Contemporary Clinical Trials xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Contemporary Clinical Trials

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

The application of group sequential stopping boundaries to evaluate the treatment effect of an experimental agent across a range of biomarker expression

Eric Holmgren

Oncomed Pharmaceuticals, Biostatistics, 800 Chesapeake Drive, United States

ARTICLE INFO

Article history:
Received 11 August 2016
Accepted 11 February 2017
Available online xxxx

Keywords:
Biomarker
Phase 2
Group sequential boundary
Strong control

ABSTRACT

Using stopping boundaries developed for group sequential trials, we control the overall type 1 error for a series of tests that evaluate the treatment effect of an experimental agent in each of several predefined marker subgroups. We then go on to present a procedure based on these group sequential stopping boundaries that provides strong control of type 1 error for testing a series of hypothesis regarding the treatment effect over a range of marker expression. Finally this use of group sequential procedures to control the type 1 error in biomarker subset testing will be compared with some other benchmark procedures in a simulation.

1. Introduction

A common problem encountered in the Phase 2 evaluation of a new oncology agent is the testing of the relationship between a clinical endpoint such as time to progression and the expression of a marker that may identify subjects who benefit from the experimental therapy. There are two approaches that may be taken to address this problem. The first is to test for a treatment effect in each of several marker defined subsets. This approach helps to answer the question of whether the new agent provides clinical benefit for some group of subjects. A second approach is to test for a difference in treatment effects between subjects in a marker defined subset and its complement. This later approach helps answer the question of whether the power of the Phase 3 trial can be enhanced by selecting for subjects who express the marker at sufficiently high levels.

In this paper we will first delve into the issue of what we are trying to learn in the context of drug development from the testing of marker defined subsets. Then we will turn to group sequential boundaries, [1] [2, 3], to control the type 1 error involved in testing for a treatment effect in multiple marker defined subsets. And finally we will look at the results of a simulation that compared the results of this group sequential procedure with some other bench mark approaches to testing marker defined subsets.

The paper assumes throughout that the clinical setting is a Phase 2 oncology trial with progression free survival (PFS) as the primary endpoint and the comparisons between treatment and control are made with the log rank test.

2. Testing for activity with a biomarker

When thinking about how to set up the formal testing for activity in a Phase 2 trial of an oncology drug when one has a biomarker that may identify who benefits from treatment, two basic approaches come to mind. One approach is to simply test for a treatment effect in several marker defined subsets. That is in several marker defined subsets PFS in the treatment group is compared with PFS in the control using the log rank test. An alternative approach is to compare the treatment effect on PFS in several marker defined subsets with the treatment effect in the complement of these subsets to see if the power of a Phase 3 trial could be enhanced by excluding subjects from the complement of one of these marker defined subsets. Jiang et al. proposed such a testing procedure [4]. This second approach, which attempts to answer a Phase 3 design question, would seem to be the preferred approach except for the fact that it is a harder question to answer. To answer the question of whether the treatment effect in a subset is greater than the treatment effect in the complement requires comparing PFS in four groups since we are comparing the results of the treatment and control subjects in a marker defined subset with the results of the treatment and control subjects in the complement of a marker defined subset. Answering the question of whether there is a treatment effect in a subset requires comparing the results of only two groups of subjects, treatment and control subjects in the marker defined subset. In essence when we are detecting a treatment effect in a subset we have $2 \cdot \sigma^2$ in our denominator for the log rank test while when are looking for a subset that will enhance the power of Phase 3, we have $4 \cdot \sigma^2$ in the denominator of the log rank test.

So from a drug development perspective, if the Phase 2 trial is used to identify which drugs to study further in Phase 3 then testing in subsets would be the preferred approach since this will provide greater power to identify drugs with activity. On the other hand if the purpose

E-mail address: eric.holmgren@oncomed.com.

http://dx.doi.org/10.1016/j.cct.2017.02.006
1551-7144/

Table 1
Hypotheses to test.

Marker expression	Hypothesis
$>p_1$ Percentile	$H_1: \lambda_1 = 1$
$>p_2$ Percentile	$H_2: \lambda_1 = 1$ and $\lambda_2 = 1$
$>p_3$ Percentile	$H_3: \lambda_1 = 1, \lambda_2 = 1$ and $\lambda_3 = 1$
All subjects	$H_4: \lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 1$, and $\lambda_4 = 1$

of Phase 2 is to help optimize the design of the Phase 3 trial rather than to act as a gate to reduce the number of drugs tested in Phase 2, then testing the relationship between the biomarker and the magnitude of the treatment effect would be the preferred approach.

This is not to say that when the primary objective in Phase 2 is to develop evidence of a treatment effect that we will not also try to determine the best subset to study in Phase 3. Rather we are saying that this question of what is the best subset to study may be answered informally in the same trial, or after looking at other Phase 2 trials in different indications or after conducting a subsequent Phase 2 trial or even that the Phase 3 trial can be designed to help answer this question as well as provide definitive evidence of clinical benefit. These efforts can commence once it has been determined that the drug may have activity worthy of further study.

Herein we will take the primary objective of testing in Phase 2 to be detecting a treatment effect and will discuss using the machinery that has been developed for group sequential analysis to facilitate testing for a treatment effect in several subsets defined by a biomarker.

3. Group sequential approach to testing marker defined subsets

First let's set some notation. We will use $\lambda_1 \ldots \lambda_4$ to represent the treatment effect measured as a hazard ratio in each of 4 marker defined subsets which are defined in terms of the quartiles of the marker distribution $p_1 > p_2 > p_3$. λ_1 represents the treatment effect in subjects with marker expression greater than the p_1 percentile, λ_2 represents the treatment effect in subjects with marker expression between the p_1 and p_2 percentiles. λ_3 represents the treatment effect between the p_2 and p_3 marker percentiles and λ_4 the treatment effect in subjects with marker expression less than the p_3 percentile.

