing down therapeutic choices based on drug activity observed in the phase 2 setting, and more precisely predicting the outcome of future trials. The future trial is now in progress to confirm the prediction.

In general, for any type of design, prior distributions (under a Bayesian framework) or model assumptions (under a non-Bayesian framework) derived from *historical data* have been used to increase the precision of parameter estimates. This has been done by choosing examples of one or more studies in a similar group of patients as those to be treated in the newly designed study. Even for single-arm trials, designs will be more robust when the parameters of study design are based on reliable and appropriate prior experiences derived from historical data. Traditionally, one-arm trial designs have been used as precursors to gather data for randomized trials when these are feasible based on the number of patients available for study enrollment. For situations where randomization is not possible because of the rarity of the tumor type, based on histologic features and site of origin or molecular genetic characteristics, one-arm trials with comparison of the experimental regimen with historical outcomes may be the only practical study

design. These one-arm designs need to be well thought out and planned in order to minimize the likelihood for drawing misleading conclusions. In particular, there is the need to avoid selection bias and confounding in single-arm trials. The novelty of the approach proposed by Eichler and colleagues is that it takes advantage of the availability of large historical data banks from patients previously treated on similar trials to bolster future study design and to maximize efficiency in a time when all tumors are rare tumors.

There are examples in which historical data were used to inform the design of randomized trials. For example, Hobbs et al. proposed adaptive adjustments of the randomization ratio using historical control data [12]. This design allows assessing concurrent and historical heterogeneity at the times of interim analyses to update the randomization ratio. The major benefit of this approach is that it maximizes the number of patients on the novel or proposed treatment or therapy and minimizes the number of study subjects who are assigned to the comparator conventional approach. This type of adaptive design enhances efficiency when implementing controlled clinical trials by facilitating more precise estimates of the treatment effect. This method uses a Bayesian design with commensurate priors derived from historical information [13, 14].

THRESHOLD CROSSING: IS THIS REALLY DIFFERENT FROM FUTILITY BOUNDARIES?

When a clinical trial is being designed, there must be extensive deliberation and discussion among clinicians, biostatisticians, and regulatory experts about the hypotheses (null and alternative), and they must reach consensus on setting appropriate efficacy and futility boundaries. The boundaries for futility can be determined from the null and alternative hypotheses, which should be defined according to power and type I error assumptions for the trial. It is not absolutely clear what threshold Eichler and colleagues refer to in the "threshold-crossing" design that they propose. Is it the critical value for the pairing of the null/alternative hypotheses? Or, is it the efficacy/futility boundary? These need to be carefully defined and clarified in the planning stages of every trial. Analysis of preliminary or

historical data can inform these decisions and refine the definition of the appropriate threshold for individual trials.

IN PRACTICAL TERMS, WHAT ARE THE TENETS FOR DEFINING THE THRESHOLD?

Investigator teams must recognize the difference between clinically relevant and significantly different effect sizes or efficacy thresholds. Simon et al. indicate that the clinical relevance of response rates may be dependent on the nature of the disease, the location of tumors, and the symptoms associated with a specific tumor [15]. Another relevant parameter is whether or not a tumor response is clinically meaningful, and that judgment can depend on the depth, duration, and type of response.

When trials are being planned, there should be a consensus among the clinicians, statisticians, and regulators in setting thresholds. Clinicians would typically prefer drugs and treatments to be highly efficacious (high "bar"); pharmaceutical companies would prefer lower efficacy thresholds so more drugs and treatments can pass the approval "finish" line, and statisticians would recommend robust designs with high power and a small probability of type I error. Prospective negotiations on endpoints are critical. Designs that incorporate large numbers of events in similar populations of patients employed as historical controls can provide the opportunity to select optimal and robust thresholds to be used for each scenario. In addition, Bayesian adaptive approaches can be used to fully take advantage of the data that accumulate in the study population as the trial is underway.

NEED FOR IMPARTIAL DATA CUSTODIANS AND VERIFIABLE PROTECTED DATA

Selection Criteria Need to Be Standardly Defined and Executed

As it is very well articulated in the Eichler et al. manuscript, the first step is to make historical clinical data available, and this is happening already. The crucial next step is to have independent non-conflicted assessors make judgments to ensure that those data are reliable, complete, and of high quality. In addition, governance committees must be established to manage, maintain, and develop proper documentation of eligibility criteria; agree upon logistics of data sharing and usage; and provide oversight of all issues associated with data banking.

Simon et al. proposed that guidelines on the methodology for prospective selection and analysis of historical control data are needed to ensure appropriate use of historical comparator groups in evaluating results from a single-arm study. These authors also add that an FDA Guidance or Best Practices document should provide such guidelines, as well as describe how adequate safety information can be developed and monitored in the post-marketing setting. Eichler and colleagues have a European perspective on this. Our perspective is that this needs to be a global effort in order for there to be international

harmonization and to ensure maximal utility of this effort. If properly conducted, the development of these guidelines will help to facilitate the use of single-arm trials that can both produce strong evidence as well as enable effective drugs to reach patients in need with maximum efficiency.

