

Review

Design and Analysis of Cancer Clinical Trials for Personalized Medicine

Sin-Ho Jung

Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27710, USA; sinho.jung@duke.edu

Abstract: Biomarkers play a key role in the development of personalized medicine. Cancer clinical trials with biomarker should be appropriately designed and analyzed reflecting the various factors, such as the phase of trials, the type of biomarker, the study objectives, and whether the used biomarker is already validated or not. In this paper, we demonstrate design and analysis of two phase II cancer clinical trials, one with a predictive biomarker and the other with a prognostic biomarker. A statistical testing method and its sample size calculation method are presented for each of the trials. We assume that the primary endpoint of these trials is a time to event variable, but this concept can be used for any type of endpoint with associated testing methods. The test statistics and their sample size formulas are derived using the large sample approximation based on the martingale central limit theorem. Using simulations, we find that the test statistics control the type I error rate accurately and the sample sizes calculated using the formulas maintain the statistical power specified at the design stage.

Keywords: enrichment trial; interaction; predictive biomarker; prognostic biomarker; progression-free survival; stratified randomization trial



Citation: Jung, S.-H. Design and Analysis of Cancer Clinical Trials for Personalized Medicine. *J. Pers. Med.* **2021**, *11*, 376. <https://doi.org/10.3390/jpm11050376>

Academic Editor: Lorenza Rimassa

Received: 7 April 2021

Accepted: 22 April 2021

Published: 4 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In many cancer clinical trials, different types of biomarkers are measured from the tumor, blood or urine using molecular, biochemical, physiological, anatomical, or imaging method at the baseline or during treatment. The observed biomarkers are used for various purposes during the diagnosis and treatment of the diseases. For example, cancer biomarkers are used to diagnose diseases (diagnostic biomarker), to predict the response to a specific treatment (predictive biomarker), to measure the aggressiveness of a disease for patients with no or a non-targeted treatment (prognostic biomarker), to monitor the recurrence of a disease, and so on.

These biomarkers can be used to select a treatment of cancer patients. However, biomarkers should be validated before being used to select a treatment in clinical trials. If a biomarker has not been validated yet, it can be used as a stratification factor of a randomized clinical trial. In such a trial, the biomarker is used for its validation, rather than for treatment selection.

The design and analysis method of a clinical trial with a biomarker-guided treatment can be very different depending on the type of the used biomarker, the biomarker's development stage, the study objective, and so on. Various design issues of randomized clinical trials with biomarkers have been widely discussed [1]. A series of statistical testing has been proposed for a randomized phase II trial with a potentially predictive biomarker which has not been strictly validated yet [2]. The efficacy of enrichment trials and stratified randomization trials with a time to event variable as the primary endpoint has been compared assuming that the treatment effect reverses between biomarker positive and negative groups and considering subset analysis within each biomarker status group [3].

Phase II trials are to screen out inefficient treatments before proceeding to a large-scale studies, such as a phase III trial. As such, phase II trials should be completed in a

short time period, so that we must choose a small sample size and a short-term surrogate endpoint, such as tumor response or progression-free survival, as the primary endpoint, rather than a confirmatory endpoint, such as overall survival.

In this paper, we demonstrate two phase II cancer clinical trials, one with a predictive biomarker and the other with a prognostic biomarker, and present analysis and sample size calculation methods for these trials. We use a survival variable as the primary endpoint in this paper, but the same concept can be used for any kind of variables including a binary variable, such as tumor response. For the purpose of sample size calculation, we assume exponential survival distributions which are most popularly used in real trial designs, although the statistical testing does not depend on any specific survival distribution. This is a review article of a biostatistical paper [4] with some modifications.

2. Materials and Methods

We consider a time to event (or survival) endpoint, progression-free survival (PFS). We use a generalized log-rank test for a trial with an imaging prognostic biomarker and a Cox proportional hazards model for a trial with a predictive biomarker, and derive their sample size formulas. To account for relatively small sample sizes of phase II trials, exact statistical methods are used for binary outcomes, but in general no exact methods are available for survival analysis. Therefore, using simulations on two real trial examples, we evaluate the small sample performance of the discussed statistical tests and their sample size formulas that are derived based on large sample approximation.

