Contemporary Clinical Trials 37 (2014) 148–154

Contents lists available at ScienceDirect

Contemporary Clinical Trials

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

A two-stage patient enrichment adaptive design in phase II oncology trials

James X. Song *

100 Tice Blvd, Woodcliff Lake, NJ 07677, USA

ARTICLE INFO

Article history:
Received 27 June 2013
Received in revised form 19 November 2013
Accepted 8 December 2013
Available online 14 December 2013

Keywords:
Phase II trials
Oncology
Predictive marker
Adaptive design

ABSTRACT

Illustrated is the use of a patient enrichment adaptive design in a randomized phase II trial which allows the evaluation of treatment benefits by the biomarker expression level and makes interim adjustment according to the pre-specified rules. The design was applied to an actual phase II metastatic hepatocellular carcinoma (HCC) trial in which progression-free survival (PFS) in two biomarker-defined populations is evaluated at both interim and final analyses. As an extension, a short-term biomarker is used to predict the long-term PFS in a Bayesian model in order to improve the precision of hazard ratio (HR) estimate at the interim analysis. The characteristics of the extended design are examined in a number of scenarios via simulations. The recommended adaptive design is shown to be useful in a phase II setting. When a short-term maker which correlates with the long-term PFS is available, the design can be applied in smaller early phase trials in which PFS requires longer follow-up. In summary, the adaptive design offers flexibility in randomized phase II patient enrichment trials and should be considered in an overall personalized healthcare (PHC) strategy.

1. Introduction

The core of the personalized healthcare (PHC) strategy is the identification of predictive biomarkers that can predict the differential efficacy of a particular therapy. The development and validation of such biomarkers, however, can be challenging. It requires the generation of the biomarker hypothesis early on, so that ideally a companion diagnostic can be tested in early drug development phase (phase I/II). Zelboraf and its companion diagnostic test Cobas which were concurrently approved in melanoma patients with Braf V600E mutation by FDA in 2011 is a good example of such co-development [1]. FDA 2012 draft guidance document on enrichment strategies for clinical trials [2] recommends the use of a patient enrichment strategy in early effectiveness studies, because it can provide clinical proof of concept that might otherwise be missed in the general, unselected population. Adaptive designs which include a prospectively planned opportunity for modification of the study design and hypotheses based on an interim analysis of patient data are most appropriate in such trials [3].

Recently, a number of adaptive designs have been proposed to incorporate a predictive marker in randomized phase III confirmatory trials. Wang, O'Neill and Hung [4] proposed a two-stage adaptive design, which evaluates treatment effects in overall population (F) and a biomarker positive subgroup (S) simultaneously. Multiplicity is adjusted by alpha allocation according to a bivariate normal distribution between the two test statistics from F and S. Using closed testing methods [5,6] to control type I error, Brannath et al. [7] described a three-stage adaptive design where Bayesian predictive powers are used to estimate the probability of success to guide the decision making. Both adaptive designs make an interim decision on whether the original study plan is to be maintained or is to be modified to recruit only the biomarker positive patients in later stage(s) and to test hypothesis only in S. Several traditional designs without adaptive features can also be applied to biomarker-driven confirmatory trials, with or without a pre-defined cut point [8–12].

* Tel.: +1 201 746 2060.
E-mail address: james_song@eisai.com.

1551-7144/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cct.2013.12.001

Compared to phase III trials, phase II trials are smaller, usually with fewer than 200 patients. A common challenge for patient enrichment adaptive design in the phase II setting is to make an adaptation decision based on the limited information available at the interim analysis. The problem is exacerbated when the primary endpoint requires long-term follow-up. One solution may rely on using a short-term response that correlates well with the long-term primary endpoint. Huang et al. [13] proposed a Bayesian approach that models the relationship between short-term tumor response and long-term overall survival in leukemia patients to facilitate the decision making in an outcome-adaptive randomization trial. They have shown that their approach leads to a higher probability of correctly selecting a more efficacious treatment arm based on posterior probability, as compared to a design without a short-term response biomarker.

In this manuscript, we explore in a randomized phase II trial the adaptive design approach proposed by Wang et al. [4]. The Bayesian model by Huang et al. [13] is also incorporated into the two-stage adaptive design as an extension to trials with a primary endpoint that requires long-term follow-up. The example used is taken from an oncology trial; but the method can be applied to other therapeutic areas as well.

