Contemporary Clinical Trials xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Statistical design for a confirmatory trial with a continuous predictive biomarker: A case study

Adarsh Joshi*, Jenny Zhang, Liang Fang

Gilead Sciences Inc., Foster City, CA, USA

ARTICLE INFO

Keywords:
Group-sequential design
Multiplicity
Personalized medicine
Biomarker cutoff selection
Subgroup identification
Clinical trial optimization

ABSTRACT

With targeted therapies, it is often hypothesized that their effect may be specific to the subpopulation in which the target pathway is activated. We consider the problem of designing a confirmatory trial when the biological hypothesis of the experimental therapy is strongly supported by the pre-clinical data but limited clinical data is available to pre-define a subpopulation based on a biomarker with continuous values. The study design is further complicated if interim evaluations of the biomarker-based subpopulations are also being considered. We compared several strategies, including a naïve threshold nomination approach, a modification of the "explore and confirm" strategy proposed by Friedlin et al. (2005), and a novel biomarker sequential testing approach, motivated by the "general bivariate normal method" discussed by Wang el al. (2007), and further discussions in Spiessens and Debois (2010) and Holmgren (2017), in a setting where all-comers and biomarker subpopulation evaluations can be performed at interim analyses as well as the end of study. Based on extensive simulations, we concluded that the novel biomarker sequential testing approach out-performed other strategies when there was limited prior information for biomarker threshold determination. This design was implemented in a recently completed clinical trial of simtuzumab (RAINIER study) and provides a useful case study for designing future confirmatory clinical trials of novel targeted therapies.

1. Introduction

With targeted therapies, it is often hypothesized that their effect may be specific to the subpopulation in which the target pathway is activated (e.g., Trastuzumab is only indicated in HER2-overexpressing breast cancers [1]). Hence, subpopulation analyses in clinical trials are becoming increasingly prominent as more targeted therapies are being developed.

The confirmatory clinical trial setting requires that the analysis population(s), endpoint(s), and statistical testing procedures be fully pre-specified. A sponsor can choose to conduct a clinical trial only in a subpopulation of patients, e.g. [1] or consider evaluation of the patient subpopulation as a co-primary endpoint along with evaluation of the all-comers population. The development strategy that the sponsor may pursue will be dependent on several factors, including the available statistical and biological evidence from pre-clinical studies and early phase clinical trials. Regardless of the development strategy, a threshold is needed to define the patient subpopulation. However, there are many challenges with pre-specifying a threshold value for a continuous biomarker in a confirmatory clinical trial. In [2] it is shown that, mathematically, the number of clinical events required to obtain a precise estimate of the biomarker threshold value is often too large for Phase I/II clinical trials prior to the initiation of the confirmatory trial. In addition, data generated from early phase trials may not be as relevant or informative as needed for defining a threshold value for a confirmatory trial. For example, early phase trials may have used an earlier version of the companion assay that is to be used in the confirmatory trial, a different comparator in the control arm, a broader or narrower patient population, or a different dosing regimen, etc. Thus, for a trial designer, it is common that we have little relevant prior information to pre-specify a biomarker threshold value for patient selection or stratification when designing a confirmatory trial.

Herein, we consider the problem of designing a confirmatory clinical trial when the preclinical data strongly supports the biological hypothesis of the experimental therapy, but limited data is available to pre-define a subpopulation based on continuous biomarker values in the context of a recently completed clinical trial of a novel investigational drug simtuzumab (the RAINIER study). The study design is further complicated because interim evaluations of the treatment benefit were also being considered.

There are multiple proposals for biomarker driven clinical trial designs in the literature. These designs range from classic designs to

* Corresponding author.
E-mail address: adarsh.joshi@gilead.com (A. Joshi).

http://dx.doi.org/10.1016/j.cct.2017.05.010
Received 8 November 2016; Received in revised form 3 May 2017; Accepted 14 May 2017

more innovative adaptive designs. In the category of classic designs, examples include biomarker-enrichment design that only enrolls patients from the "biomarker positive" ([1], also discussed in [3]), biomarker stratified design that stratifies enrollment based on the binary biomarker status (positive or negative) [3], sequential testing strategy design [3] that is a traditional design for randomized clinical trials but employs a sequential testing procedure, and some variants of these designs. For adaptive designs, examples include adaptive enrichment design [4], adaptive accrual design [5], adaptive threshold design [6], adaptive signature design [7], etc. Although adaptive designs have certain statistical advantages that can help maximize the success probability of a confirmatory trial, this group of designs have not been fully utilized by the pharmaceutical industry mainly due to operational and regulatory concerns. Operationally, adaptations during the conduct of a clinical trial present challenges in sample and data collection, site enrollment, etc., although most of these challenges can be overcome with careful planning. In addition, although there are increasing experience and interests from regulatory bodies on adaptive designs, industry sponsors tend to be more conservative and are hesitant to proceed with designs that do not have many precedent successful examples. In [8] FDA CBER's experience with adaptive design clinical trials was discussed and was noted that adaptive designs submitted to FDA CBER between 2008 and 2013 are mainly for Phase II studies or fall into the category of "well-understood" adaptive designs such as group sequential design or sample size re-estimation. Based on this report and to our knowledge, unfortunately, the adaptive biomarker designs mentioned above have rarely been implemented in a confirmatory trial that is designed for the purpose of obtaining a drug label. For these reasons, when designing RAINIER study, we primarily focused on exploring design options in the category of classic designs which are easier to implement and have lower regulatory risks, however, may not have the best statistical performance.