3. Results

3.1. A Phase II Trial with a Predictive Biomarker

Predictive biomarkers help provide information on the likelihood of response to a specific chemotherapy. For example, tumors expressing high thymidylate synthase (TS) levels were shown to be resistant to pemetrexed in a preclinical study [5], but it was not validated by a clinical study yet. Suppose that we want to investigate whether TS expression is a predictive marker for the clinical outcome of pemetrexed/cisplatin (PC) in patients with nonsquamous non-small-cell lung cancer (NSCLC) through a phase II trial. The control non-targeted treatment is gemcitabine/cisplatin (GC). Compared to GC, PC is expected to be similarly efficacious for TS-positive group, but to be more efficacious in TS-negative group.

To investigate this hypothesis, we want to randomize patients between the two treatment arms stratifying by TS-positivity vs. TS-negativity. This trial was designed and published with overall response as the primary endpoint [6], but in this paper, we demonstrate how to design and analyze a trial using PFS as the primary endpoint based on the estimates from the trial.

3.1.1. Statistical Testing

When this study is completed, PFS will be regressed on treatment allocation z_1 (=0 for GC arm; =1 for PC arm) and TS-positivity z_2 (=0 for TS-negative group; =1 for TS-positive group) using a proportional hazards model [7]

$$\lambda(t) = \lambda_0(t) \exp(\beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1 z_2) \quad (1)$$

Please note that we have an interaction term $z_1 z_2$ in the model.

From model (1), the hazard functions of four patient groups defined by treatments and TS status are given as $\lambda(t|z_1 = 0, z_2 = 0) = \lambda_0(t)$, $\lambda(t|z_1 = 1, z_2 = 0) = \lambda_0(t) \exp(\beta_1)$, $\lambda(t|z_1 = 0, z_2 = 1) = \lambda_0(t) \exp(\beta_2)$, and $\lambda(t|z_1 = 1, z_2 = 1) = \lambda_0(t) \exp(\beta_1 + \beta_2 + \beta_3)$. For TS-positive patients ($z_2 = 1$), the hazard ratio between PC and GC is $\lambda(t|z_1 = 1, z_2 = 1) / \lambda(t|z_1 = 0, z_2 = 1) = \exp(\beta_1 + \beta_3)$, so that we expect $\beta_1 \approx -\beta_3$ if GC and PC are similarly efficacious for TS-positive patients. For GC arm ($z_1 = 0$), the hazard ratio between TS-positivity group and TS-negativity group is $\lambda(t|z_1 = 0, z_2 = 1) / \lambda(t|z_1 = 0, z_2 = 0) = \exp(\beta_2)$, so that we will have $\beta_2 = 0$ if GC is non-targeted against TS. With $\beta_2 = 0$, $\lambda_0(t)$

is the hazard function for GC arm. On the other hand, for PC arm ($z_1 = 1$), the hazard ratio between TS-positivity group and TS-negativity group is $\lambda(t|z_1 = 1, z_2 = 1)/\lambda(t|z_1 = 1, z_2 = 0) = \exp(\beta_3)$ since GC is non-targeted treatment (i.e., $\beta_2 = 0$). If TS-positive tumors are resistant to pemetrexed, we will have $\beta_3 > 0$. Therefore, the hypotheses of interest are $H_0 : \beta_3 = 0$ and $H_1 : \beta_3 > 0$.

For patient $i (= 1, \dots, n)$, let X_i be the minimum of censoring time and survival time, δ_i be the event indicator taking 1 if tumor progression has occurred and 0 otherwise, and $Z_i = (z_{1i}, z_{2i}, z_{1i}z_{2i})^T$ be the covariate vector. Partial score and information functions for regression coefficients $\beta = (\beta_1, \beta_2, \beta_3)^T$ are given as

$$U(\beta) = \sum_{i=1}^n \int_0^\infty \left\{ Z_i - \frac{\sum_{j=1}^n Y_j(t) Z_j e^{\beta^T Z_j}}{\sum_{j=1}^n Y_j(t) e^{\beta^T Z_j}} \right\} dN_i(t)$$

and

$$I(\beta) = \sum_{i=1}^n \int_0^\infty \left[\frac{\sum_{j=1}^n Y_j(t) Z_j^{\otimes 2} e^{\beta^T Z_j}}{\sum_{j=1}^n Y_j(t) e^{\beta^T Z_j}} - \frac{\{\sum_{j=1}^n Y_j(t) Z_j e^{\beta^T Z_j}\}^{\otimes 2}}{\{\sum_{j=1}^n Y_j(t) e^{\beta^T Z_j}\}^2} \right] dN_i(t),$$

respectively, where $N_i(t) = \delta_i I(X_i \leq t)$ is the event process, $Y_i(t) = I(X_i \geq t)$ is the at-risk process, $I(\cdot)$ is an indicator function, and $z^{\otimes 2} = zz^T$ for a vector z .