The hypotheses that we are interested in testing are listed in Table 1 and involve testing for a treatment effect in overlapping marker subsets. That is testing is undertaken to detect a treatment effect (hazard ratio < 1) in subjects with marker expression greater than the p_1, p_2 and p_3 percentiles as well in all subjects enrolled in the study. The rational for testing these subsets rests on the underlying belief that higher levels of marker expression should lead to greater treatment efficacy.

In order to draw a parallel between group sequential testing and testing in these marker subsets, let's suppose for the sake of simplicity that we are evaluating a study where once subjects receive a single dose of study treatment the primary endpoint is immediately observed. To reconstruct the interim analyses after such a study is completed, subjects would be ordered by the calendar time at which they entered the study. Subjects whose calendar time of study entry fell before the calendar time of the first interim would be included in the first interim analysis, subjects whose calendar time of study entry fell before the second interim would be included in the second interim analysis etc.

Now instead of ordering subjects by the time they enter the study, we can order them by their marker expression. For example to test the hypotheses listed in Table 1 we can include all the subjects with marker expression greater than the p_1 percentile in the first analysis, all the subjects with marker expression greater than the p_2 percentile in the second analysis etc. Thereby we can use the machinery developed for group sequential analysis to control type 1 error and calculate power.

Table 2 describes the basic probability calculations for group sequential error probabilities in terms of the Z statistic for the mean of normally distributed data. Z_k and Z_{k-1} are the simple Z statistics for two overlapping groups of subjects corresponding to two analysis times with respective samples of sizes n_k and n_{k-1} in the treatment and control arms as displayed in the first line of Table 2. Here n_k is taken to be greater than n_{k-1}, that is the k'th Z statistic summarizes all the data in the study at a later time than the $k-1$'th Z statistic. The conditional probability that Z_k is greater than z_α given Z_{k-1} can be written as in the middle portion of the table, and the probability that $Z_k > z_\alpha$ is then just the integral of this probability with respect to Z_{k-1} as is displayed in the last part of the table. The last equation forms the basis of a recursive algorithm for calculating type 1 error and power for an interim analysis plan. First note that we know the distribution of Z_1 since it is a simple Z statistic. Now given the distribution of Z_1 and the fact that if the distribution of Z_{k-1} is known one can determine the distribution of Z_k, by induction the distribution of Z_k can be determined for $k \geq 2$. From these equations it can be seen all that is required to carry out this calculation is the definition of the overlapping groups of subjects. With interim analyses the overlapping groups are defined with respect to study entry time. With testing for a treatment effect in biomarker subsets, the overlapping groups can be defined in terms of marker expression. The equations for calculating the error probabilities are exactly the same in both cases.

It is not immediately clear that the calculations presented in Table 2 can be applied to the log rank test. However using the representation of the log rank test as a martingale process [5] and the independent increment property of martingales, a similar argument to that presented in Table 2 can be made for the log rank test as well.

4. Example

Next we will construct a testing procedure that can detect a treatment effect in a marker subset of the study population or in all subjects

Table 2
Basic probability calculations for group sequential stopping boundaries.

Two overlapping Z statistics $n_k > n_{k-1}$	$Z_k = \frac{\sum_{i=1}^{n_k} X_i^T - \sum_{i=1}^{n_k} X_i^C}{\sqrt{2\sigma^2 \cdot n_k}} \quad Z_{k-1} = \frac{\sum_{i=1}^{n_{k-1}} X_i^T - \sum_{i=1}^{n_{k-1}} X_i^C}{\sqrt{2\sigma^2 \cdot n_{k-1}}}$
Probability Z for larger group is significant given Z for smaller group	$P(Z_k > z_\alpha \mid Z_{k-1}) = P(\sqrt{n_k} \cdot Z_k > \sqrt{n_k} \cdot z_\alpha \mid Z_{k-1}) =$ $P(\sqrt{n_k} \cdot Z_k - \sqrt{n_{k-1}} \cdot Z_{k-1} > \sqrt{n_k} \cdot z_\alpha - \sqrt{n_{k-1}} \cdot Z_{k-1} \mid Z_{k-1}) =$ $P(\frac{\sqrt{n_k} \cdot Z_k - \sqrt{n_{k-1}} \cdot Z_{k-1}}{\sqrt{n_k - n_{k-1}}} > \frac{\sqrt{n_k} \cdot z_\alpha - \sqrt{n_{k-1}} \cdot Z_{k-1}}{\sqrt{n_k - n_{k-1}}} \mid Z_{k-1}) =$ $P(Z > \frac{\sqrt{n_k} \cdot z_\alpha}{\sqrt{n_k - n_{k-1}}} - \frac{\sqrt{n_{k-1}} \cdot Z_{k-1}}{\sqrt{n_k - n_{k-1}}} \mid Z_{k-1})$
Probability Z for larger group significant given Z for smaller group does not reject	$P(Z_k > z_\alpha) = \int_{Z_{k-1}\ \text{not reject}} P(Z > \frac{\sqrt{n_k} \cdot z_\alpha}{\sqrt{n_k - n_{k-1}}} - \frac{\sqrt{n_{k-1}} \cdot Z_{k-1}}{\sqrt{n_k - n_{k-1}}} \mid Z_{k-1}) \cdot df_{Z_{k-1}}$

Please cite this article as: E. Holmgren, The application of group sequential stopping boundaries to evaluate the treatment effect of an experimental agent across a range o..., Contemp Clin Trials (2017), http://dx.doi.org/10.1016/j.cct.2017.02.006