ISSUE BRIEF

Friends of Cancer Research Annual Meeting – November 2016

Optimization of Exploratory Randomized Trials

CONTRIBUTORS

Eric Kowack
Ignyta

Lisa LaVange
U.S. FDA

Amy McKee
U.S. FDA

Cyrus Mehta
Cytel Corporation; Harvard University

Samantha Roberts
Friends of Cancer Research

Richard Simon
National Cancer Institute

Rajeshwari Sridhara
U.S. FDA

Mark Stewart
Friends of Cancer Research

Summary

In recent years, the field of oncology has benefitted from the development of several highly effective new therapies for some forms of cancer. These therapies may demonstrate profound treatment effects that are apparent in early phase clinical trials, necessitating expedient clinical trial approaches that move beyond the traditional stepwise drug development paradigm. These approaches must maintain rigor while improving efficiency to ensure that truly effective drugs can quickly reach patients in need without compromising patient safety. Different approaches may be needed for different scenarios. This panel will address potential paths forward when unexpectedly large improvements in overall survival (usually included as a secondary endpoint) are observed in early phase randomized studies, a scenario that is occurring with increasing frequency. Because these are exploratory trials, primarily initiated to guide "go/no-go" decisions in product development, they are typically not designed with the necessary statistical rigor for definitive assessments of clinical benefit. Therefore, using these trials as the basis of a regulatory decision without further study may present challenges. However, only using the data for a "go/no go" decision and initiating a separate randomized phase 3 trial may also be problematic depending on how compelling the exploratory trial results were as well as the level of unmet need in the disease under study. We will provide recommendations for the optimal conduct of early phase randomized trials, potential frameworks that can be put in place prospectively for the controlled expansion of exploratory trials, and statistical approaches that can be used by sponsors or FDA reviewers to help interpret the results in the absence of pre-specification and determine how to proceed in the event of unexpected but promising survival signals.

Exploratory Randomized Trials

Although exploratory trials are often single-arm studies, in some cases randomized trials are employed early in development with the objectives of providing proof-of-concept or generating hypotheses. In these trials, the patient population under study may be limited for safety reasons or to improve the chance of detecting an efficacy signal. The requirements for the trial's operating characteristics such as power and Type I error (concluding that a drug has a certain effect, when it, in fact, does not) may also be less restrictive than in later stage trials, or may not even be pre-specified. There may be multiple looks at the data, potentially introducing bias, and informal interim analyses with no planned adjustments to avoid inflation of the Type I error rate. In our scenario, compelling survival outcomes may be observed, but survival is not the primary endpoint and is rather a secondary endpoint. In fact, a variety of endpoints may be specified to assess pharmacological activity and tolerability and to provide early evidence of efficacy with respect to clinical or patient reported outcomes, typically with no plans to account for multiplicity due to the numerous outcomes in place. In general, these "looser" operating characteristics are accepted for exploratory trials, with the assumption that clinical benefit will be rigorously assessed in later phase trials.

Consequently, if unexpected and potentially exceptional survival signals are observed in an exploratory randomized study, these issues can lead to difficult decisions about whether to expand the ongoing trial, initiate a subsequent phase 3 trial, or seek regulatory approval. Options available to sponsors could include:

Traditional approach:

1. Use the exploratory data results solely for a "go/no go" decision and initiate confirmatory trials.

Alternative approaches:

2. Expand the exploratory randomized trial, and if the survival benefit is maintained, seek regulatory approval.

3. For exceptional survival data in exploratory randomized trial, submit for regulatory approval, and potentially initiate a phase 3 confirmatory study at the same time.

There are many factors to consider as sponsors determine how best to proceed with unexpected data that includes both statistical and non-statistical issues.

Statistical Approaches for Interpretation of Unexpected Findings

In instances where there is little to no pre-specification in the exploratory randomized trial and an unexpectedly large improvement in overall survival is observed, sponsors and the FDA can be faced with the challenging scenario of interpreting the results. We will discuss a potential statistical model that can be useful in these scenarios. This is an adaptation of a previously published Bayesian approach which accounts for the clinical significance of the results and for the fact that the survival results were unexpected in a phase 2 trial and often not specified as the primary endpoint for analysis.[1] Bayesian approaches can be very useful in looking across multiple endpoints, analysis times, or studies. These factors are not accounted for in the calculation of a "p value".