Let $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)^T$ denote the solution to $U(\beta) = 0$. Then, $\hat{\beta}$ is approximately normal with mean 0, and variance–covariance $I^{-1}(0)$ under the global null hypothesis of $\beta = 0$ [8]. Hence, with a one-sided type I error rate of α , we reject $H_0 : \beta_3 = 0$ in favor of $H_1 : \beta_3 > 0$ if $\hat{\beta}_3/\hat{\sigma}_3 > z_{1-\alpha}$, where $\hat{\sigma}_3^2$ is the (3,3)-component of $I^{-1}(0)$ and $z_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard normal distribution.

3.1.2. Sample Size Calculation

For sample size calculation of this study, we need to specify following design parameters.

- Type I error rate and power, $(\alpha, 1 - \beta)$
- Allocation proportion for GC arm, p_0 , and for PC arm, p_1 ($p_0 + p_1 = 1$)
- TS-negativity q_0 and TS-positivity q_1 based on the prevalence in the study population ($q_0 + q_1 = 1$)
- Assuming exponential distributions for PFS, the hazard rates, $\lambda_{z_1 z_2}$, of the four patient groups, $\lambda_{00}, \lambda_{01}, \lambda_{10}$, and λ_{11}
- Accrual period a (or accrual rate r) and additional follow-up period b

Assuming exponential distributions for PFS with hazard rates $\lambda_{z_1 z_2}$, model (1) is simplified to

$$\lambda = \lambda_0 \exp(\beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1 z_2)$$

with

$$\lambda_{00} = \lambda_0, \quad \lambda_{10} = \lambda_0 \exp(\beta_1), \quad \lambda_{01} = \lambda_0 \exp(\beta_2), \quad \lambda_{11} = \lambda_0 \exp(\beta_1 + \beta_2 + \beta_3).$$

By solving these equations with respect to $(\lambda_0, \beta_1, \beta_2, \beta_3)$, we have

$$\lambda_0 = \lambda_{00}, \quad \beta_1 = \log \lambda_{10} - \log \lambda_{00}, \quad \beta_2 = \log \lambda_{01} - \log \lambda_{00},$$

$$\beta_3 = \log \lambda_{11} - \log \lambda_{10} - \log \lambda_{01} + \log \lambda_{00}. \tag{2}$$

Hence, we can calculate the values of β_3 under H_1 in terms of the hazard ratios that are specified as design parameters above.

To derive a sample size formula, we need to calculate the limit of $\hat{\sigma}_3^2$ or $I^{-1}(0)$ as $n \rightarrow \infty$ in terms of the design parameters. Let $p_{kl} = P(z_1 = k, z_2 = l)$ for $k, l = 0$ or 1 denote the relative frequency of each cell of the 2×2 table defined by treatment and TS status. Under the stratified randomization scheme, we have $p_{kl} = p_k q_l$ for $k, l = 0$ or 1 .

Appendix A shows that $I(0)$ converges to DA , where $D = nd$ denotes the expected number of events (or number of patients with tumor progression), $d = p_{00}d_{00} + p_{10}d_{10} +$

$p_{01}d_{01} + p_{11}d_{11}$ denotes the probability that a patient has a progression during the study period, $d_{kl} = 1 - \exp(-\lambda_{kl}b)\{1 - \exp(-\lambda_{kl}a)\}$ denotes the probability that a patient in group $(z_1, z_2) = (k, l)$ has a progression for $k, l = 0, 1$ as derived based on an exponential PFS distribution and $U(b, a + b)$ censoring distribution, and

$$A = \begin{pmatrix} p_0p_1 & p_{11} - p_1q_1 & p_0p_{11} \\ p_{11} - p_1q_1 & q_0q_1 & q_0p_{11} \\ p_0p_{11} & q_0p_{11} & p_{11}(1 - p_{11}) \end{pmatrix}$$

Please note that $d_{z_1z_2}$ is derived from an exponential PFS distribution with hazard ratio $\lambda_{z_1z_2}$ and a censoring distribution of $U(b, a + b)$. Hence, the limit of $\hat{\sigma}_3^2$ is $\sigma_3^2 = A_{(3,3)}/D$, where $A_{(3,3)}$ is the (3, 3) component of A^{-1} .