In order to maintain a strong control over the type I error, an appropriate family-wise error rate (FWER) control method needed to be incorporated in the trial. Commonly used FWER control methods, such as Bonferroni, Holm, Hochberg, may be too conservative for trials that have hypothesis tests on both all-comers and subpopulations at study interim as well as study final because these methods ignore the positive correlations between the tests. The "general bivariate normal method" first discussed in [5] and further developed in [10], provides a framework for evaluating multiple populations in a clinical trial while taking into consideration the correlations between the associated tests (see also [9]).

Herein, we will discuss comparison of three different approaches for testing patient subpopulations as co-primary endpoints in the context of the RAINIER study with multiple testing performed according to this "general bivariate normal method" The first approach is to nominate a specific biomarker value, based on the limited prior information, to define the subpopulation in the interim and final analyses in the study protocol or statistical analysis plan at the design stage. The second approach is an "explore-and-confirm" strategy as provided in [7], but with a revised multiplicity adjustment method (a Bonferroni adjustment was used in [7]). Lastly we considered a novel biomarker sequential testing approach by extending the proposal in [9] to evaluate multiple patient subpopulations at both the interim and the final.

2. RAINIER study background and design

Simtuzumab (GS-6624) is a monoclonal antibody which inhibits LOXL2, an enzyme that has been implicated in several types of cancers and fibrotic diseases. The RAINIER study (GS-US-322-0207) was designed to evaluate the efficacy of simtuzumab in patients with idiopathic pulmonary fibrosis (IPF), a progressive, fibrotic lung disease that affects middle-aged and elderly patients. RAINIER was a randomized (1:1), double-blind, placebo-controlled study with a planned sample size of 500 patients. The primary endpoint was progression-free survival (PFS), defined as the time from randomization to disease progression or death, whichever occurred earlier. An interim analysis for efficacy was planned after 200 events were observed and the final efficacy analysis was planned after 250 events were observed.

Due to its biological mechanism and literature data suggesting the prognostic value of LOLX2 biomarker, the sponsor was interested in exploring benefit and risk profiles in both all comers and subpopulations with high LOLX2 values. However, due to lack of samples from a control group in early clinical trials of simtuzumab and the fact that the market-ready assay was not available at the time of initiation of this study, specifying a threshold for defining this subpopulation was a major challenge. In addition, the sponsor would like to maximize the power to detect a treatment effect in the test for all comers, preserve the ability of stopping the trial for a drug label application at the interim analysis, while still maintaining a certain power for detecting a large treatment effect in the LOLX2 subpopulations. This "wish list" is common for drug developers when the biomarker hypothesis is still under investigation (i.e., data evidence suggesting LOLX2 is a predictive biomarker have emerged but have not been fully established), and presented a unique challenge for statisticians to propose innovative clinical trial designs that could meet these goals.

The RAINIER study was terminated at the interim due to lack of efficacy.

3. Clinical trial design options considered

We considered three approaches to determining threshold values for defining biomarker subpopulations. We refer to these as the nomination approach, the explore-and-confirm approach, and the biomarker sequential testing approach. We evaluated the performance characteristics of these approaches in a simulation study to inform the decision-making of the design.

3.1. Nomination approach

The biomarker threshold defining the subpopulation to be tested at both the interim and the final is set at a pre-specified biomarker value at the design stage of the study. As discussed in the Introduction section, it is often not possible to obtain a reliable biomarker threshold value for this purpose. Thus, this approach, although operationally straightforward, could be rather risky.

3.2. Explore-and-confirm approach

This approach is motivated by [7]. First, a nomination approach is used to pre-specify a biomarker threshold value at the interim analysis. If the efficacy or futility boundaries are not crossed at the interim analysis and the trial proceeds to the final analysis, a subset of patients are randomly assigned to a Marker Exploratory (ME) group while the rest patients are assigned to a Marker Confirmatory (MC) group. Only the MC set is utilized for the hypothesis testing in the subpopulation, while the ME set is used to search for the best threshold to pre-specify the subpopulation to be tested in the MC set. Fig. 1 shows how this design would be conducted in a setting like that of the RAINIER study.

Similar to the justification in [7], since the ME patients would not be included in the hypothesis testing of the subpopulation at the final