Instead of a p value, the tool described here provides a posterior probability that the treatment effect (treatment relative to control) exceeds a minimal clinically significant threshold. For survival in oncology studies that threshold might be a 20% to 30% relative reduction in the hazard of death but will depend on the disease and line of therapy. The posterior probability depends on the observed treatment effect in the clinical trial, the size of the trial, and on the prior probability distribution of the treatment effect (i.e., the likelihood of different treatment effect sizes one would expect before seeing the results of the clinical trial). It is the prior distribution which enables one to express the fact that an extreme treatment effect on survival is unexpected for a phase 2 trial with a PFS or response endpoint. The prior distributions have decreasing effect on the posterior probabilities as the sample size of the trial increases, and in the example case studies discussed below, the results are not critically dependent of the prior distribution.

The model we have investigated is based on the estimation of an unknown hazard ratio (HR) for treatment on survival. An HR of 1.0 means no treatment effect on survival and an HR of 0.75 represents a 25% reduction in the hazard of death by treatment. The prior probability of the null hypothesis of no treatment effect on survival is denoted by $1-\theta$. For an early phase 2 trial of a drug of unknown efficacy, one would generally set this null prior probability to be .90 or some suitably large figure. The prior distribution when the null hypothesis is false is based on a standard deviation parameter τ as described in the Appendix.

[1] Simon R. "[Clinical Trials and Sample Size Considerations: Another Perspective]: Comment." Statistical Science. 2000 15(2): 103-5.

To compute the posterior probability that an observed difference in survival is clinically significant, one must specify the survival results of the trial, θ, τ, and the threshold for clinical significance (e.g. 25% reduction in hazard of death). In our simulations, we have summarized the trial results for survival by indicating the observed HR and the total number of deaths observed in the trial. For details on computing the posterior probability of a clinically significant treatment effect, see the Appendix.

It is important to note that this model helps to interpret the level and confidence of evidence in trials with unexpected results, but it does not alleviate issues related to robustness of results, sensitivity analyses, uncontrolled interim looks, and trial conduct.

Application of the Bayesian Statistical Approach to Real-World Case Studies

Iniparib is an inhibitor of the enzyme poly ADP-ribose polymerase (PARP). This example is chosen because it is a well-known example where preliminary trial results indicated a significant survival advantage which generated considerable enthusiasm. However, follow-up studies failed to confirm this effect and, in fact, demonstrated that iniparib did not inhibit PARP at clinically relevant doses. Early approval based on the initial phase 2 results would have put patients at risk by exposing them to an ineffective drug.

In a phase 2 open-label, randomized study of patients with metastatic triple negative breast cancer iniparib combined with chemotherapy improved the rate of clinical benefit from 34% to 56% (P=0.01) and the rate of overall response from 32% to 52% (P=0.02).[2] The addition of iniparib also prolonged the median progression-free survival from 3.6 months to 5.9 months (HR for progression, 0.59; P=0.01) and the median overall survival from 7.7 months to 12.3 months (HR for death, 0.57; P=0.01). A subsequent randomized phase 3 trial enrolled 519 women who had previously received at least two rounds of chemotherapy. This trial was designed with overall survival and progression-free survival as co-primary endpoints and was unable to demonstrate significant improvements in these endpoints.

We applied the statistical model described above to determine how it might have influenced decision making based on the phase 2 results. The total number of deaths was not reported but was estimated to be approximately 73 from the confidence interval given for the HR. From that value and the reported HR, the Bayesian analysis was performed using θ=.9 and τ=1. The resulting posterior probability distribution for the true HR for survival is shown in Figure 1. For any HR on the x axis, the y axis shows the posterior probability that the x-axis value is the true HR. An HR value of 1.0 corresponds to no treatment benefit on survival. A vertical line is drawn at 0.70 as a potential threshold for clinical significance; that is an HR < 0.70 would represent a clinically significant treatment effect on survival. The area under the curve to the left of the vertical line is the posterior probability that the treatment effect is clinically significant. In this case that area is 0.71; and may indicate that additional data is needed to ensurethe treatment effect on survival is clinically significant. If we use a threshold of clinical significance of 0.75, the area under the curve to the left of the x-axis point HR=0.75 is 0.82. Thus, even with a threshold of clinical significance of 0.75, the data is not strongly convincing that there is a clinically significant treatment effect on survival. The posterior probability of the null hypothesis that iniparib has no effect on survival was 0.044 as can be seen by the point at an x-axis value of 1.0. The posterior probability of the null hypothesis is however not very robust to changes in the model parameters, and we do not recommend using it for decision making in this context. A posterior probability of .82 that the HR for survival is less than .75 may not be sufficiently strong to support the conclusion that the

[2] SO'Shaughnessy J, Osborne C, Pippen JE, et al. "Iniparib plus chemotherapy in metastatic triple-negative breast cancer." N Engl J Med. 2011 364:205-14.

treatment is effective to a clinically significant degree. Consequently, this approach as part of the evaluation of the study would not have suggested consideration of approval of the drug without a follow-up phase 3 trial. This use of this approach would have been appropriate to guide decision making in this example.