From (2), $\bar{\beta}_3 = \log \lambda_{11} - \log \lambda_{10} - \log \lambda_{01} + \log \lambda_{00}$ is the β_3 value specified under H_1 . Since $(\hat{\beta}_3 - \bar{\beta}_3)/\sigma_3$ has the standard normal distribution under H_1 , the power for a local alternative hypothesis $H_1 : \beta_3 = \bar{\beta}_3$ is given as

$$\begin{aligned} 1 - \beta &= P(\hat{\beta}_3/\hat{\sigma}_3 > z_{1-\alpha} | \beta_3 = \bar{\beta}_3) \\ &= P\left(\frac{\hat{\beta}_3 - \bar{\beta}_3}{\sigma_3} > z_{1-\alpha} - \frac{\bar{\beta}_3}{\sigma_3} \mid \beta_3 = \bar{\beta}_3\right) \\ &= \Phi(z_{1-\alpha} - \bar{\beta}_3/\sigma_3) \end{aligned} \tag{3}$$

where $\Phi(\cdot)$ is the survivor function of the standard normal distribution.

Noting that $\sigma_3^2 = DA_{(3,3)} = ndA_{(3,3)}$, we obtain the required number of events

$$D = A_{(3,3)} \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\bar{\beta}_3}\right)^2$$

or the required sample size

$$n = \frac{A_{(3,3)}}{d} \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\bar{\beta}_3}\right)^2 \tag{4}$$

by solving Equation (3).

Formula (4) requires specification of accrual period a together with $(\alpha, 1 - \beta)$, $\bar{\beta}_3$, b , p_0 and q_0 . In designing a clinical trial, however, we can estimate the accrual pattern, rather than an accrual period. Suppose that patients are expected to be enrolled to the study at a rate of r during an accrual period based on the number of patients treated by the study member sites recently. Assuming uniform patient accrual during period a , we have $n \approx a \times r$. Noting that $d = d(a)$ is a function of a , (4) is expressed as

$$a \times r = \frac{A_{(3,3)}}{d(a)} \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\bar{\beta}_3}\right)^2 \tag{5}$$

By solving (5) with respect to a using a numerical method, such as the bisection method, we obtain the required accrual period, say a^* , and the required sample size $n = a^*r$.

3.1.3. Example 1

We demonstrate our sample size calculation method with the NSCLC trial that is introduced above. We will randomize patients between the two treatment arms with 1-to-1 fashion, i.e., $p_0 = p_1 = 1/2$ stratified by TS status. The expected TS-positivity is 50% (i.e., $q_0 = q_1 = 0.5$) because the median TS level was selected as the cutoff value for TS-positivity from a previous study [9]. Hence, we have $p_{z_1z_2} = p_{z_1}q_{z_2} = 1/4$ for $z_1, z_2 = 0$ or 1. The 6-month PFS is expected to be about 35% for GC arm regardless of TS level and for PC arm with TS-positivity, and 55% for PC arm with TS-negativity. For an

exponential distribution, t -year survival probability $S(t)$ is associated with its hazard rate λ by $S(t) = \exp(-\lambda t)$. Therefore, the annual hazard rates under the alternative hypothesis are given as $\lambda_{00} = \lambda_{01} = \lambda_{11} = 2.100$ and $\lambda_{10} = 1.196$ under the exponential PFS assumption. For these hazard rates, we have the baseline hazard rate $\lambda_0 = 2.100$, $\beta_2 = 0$, and $\beta_3 = -\beta_1 = 0.563$ from (2). Suppose that about 10 patients per month are expected to be entered to the study, i.e., an annual accrual rate of $r = 120$. We plan to follow the patients for additional $b = 1$ year after the last patient enters. Then, the 1-sided $\alpha = 0.1$ test for $H_0 : \beta_3 = 0$ against $H_1 : \beta_3 = 0.563$ in model (1) requires $n = 345$ patients for a power of $1 - \beta = 0.9$. The expected number of events (i.e., number of patients with a disease progression) at the analysis will be $D = 333$. As an effort to lower the sample size for this phase II trial, we use a large α level compared to the standard two-sided $\alpha = 0.05$. We observe an empirical power of 0.897 from 10,000 simulation samples of size $n = 345$ that are generated at the design setting. This trial recruited 321 patients using overall response as the primary endpoint [6].