2. A motivating example

A randomized phase II trial in metastatic hepatocellular carcinoma (HCC) patients in either the 1st line setting or the 2nd line setting is considered to investigate the efficacy and safety of an experimental antibody (EXP). Based on the preclinical and phase I data, the responsiveness to EXP appears to depend on the expression level of its binding target protein, which is measured by immunohistochemistry (IHC) from 0 to 3+. Patients with 0 IHC showed no anti-tumor activity, as expected, due to lack of target expression. Patients with low expression (IHC = 1+) had a short treatment duration and short time to progression. The data suggested that the responder population for the new therapy is most likely to be tumors with high target expressions (IHC = 2+/3+). Although one might hypothesize that strong expression is a prerequisite for activity, it has not yet been rigorously tested. The goal of the trial design is to demonstrate a high degree of benefit in a targeted subset of the HCC population, while not prematurely excluding those patients whose tumors had weak target expression, identified by a rigorous, central diagnostic assessment. The design should differ from an "all comers" design by virtue of statistical power and robust evaluation of the 1+ patient population, while minimizing the likelihood of exposing large numbers of patients to a potentially ineffective therapy.

A prototype diagnostic assay and its scoring algorithm have been developed and available at the beginning of the trial for patient screening. The prevalence of 0 to 3+ IHC score in all HCC patients were estimated as 30% (0), 20% (1+), 20% (2+) and 30% (3+) from literature review, which have been confirmed in our own tumor biopsy study. It was also decided that a strong PFS HR in a target population is a prerequisite for further development of the agent.

Based on the above considerations, a two-stage adaptive design is considered to allow for a more informative evaluation in patients with potentially different degrees of benefits. The efficacy is evaluated in two related study populations: 1) patients with any target expression 1+/2+/3+ (F) and 2) patients with medium/high expression 2+/3+ (S). The study will also determine the appropriate patient population for the subsequent pivotal trial if treatment benefit is observed in this trial. The adaptive design is chosen to minimize the number of patients exposed to ineffective therapy and to maximize the power of the study to demonstrate benefit in the appropriate target population.

In the 2nd line trial, EXP will be compared to Placebo (PLA). The required number of PFS events can be observed rather more quickly as the median PFS are only 2 months in PLA arm and 4 months in EXP arm. The short PFS allows a timely interim decision to be made without causing much disruption in trial operation. In the 1st line trial, the standard of care therapy sorafenib (SOR) will be the comparator. With longer median PFS in the 1st line trial, it would take much longer time to accumulate the required number of PFS events at the interim analysis. In Section 3.2, the Bayesian model proposed by Huang et al. is applied to the adaptive design. It is shown to be a valuable tool in facilitating the interim decision making when the number of observed primary endpoint events is low.

3. Design details

3.1. HCC trial in the 2nd line setting

The two null hypotheses of interest are as follows,

H_F: PFS $HR_{123+} = 1$ in patients with 1+/2+/3+
H_S: PFS $HR_{23+} = 1$ in patients with 2+/3+

The multiplicity adjustment was done by alpha allocation. Under null hypotheses, the two Z statistics from log rank test in each population (Z_F and Z_S) follow a bivariate normal distribution with zero mean vector and covariance $(N_S/N_F)^{1/2}$ where N_S and N_F are numbers of PFS events in S and F. Assuming the number of PFS events in 2+/3+ is 67% of all biomarker positive patients based on the prevalence data, alpha levels of 0.011 and 0.02 are specified in H_F and H_S, respectively, in order to control the overall type I error at 0.025. Most of the alpha (0.02) is spent on testing H_S since 2+/3+ patients are considered most likely to benefit from the new therapy. The final PFS analysis will take place after the numbers of PFS events defined in Table 1 are observed

Table 1
Decision rules at the end of stage 2, assuming no futility for H_F in stage 1.

Analysis	HR	# of PFS events	# of patients	Alpha	Power
H_F	0.575[a]	112[b]	150[b]	0.011	0.68
H_S	0.5	79[b]	108[b]	0.02	0.80

[a] Assuming HR = 0.5 in 2+/3+ and HR = 0.8 in 1+, the HR in 1+/2+/3+ is calculated based on the estimated prevalence of target expression.

[b] For a given HR, alpha and power levels, the number of PFS events using a log rank test and its corresponding number of patients are calculated based on the following assumptions: median PFS of 4 months and 2 months in EXP and PLA arms, 20 month enrollment with an additional 2 month follow-up, a 2:1 randomization in EXP:PLA, and a 5% attrition rate.