NEED FOR MATCHING TISSUE WITH DATA FOR FUTURE INQUIRIES (ORIEN)

Initiatives like the Oncology Research Information Exchange Network (ORIEN), developed by investigators at Moffit Comprehensive Cancer Center and an enlarging group of collaborators in the U.S., are being designed to provide cancer patients, investigators, and companies greater access to clinical trials specific to an individual's genomic cancer type. This initiative gathers clinical data, permits it to be updated, and also gets tumor and germline DNA for prospective tumor typing. By doing these analyses early, driving mutations can be identified, and patients with them categorized, so that when relevant trials become available, the patient can be matched to the study. The enrolled patients will also contribute to the data bank that could be used to design new trials. This collaboration and cooperation among a growing number of ORIEN centers implies that patients may not need to travel far from home to participate in studies and promises to discover targeted treatments. Research based on the ORIEN genomic data or similar efforts would result in a better understanding of cancer biology at the molecular level and hopefully would enable development of more targeted cancer treatments. This network will also offer the opportunity to identify patients with unusual mutations or subtypes who would most likely benefit from targeted therapies so that they can be rapidly enrolled in promising studies.

Data-sharing plans through networks like ORIEN will provide more reliable and extensive historical data, which will lead to better study designs and thresholds determination.

CONCLUSION

Eichler and his colleagues have made a bold proposal that is a call to international collaborations among researchers, regulators, and patients to speed the evolution of the way that we conduct clinical trials in patients with cancer in the age of genomic medicine. The author's idea about borrowing information from previous trials and historical data will be critical for the design of new "counterfactual" trials using either the frequentist or Bayesian approaches. In addition, this threshold-crossing approach requires strong planning and effective communications among all team members throughout the entire trial's process: (a) design, (b) conduct and oversight, (c) data collection, (d) data analysis, and (e) conclusion. This will result in high-quality clinical research (i.e., trials that would lead to highly effective treatments). We applaud him and his coauthors for presenting us with this challenge. As a community, we should rise to the occasion.

REFERENCES

1. Marshall G, Blacklock JS, Cameron C, et al. Streptomycin treatment of tuberculous meningitis. Lancet 1948;1:582–596.

2. Fisher B, Bauer M, Margolese R et al. Five-year results of a randomized clinical trial comparing total mastectomy and segmental mastectomy with and without radiation in the treatment of breast cancer. N Engl J Med 1985;312:665–673.

3. Venook AP, Niedzwiecki D, Lenz HJ et al. CALGB/SWOG 80405: Phase III trial of irinotecan/5-FU/ leucovorin (FOLFIRI) or oxaliplatin/5-FU//leucovorin (mFOLFOX6) with bevacizumab (BV) or cetuximab (CET) for patients (pts) with KRAS wild-type (wt) untreated metastatic adenocarcinoma of the colon or rectum (MCRC). J Clin Oncol 2014 [Epub ahead of print].

Oncologist

4. Lillie EO, Patay B, Diamant J et al. The n-of-1 clinical trial: The ultimate strategy for individualizing medicine? Per Med 2011;8:161–173.

5. Renfro LA, Shi Q, Sargent DJ. Mining the ACCENT database: A review and update. Chin Clin Oncol 2013;2:18.

6. Sargent DJ, Buyse M, Matheson A et al. ARCAD clinical trials program: An update and invitation. *The Oncologist* 2012;23:188–191.

7. American Association for Cancer Research. Project Data Sphere takes off. Cancer Discov 2016;6:464.

8. Eichler HG, Bloechl-Daum B, Bauer P et al. "Threshold-crossing": A useful way to establish the counterfactual in clinical trials? Clin Pharmacol Ther 2016;100:699–712.

9. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov 2006;5:27–36.

10. Biswas S, Liu DD, Lee JJ et al. Bayesian clinical trials at the University of Texas M. D. Anderson Cancer Center. Clin Trials 2009;6:205–216.

11. 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.

12. Hobbs BP, Carlin BP, Mandrekar SJ et al. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics 2011;67:1047–1056.

13. Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. Bayesian Anal 2012;7:639–674.

14. Hobbs BP, Carlin BP, Sargent DJ. Adaptive adjustment of the randomization ratio using historical control data. Clin Trials 2013;10:430–440.

15. Simon R, Blumenthal GM, Rothenberg ML et al. The role of nonrandomized trials in the evaluation of oncology drugs. Clin Pharmacol Ther 2015; 97:502–507.

For Further Reading:

Gideon M. Blumenthal, Paul G. Kluetz, Julie Schneider et al. Oncology Drug Approvals: Evaluating Endpoints and Evidence in an Era of Breakthrough Therapies. *The Oncologist* 2017;22:762–767.

Excerpt:

In the past few decades, with improved understanding of the genomic and immunologic underpinnings of cancer, better molecular characterization of tumors, and more precisely targeted agents, new and innovative therapeutics have altered the natural histories of certain cancer types such as chronic myeloid leukemia (CML), multiple myeloma, and melanoma. Recognizing a need to further expedite development of drugs that show promising early clinical evidence of benefit over available therapy, the U.S. Congress, in 2012, established the Breakthrough Therapy Designation program. The U.S. Food and Drug Administration (FDA) uses this program frequently for transformative therapies that show great promise in early clinical trials. . . With the Breakthrough Therapy Designation program adding to the tools that the FDA has for expediting drug development, the FDA reassessed the endpoints needed for approval of transformative therapies. Although the demonstration of an improvement in overall survival remains the gold standard for drug approval, innovation in cancer research has led to use of other endpoints in regulatory decision-making. These endpoints include substantially delaying tumor progression or extending progression-free survival, substantially reducing tumor size for a prolonged time, improving objective response rate and duration of response, or improving cancer-related symptoms and patient function.