A stratified randomized trial of a treatment with a predictive biomarker requires a large sample size for testing on the interaction term. Sample size of a trial for testing the interaction term with 50% of biomarker positivity may be compared to that of a trial for an arm-to-arm comparison with 1-to-3 randomization in the setting of the NSCLC trial expecting a higher efficacy of PC only for TS-negative group. Let us consider a randomized trial to compare two treatment arms with $(\alpha, 1 - \beta, r, b) = (0.1, 0.9, 120, 1)$ as above and 6-month PFS of 35% for the control treatment and 55% for the experimental treatment. In this case, we need only $n = 122$ ($D = 102$) patients by 1-to-3 randomization by a sample size formula for the standard log-rank test [10]. If TS had been already validated to be a predictive biomarker of pemetrexed before this trial, then we could have chosen an enrichment trial for TS-negative patients that would require a much smaller sample size. The efficiency has been compared between an enrichment design and a stratified randomization design has been for predictive biomarker in terms of a continuous outcome [11].

3.2. A Phase II Trial with a Prognostic Biomarker

Prognostic biomarkers provide information on the overall cancer outcome in patients to facilitate cancer diagnosis regardless of selected treatments. In this section, we consider a phase II trial with an imaging prognostic biomarker. Chemotherapy B has been a standard regimen for patients with non-bulky stage I and II Hodgkin lymphoma. In a previous study on 6 cycles of B, each patient had a FDG-PET (fluorodeoxyglucose positron-emission tomography) imaging after 2 cycles of B. It was found that the patients with a negative PET image (group 1) and those with a positive PET image (group 2) had a 3-year PFS of $S_1(3) = 0.86$ and $S_2(3) = 0.52$, respectively, and the hazard ratio, $\Delta = \lambda_2/\lambda_1$, was estimated as $\Delta_0 = 4.3$.

In a new single-arm phase II trial, the patients with a negative PET image after 2 cycles of B will be treated by additional 4 cycles of the chemotherapy B as in the previous study, whereas those with a positive PET image after 2 cycles of B will be treated by 4 cycles of a more aggressive chemotherapy C plus radiation therapy (C+RT).

In this trial, we want to show that by treating PET positive patients with the more aggressive therapy C+RT, their PFS will become closer to that of PET negative patients who are treated by the standard chemotherapy B. To this end, we test $H_0 : \Delta = \Delta_0$ against $H_1 : \Delta < \Delta_0$. Although the PFS of group 2 will be different between H_0 and H_1 , that of group 1 is expected to be identical since PET negative patients receive the same treatment as that of the previous study.

Statistical Testing

Let n_k denote the sample size in group k , $n = n_1 + n_2$ the total sample size, and T_{ki} the time to progression for subject i in group k ($1 \leq i \leq n_k; k = 1, 2$). We observe (X_{ki}, δ_{ki}) , where X_{ki} is the minimum of T_{ki} and the censoring time and δ_{ki} is an event (or progression)

indicator taking 1 if the subject had a tumor progression and 0 otherwise. For group k , T_{k1}, \dots, T_{k,n_k} are distributed with hazard function $\lambda_k(t)$. Under the proportional hazards assumption, $\Delta = \lambda_2(t)/\lambda_1(t)$ denotes the hazard ratio between the two patient groups.

