Contemporary Clinical Trials 71 (2018) 181–185

Contents lists available at ScienceDirect

Contemporary Clinical Trials

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

Adaptive expansion of biomarker populations in phase 3 clinical trials

Cong Chen[a,*], Xiaoyun (Nicole) Li[a], Wen Li[a], Robert A. Beckman[b]

[a] Biostatistics and Research Decision Sciences, Merck & Co., Inc., Kenilworth, NJ 07033, USA
[b] Departments of Oncology and of Biostatistics, Bioinformatics, and Biomathematics, Lombardi Comprehensive Cancer Center and Innovation Center for Biomedical Informatics, Georgetown University Medical Center, Washington, DC 20007, USA

ARTICLE INFO

Keywords:
Adaptive design
Bonferroni correction
Intermediate endpoint
Predictive biomarker
Two-stage design

ABSTRACT

It is well documented in this genomic era that an investigational new drug may have greater treatment effect in a biomarker positive population than in the biomarker negative population. However, limited by preclinical data and early phase clinical data, a lot of Phase 3 confirmatory trials are initiated without fully understanding the biomarker effect. In this article, we will investigate the impact of adaptive population expansion on the overall Type I error in two statistical designs. The endpoint for making the adaptive decision can be different from the primary endpoint of the study. The first design allows expansion of study population from biomarker positive patients to all-comers if the treatment effect in the biomarker positive population is more impressive than expected, suggesting broader activity of the study drug. We show that, under this design, the trial outcome can be tested at the desired alpha level without inflating the Type I error when the adaptive decision is based on the primary endpoint of the study or based on an endpoint non-negatively correlated with the primary endpoint, an assumption that generally holds in practice. The second design allows addition of biomarker positive patients in an all-comer study if the treatment effect in the biomarker negative population is less impressive than expected, suggesting lower probability of success in the all-comer population. We show that, under this design, the trial outcome can always be tested at the desired alpha level without inflating the Type I error.

1. Introduction

With increasing knowledge of molecular biology and human disease in the genomic era, a number of predictive biomarkers are being investigated in drug development. Limited by preclinical data and early phase clinical data, a lot of Phase 3 randomized confirmatory trials are initiated without fully understanding the biomarker effect. The biomarker measurements can be categorical or continuous. The treatment effect is assumed to increase with the biomarker level in this article. We will focus on studies with only two non-overlapping biomarker subpopulations, a biomarker positive population with higher biomarker level and a biomarker negative population with lower biomarker level. The cut-point for separating the two subpopulations is pre-specified and well-defined at the time of Phase 3, but it is often based on a short-term endpoint (e.g., tumor response rate in oncology) using data from single arm studies. It is unclear how the biomarker cut-point correlates with the treatment effect based on a long-term endpoint (e.g., overall survival in oncology). The treatment by biomarker interaction effect is assumed to be quantitative (i.e., the study drug benefits both biomarker subpopulations but less positive or zero in the biomarker negative population than in the biomarker positive population) but not qualitative (i.e., the study drug is beneficial to the biomarker positive patients but detrimental to the biomarker negative patients). Epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors benefited non-small cell lung cancer (NSCLC) patients with EGFR mutation but not others, and demonstrated qualitative interaction when compared with standard of care (SOC) as a monotherapy [10,15]. Response rate to pembrolizumab (a PD-1 checkpoint inhibitor) in NSCLC patients increased linearly with PD-L1 expression level without a clear inflection point [8]. A quantitative interaction effect with PD-L1 expression may be expected for similar PD-1 checkpoint inhibitors especially when testing combination with SOC vs SOC alone in NSCLC. Under the quantitative interaction assumption, a lower than expected treatment effect in one population, which may be due to an underestimation of the treatment effect in the control arm, may suggest a lower than expected treatment effect in the other, and vice versa.

To mitigate the risk of Phase 3 failure, two different approaches to population selection are often implemented in practice. One approach is to only enroll the biomarker positive patients. KN-024 [17], a randomized Phase 3 study of pembrolizumab in 1st line NSCLC patients, took this approach by only enrolling patients with IHC PD-L1 expression $\geq 50\%$ despite the evidence from other studies that the drug may

* Corresponding author at: MAILSTOP UG-1CD44, 351 North Sumneytown Pike, North Wales, PA 19454, USA.
E-mail address: cong_chen@merck.com (C. Chen).

https://doi.org/10.1016/j.cct.2018.07.001
Received 14 November 2017; Received in revised form 3 April 2018; Accepted 4 July 2018
Available online 05 July 2018
1551-7144/

also benefit patients with lower PD-L1 expression level. The other approach is to enroll an all-comer population consisting of both biomarker positive and negative patients, and the Type I error rate is often split between the biomarker positive population and the all-comer population so that the trial can be declared positive in either population. This split may be optimized to maximize the study power as appropriate [4]. A special alpha-splitting method is to analyze the biomarker positive population first at the full alpha level and, if positive, followed with the analysis of the all-comer population (i.e., a step-down procedure). KN-010 [9], a randomized Phase 3 study of pembrolizumab in 2nd line NSCLC patients, took this approach by enrolling a patient population with lower PD-L1 expression level ($\geq$1%) than in KN-024. The randomization was stratified by PD-L1 expression level ($\geq$50% or $<$ 50%). Patients with PD-L1 expression level $<$ 1% were not enrolled in the study, which makes it short of a bona-fide all-comer study. In this article, we will ignore the difference for ease of the presentation and would consider KN-010 an all-comer study (in patients with PD-L1 expression $\geq$1%). The prevalence of the biomarker positive signature in the all-comer population can take any value from 0% to 100% in theory. When the prevalence is high, barring clear lack of activity in the biomarker negative population, there is less incentive to incorporate a predictive biomarker into study design in practice. For the purpose of this article, we focus on low to moderate prevalence (e.g., $<$ 50%), a more common scenario of interest in oncology drug development.