Olaratumab is a platelet-derived growth factor (PDGF) receptor alpha blocking antibody. Olaratumab received fast track and breakthrough therapy designation, priority review status, and accelerated approval for its use in soft tissue sarcoma (STS). This example illustrates a study where there was highly significant improvement in survival outcomes, a secondary endpoint, in an early phase trial when progression-free survival was the primary endpoint. Though the study was not necessarily designed to be a pivotal trial, it did lead to its approval.

Data came from a randomized phase 2 trial involving 133 patients with multiple subtypes of metastatic STS.[3] Patients were randomized in a 1:1 ratio to receive either combination therapy comprising of olaratumab and doxorubicin, or the standard of care treatment of doxorubicin monotherapy. Patients in the combination treatment arm had a median overall survival of 26.5 months, compared with 14.7 months for those treated with doxorubicin monotherapy (HR 0.46; P=0.0003). In contrast, progression-free survival was extended by only 2.5 months in the olaratumab arm (6.6 months versus 4.1 months). Though the primary endpoint of the study, a 50% increase in progression-free survival, was met, it was not significant by investigator assessment (HR 0.67; P=0.0615) or independent radiological review (HR ratio 0.67; P=0.1208).

To evaluate how the statistical model described above might have influenced decision making, data from the phase 2 trial were evaluated using an HR of 0.46 and 91 observed deaths (Figure 1). A vertical line is drawn at a true HR of 0.70 which might correspond to a minimal clinically significant effect. The Bayesian analysis was performed and the posterior distribution of HR for survival is shown in Figure 1. The area under the curve to the left of 0.70 is approximately 0.95. This means that there is a 95% posterior probability that the true HR is 0.70 or less indicating that the evidence is convincing and supports the FDA decision. If the line were drawn at 0.75, representing a lesser reduction in survival, the posterior probability to the left of that would be 0.98.

The posterior distribution was computed based on an assumption that the prior probability of no treatment benefit was 0.90; so the survival effect was unexpected. However, the data was sufficiently strong that with 91 deaths the high prior probability of no treatment effect is overridden by the data. We also used the parameter τ=1 for the standard deviation of the treatment effect under the alternative. The results were little changed however if we used τ=2 or 0.5.

[3] Tap WD, Jones RL, Van Tine BA, et al. "Olaratumab and doxorubicin versus doxorubicin alone for treatment of soft-tissue sarcoma: an open-label phase 1b and randomized phase 2 trial." Lancet. 2016 388:488-97.