Let $\hat{\Lambda}_k(t) = \int_0^t Y_k^{-1}(t) dN_k(t)$ denote the Aalen–Nelson estimator [12,13] for the cumulative hazard function $\Lambda_k(t) = \int_0^t \lambda_k(s) ds$, $Y_k(t) = \sum_{i=1}^{n_k} I(X_{ki} \geq t)$ and $N_k(t) = \sum_{i=1}^{n_k} \delta_{ki} I(X_{ki} \leq t)$ are the at-risk process and the event process for group k , respectively, and $N(t) = N_1(t) + N_2(t)$. It was shown [14] that

$$W(\Delta) = \int_0^\infty \frac{Y_1(t)Y_2(t)}{Y_1(t) + \Delta Y_2(t)} \{ \Delta d\hat{\Lambda}_1(t) - d\hat{\Lambda}_2(t) \}$$

is increasing in Δ , and $W(\Delta)/\sigma_n(\Delta)$ is asymptotically $N(0, 1)$, where

$$\sigma_n^2(\Delta) = \Delta \int_0^\infty \frac{Y_1(t)Y_2(t)}{\{Y_1(t) + \Delta Y_2(t)\}^2} dN(t)$$

Hence, we reject $H_0 : \Delta = \Delta_0$, in favor of $H_1 : \Delta < \Delta_0$, if $W(\Delta_0)/\sigma_n(\Delta_0) > z_{1-\alpha}$ with one-sided type I error rate α . The test statistic with $\Delta_0 = 1$ is the standard log-rank test [15].

3.3. Sample Size Calculation

We want to estimate the sample size n under a local alternative hypothesis $H_1 : \Delta = \Delta_1 (<\Delta_0)$ with a desired power. For sample size calculation of this trial, we need to specify following design parameters.

- Type I error rate and power, $(\alpha, 1 - \beta)$
- PET-negativity and PET-positivity, p_1, p_2
- Distributions of PFS for PET negative and PET positive patient groups: exponential distributions with hazard rates λ_1 for PET negative group; λ_{20} under H_0 and λ_{21} under H_1 for PET positive group
- Accrual period a (or accrual rate r) and additional follow-up period b

Using the specified hazard rates, we have $\Delta_0 = \lambda_{20}/\lambda_1$ and $\Delta_1 = \lambda_{21}/\lambda_1$. A sample size formula with $\Delta_0 > 1$ and $\Delta_1 = 1$ for designing non-inferiority trials was proposed [14]. This sample size formula was further extended for general Δ_0 and Δ_1 with $\Delta_1 < \Delta_0$ [16].

Appendix B derives a sample size formula by adapting Jung and Chow’s formula [16] for this trial,

$$n = \frac{(\sigma_0 z_{1-\alpha} + \sigma_1 z_{1-\beta})^2}{\omega^2} \tag{6}$$

where

$$\begin{aligned} \omega &= p_1 p_2 \int_0^\infty \frac{G(t) S_1(t) S_{21}(t) \{ \lambda_1 \Delta_0 - \lambda_{21} \}}{p_1 S_1(t) + p_2 \Delta_0 S_{21}(t)} dt \\ \sigma_0^2 &= \Delta_0 p_1 p_2 \int_0^\infty \frac{G(t) S_1(t) S_{21}(t) \{ p_1 \lambda_1 S_1(t) + p_2 \lambda_{21} S_{21}(t) \}}{\{ p_1 S_1(t) + p_2 \Delta_0 S_{21}(t) \}^2} dt \\ \sigma_1^2 &= p_1 p_2 \int_0^\infty \frac{G(t) S_1(t) S_{21}(t) \{ p_2 \lambda_1 \Delta_0^2 S_{21}(t) + p_1 \lambda_{21} S_1(t) \}}{\{ p_1 S_1(t) + p_2 \Delta_0 S_{21}(t) \}^2} dt \end{aligned}$$

$S_1(t) = \exp(-\lambda_1 t)$, $S_{21}(t) = \exp(-\lambda_{21} t)$, $G(t)$ is the survivor function of the $U(b, a + b)$ censoring distribution with

$$G(t) = \begin{cases} 1 & \text{if } t \leq b \\ -t/b + (a + b)/a & \text{if } b < t \leq a + b \\ 0 & \text{if } t > a + b \end{cases}$$

and $p_k = n_k/n$. The integrals for ω , σ_0^2 , and σ_1^2 are calculated using a numerical method.