The first approach of only enrolling biomarker positive patients is taken under the premise that inclusion of the biomarker negative patients may reduce the study power in the all-comer population. However, this decision may be revisited during the trial. In case the treatment effect in the biomarker positive population is more impressive at an interim analysis than expected, suggesting that the study drug may benefit more patients than initially expected, the study population will be expanded to enroll an all-comer population (or a broader population with lower biomarker cut-point as in KN-010). This is one of the adaptive biomarker threshold designs considered in [1,13], but is otherwise less investigated. The second approach is taken under the premise that inclusion of the biomarker negative population will not substantially reduce the study power in the all-comer population. However, if the treatment effect in the biomarker negative population is less impressive at an interim analysis than expected, suggesting the study may have lower probability of success in the all-comer population than initially expected, additional biomarker positive patients will be enrolled to ensure adequate power in this population. This design has been less investigated in the literature as well, as compared to an alternative design that terminates the enrollment of biomarker negative patients during the trial. Termination of biomarker negative patients eliminates the chance for declaring a positive outcome in the all-comer population. Barring a detrimental effect in this population, it is a challenging business decision to make because the trial is often in an advanced stage by the time of the decision and there is little cost-saving afterwards. Moreover, early data may not predict clinical outcome and it is a risky proposition to quit prematurely. Finally, it is well established that early termination of non-performing subpopulations inflates the overall Type I error and the nominal alpha at the final analysis has to be adjusted downward. All these considerations mitigate against the implementation of this alternative adaptive design in practice. Early research work on biomarker subpopulation de-selection can be found in [19,20] and more recent work can be found in [11,12].

In this article, we will investigate the impact of adaptive population expansion on Type I error in the above two approaches. The interim decision under the two designs can be based on the primary endpoint of the trial or any other endpoint. In practice, use of an intermediate endpoint sensitive to intervention is essential to cost-effective adaptations. A general discussion about the utility of early endpoint can be found in [2]. Unless otherwise specified, X denotes the endpoint for adaptation decision, Y denotes the primary endpoint of the Phase 3 study, and Z denotes the quantile of a standard normal variable at a respective level.

We will show in Section 2 that, under the first approach (addition of biomarker negative patients), the trial outcome can be tested at the desired alpha level without inflating the Type I error when the adaptive expansion decision is based on the primary endpoint of the study or based on an endpoint non-negatively correlated with the primary endpoint, an assumption that generally holds in practice. We will also show in Section 3 that, under the second approach (addition of biomarker positive patients), the trial outcome can always be tested at the desired alpha level without inflating Type I error, irrespective of which endpoint is used for the adaptive expansion decision. Section 4 will conclude with discussions. Immunotherapies are revolutionizing cancer treatment. The motivating hypothetical trials in the following presentation are based on our experience in developing new immune-oncology therapies.

2. Expansion of a biomarker positive study to an all-comer study

A hypothetical 1:1 randomized Phase 3 oncology survival trial is conducted in a biomarker positive patient population. The study targets to enroll 350 patients in 15 months and completes after 230 death events are observed. With 230 events, the study has approximately 90% power to detect a hazard ratio of 0.65 at a Type I error of 0.025 (one-sided). An interim analysis is conducted after the first 150 patients are enrolled. At this analysis, if the test statistic for X does not exceed a pre-specified threshold C, the study will finish as planned and the null hypothesis will be tested at the usual 0.025 level. But, if the test statistic for X does exceed the threshold, suggesting greater effect than planned for, the study will be expanded to enroll an all-comer population. The accrual of the study still completes after 350 biomarker positive patients are enrolled as originally planned. The prevalence of biomarker positive patients is expected to be 50%. This means that an unbiased sample of the all-comer population consisting of approximately 200 biomarker positive patients and 200 biomarker negative patients will be enrolled after the expansion. As a result, the study will now enroll approximately 550 patients in total. The 150 biomarker positive patients enrolled before the adaptation decision will not be included in the analysis of all-comer population to avoid bias of the treatment effect estimation in this population due to disproportional sample of the biomarker positive patients. However, the 150 biomarkers positive patients will be included in the analysis of biomarker positive population. A usual alpha-splitting strategy is considered in case of expansion, which allocates α_1 (>0) to the biomarker positive population (N = 350) and α_2 (>0) to the all-comer population (N = 400) such that $\alpha_1 + \alpha_2 = 0.025$ (i.e., Bonferroni correction). With an addition of 200 biomarker negative patients to the study, it now has a chance to declare success not only in the biomarker positive population as originally planned but also in the all-comer population. Since a smaller nominal alpha is used for hypothesis testing in the biomarker positive population, the power for this population is now lower than planned for. But it may still remain adequate because the actual treatment effect may be greater than expected. Moreover, the loss of power in the biomarker positive population is offset by the gain of power in the all-comer population as a positive outcome in the all-comer population leads to a broader label indication, rendering power loss in the biomarker positive population less a concern. Compared to the conventional approach that conducts a separate study in an all-comer population after a positive outcome from the biomarker positive study, the greatest benefit of the adaptive design is to accelerate the development timeline for a broader label and reduce the sample size for the overall program.

The primary analysis for the biomarker positive population is still based upon 230 events as pre-specified. In order to have adequate number of events in the all-comer population, the accrual of this population can continue as needed after 230 events in the biomarker positive population are observed. And in this case, the primary analysis