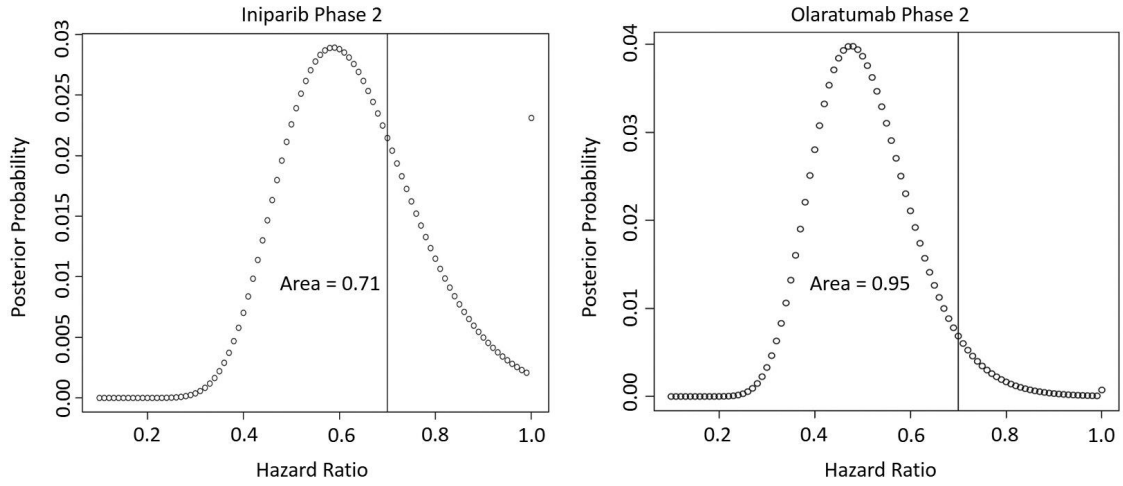

Figure 1. Statistical model evaluating data from phase 2 trial data of iniparib and olaratumab. *For any HR on the x-axis, the y-axis shows the posterior probability that the x-axis value is the true HR. An HR value of 1.0 corresponds to no treatment benefit on survival. A vertical line is drawn at 0.70 as a potential threshold for clinical significance; that is an HR < 0.70 would represent a clinically significant treatment effect on survival. The Bayesian analysis was performed and the posterior probability calculated as the area under the curve. For iniparib, there is a 71% posterior probability that the true HR is 0.70 or less indicating that additional evidence may be needed to sufficiently support the conclusion that the treatment is effective to a clinically significant degree. For Olaratumab, there is a 95% posterior probability that the true HR is 0.70 or less indicating that the evidence is convincing, but would need to be considered as part of the evaluation of the entire study.*

Additional Factors to Consider When Interpreting Findings from an Exploratory Trial

As sponsors navigate these various options, factors other than statistical analyses will also need to be considered, such as the strength of evidence from the phase 2 study, the feasibility of restarting enrollment once preliminary results are known, the role of an independent monitoring committee in triggering further enrollment, and potential drifts in patient population due to expanding the number or location of study sites.

Additionally, recruitment for a subsequent study may be difficult, once the early trial's results are publicly available. Other issues critically important include chemistry, manufacturing, and controls (CMC) readiness and the adequacy of the safety database

Standard considerations around the interpretation of results from any randomized clinical trial, such as those described in ICH E9, apply across all trials in general.[4] Several considerations are highlighted here as particularly relevant in the context of observing an unexpectedly large benefit in survival in an exploratory randomized study. The extent to which many of these can be adequately addressed will help determine the interpretation of the strength of the results, and hence help determine the best appropriate path forward.

[4] Guidance for Industry: E9 Statistical Principles for Clinical Trials. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073137.pdf (accessed Oct. 31, 2016)

Potential sources of concern in the design, conduct, analysis, and interpretation of the results include:

- Multiplicity due to overall survival typically being a secondary endpoint, multiple interim analyses; other possible sources of multiplicity such as multiple arms, statistical methods for analyzing the data, and the primary analysis population
- The robustness of the survival results, in view of the likely small sample size of an exploratory study, especially as it relates to potential imbalances in important prognostic factors, the impact of post-discontinuation anti-cancer therapy, and the ability to evaluate consistency across subgroups
- The conduct of the study, including the level of blinding, the sufficiency of the description of the process and procedures for examining data, including informal and formal analyses, and whether crossover to the experimental therapy has been allowed

In contrast to confirmatory trials, exploratory trials often have less pre-specification allowing for more flexibility in the design and conduct of the study. This is often appropriate depending on the specific objectives of the study, the extent to which the safety and efficacy of the experimental treatment is understood, and the extent of ongoing biomarker evaluation. However, flexibility in the design and conduct of these studies can pose challenges to evaluating the results in a regulatory approval setting.

There is benefit in sponsors considering the incorporation of some standard features found in confirmatory trials when designing, conducting, and analyzing exploratory randomized studies. The addition of some pre-specification and rigor around the timing and assessment of interim analyses or the planned timing of the final overall survival analysis can often bring additional scientific rigor with little downside. Furthermore, recently observed faster-than-expected enrollment in studies (e.g., checkpoint inhibitor trials) coupled with the time lag on obtaining survival data may make it necessary to add language that allows extended enrollment if early efficacy results are quite favorable to help reduce delays in global protocol amendment.

Opportunities for Prospective Planning and Trial Expansions

In instances where there is a promising benefit in overall survival but the data is not quite strong enough or requires additional patient populations before submission for regulatory approval, a trial expansion may be one approach to efficiently collect more evidence and data. The points to consider in the section on "Additional Factors to Consider When Interpreting Findings from an Exploratory Trial" are applicable both to the original study (in terms of whether it provides a solid basis for expansion) as well as to the design of the expansion. It is also important to note that having a prospective plan in place does not guarantee positive data; it simply improves the ability to appropriately interpret the data.

Adaptive designs that prospectively incorporate a trial expansion, using appropriate statistical methods to control the Type I error rate are available and are described in the extensive literature on this topic. These designs are valuable in the context of designing a Phase 2/3 study. Few exploratory studies are designed in this way, and it would not be desirable or feasible to design all randomized phase 2 exploratory studies in this manner. However, trials for new classes of drugs that have shown exceptional promise in early exploratory settings in terms of objective

tumor response (e.g., immune checkpoint inhibitors), may warrant greater consideration for provisions to be in place at the start of the trial.

Some pre-planning around potential outcomes, associated decisions, and resultant actions into the protocol can be beneficial in terms of reducing the need for protocol amendments and improving the understanding of the operating characteristics of the study. At this stage of development though, maintaining the ability of the sponsor to also incorporate ongoing, and maybe unexpected, learnings from this trial and external data will continue to be important.

A major concern in regards to unplanned adaptations in clinical trial design or conduct during the trial is the loss of control over Type I error rates. As a starting point, it is worthwhile to consider the simplest case of expanding the trial (either in terms of number of patients or number of events) to collect additional overall survival data in the same patient population as defined by the protocol. In the context of this paper, it is assumed that a large overall survival has been observed in the original (unexpanded) study, and what might be defined as an unexpectedly large effect, is also likely to be statistically significant at the 5% level. With a statistically significant effect on overall survival acting as a gate-keeper, a study expansion in the same population would not inflate the Type I error rate above the standard 5%. Methods to assess the evidence, such as the interpretation of the p-value (and point estimate), from the expanded trial with reference to an even higher bar for remarkable results (such as that outlined in the Bayesian statistical approach above as guidance) will then need to be evaluated.[5] In instances where the threshold for statistical significance is not met, it may be desirable to increase patient follow-up time or recruit additional patients. However, in scenarios where statistical significance is not met but is still promising, it may be necessary for sponsors to have some pre-specification in place prior to unblinding (e.g., number of overall survival events to collect if the trial were to proceed) and to utilize adaptive methods to ensure statistical validity is maintained.[6,7]

Expansion from a study that made significant alterations to the patient population based on results might be more challenging and could lead to significant bias. This could be somewhat mitigated if the study was originally designed to evaluate the populations—for instance, a study designed to evaluate in a specific biomarker positive and negative population, and subsequently dropping the biomarker negative population for the expansion.

Further, expansion based on data collected from a study with insufficient quality is also a concern. Making alterations to the choice of study endpoints, patient populations, or treatment allocations based on unblinded interim results may lead to biases toward favorable study outcomes or add unwanted variability to the study characteristics. A framework designed with operating characteristics that permit trial expansion or potential drug approval depending on the outcome of the exploratory trial may help minimize uncertainty in the assessment of the results of these types of trials.

A template for pre-specified expansion could include options to modify inclusion/exclusion criteria to increase generalizability of data within a single trial, or the template could allow for adaption of specific trial features, if warranted based on accumulating data, without starting a new trial. Additionally, it could include options to increase the follow-up time and the sample size possibly through expanding the number and geographic spread of trial sites. This framework could guide the development and use of pre-specified triggers for expansion in the event of

[5] Gao P, Liu L, Mehta C. "Exact inference for adaptive group sequential designs." Stat Med. 2013 Oct;32(23):3991-4005.

[6] Denne JS. "Sample size recalculation using conditional power." Stat Med. 2001 Sep;20(17-18):2645-60.

[7] Müller HH, Schäfer H. "Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches." Biometrics. 2001 Sep;57(3):886-91.

observing a surprising survival benefit, the statistical considerations necessary for a robust analysis, and the operating characteristics necessary to minimize uncertainty in the results in these types of trials.

Conclusions

Oncology drug development is benefiting from improved research capabilities and techniques that help to better identify appropriate patient populations for clinical trials. Thus, scenarios where unexpectedly large improvements in overall survival observed in exploratory randomized studies are becoming more frequent as our scientific understanding continues to advance. New therapies have necessitated the need for innovative clinical trial designs and expedient pathways for drug approval. It is clear that full pre-planning for registration for every early phase trial is not feasible or even possible because exploratory trials need to be able to have a reasonable sample size for the phase of development; be able to address multiple exploratory objectives; and if needed, evolve in response to the data being generated. However, both sponsors and the FDA can be better equipped to handle and evaluate trials with unexpectedly large improvements in overall survival. It is important to get the most information from data collected in these trials. Therefore, we have provided considerations for sponsors to consider at the design and conduct stages and for both sponsors and the FDA to use to help evaluate these types of results, which include a Bayesian analysis approach to assist with decision-making.