The number of events D expected at the analysis time under H_1 is calculated by $D = n(p_1d_1 + p_2d_2)$, where $d_1 = 1 + \int_0^\infty S_1(t)dG(t) = 1 - \exp(-\lambda_1b)\{1 - \exp(-\lambda_1a)\}$ and $d_2 = 1 + \int_0^\infty S_{21}(t)dG(t) = 1 - \exp(-\lambda_{21}b)\{1 - \exp(-\lambda_{21}a)\}$.

Sample size formula (6) assumes that the accrual period a is specified. Suppose that accrual rate r is specified instead of accrual period a . Given $(\lambda_1, \Delta_0, \Delta_1, \alpha, 1 - \beta, p_1, b)$, $\omega = \omega(a)$ and $\sigma_h = \sigma_h(a)$ for $h = 0, 1$ are functions of a . Under uniform accrual assumption, we have $n = a \times r$. Hence, (6) is expressed as

$$a \times r = \frac{\{\sigma_0(a)z_{1-\alpha} + \sigma_1(a)z_{1-\beta}\}^2}{\omega^2(a)} \tag{7}$$

By solving (7) with respect to a , using a numerical method such as the bisection method, we obtain the required accrual period a^* and the required sample size $n = a^* \times r$.

Example 2

We consider the PET-guided Hodgkin lymphoma trial introduced in the beginning of this section. Under H_0 , we assume a 3-year PFS of 86% and 52% for PET negative and positive groups, respectively, which correspond to annual hazard rates of $(\lambda_1, \lambda_{20}) = (0.050, 0.218)$ under an exponential PFS model, resulting in a hazard ratio of $\Delta_0 = 4.3$. By treating the PET positive patients with an aggressive treatment C+RT, we expect to increase their 3-year PFS up to 74% (from 52%), resulting in an annual hazard rate of $\lambda_{21} = 0.100$ and hazard rate of $\Delta_1 = 2.0$. The previous study observed about $p_2 = 20\%$ of PET-positivity. Assuming an annual accrual rate of $r = 60$ patients and $b = 3$ years of additional follow-up after completion of accrual, we need $n = 191$ patients for $1 - \beta = 90\%$ power for detecting $H_1 : \Delta_1 = 2$ by the generalized log-rank test $W(\Delta_0)/\sigma_n(\Delta_0)$ with one-sided $\alpha = 10\%$ under $H_0 : \Delta_0 = 4.3$. Under this specific alternative hypothesis, we expect about 46 events (progressions or deaths) at the data analysis. This trial was conducted with this study objective as a second objective [17]. Simulation studies are conducted to evaluate the performance of the calculated sample size under the above design settings under H_0 and H_1 . Using 10,000 simulation samples of size $n = 191$ under each hypothesis, the empirical type I error rate and power are observed as 0.0984 (to be compared to $\alpha = 0.1$) and 0.8749 (to be compared to $1 - \beta = 0.9$), respectively.

4. Discussion

We have presented design and analysis methods of two phase II trials for biomarker-guided treatments.

The power of the statistical tests of the trials discussed above depends on the prevalence of the biomarker positivity. Therefore, we need to check the observed prevalence during the patient accrual, and to recalculate the sample size if the observed prevalence is very different from the one specified at the design stage. For both of our example trials, the initial sample size will be under-powered if the observed prevalence is farther from 1/2 than the specified one at the design stage. In this case, we may plan a sample size recalculation reflecting the observed prevalence in the middle of the trial, and modify the sample size of the trial if necessary.

For the sample size calculations, we have assumed exponential survival distributions and an accrual pattern with a constant accrual rate, but we can easily extend the formulas for any survival distributions and any accrual pattern [18]. We have considered a survival endpoint as the primary endpoint in this paper, but the concept can be used to design and analysis for biomarker-driven phase II trials with other type of endpoint, such as a binary outcome for tumor response.

As an effort to lower the sample size of a phase II trial from that of a phase III trial, we use a high type I error rate [19,20], such as 1-sided $\alpha = 5\%$ or 10% (compared to 2-sided $\alpha = 5\%$), a surrogate short-term outcome, such as tumor response or progression-free survival (compared to a confirmatory endpoint such as overall survival), a larger treatment effect, and a single-arm design (compared to a randomized design). Despite these efforts,

we observe that a randomized trial stratified by a predictive biomarker requires a relatively large sample size for a phase II trial. This fact is pointed out by literature [3,11,21].

5. Conclusions

Through simulations on the two real study examples, we find that the proposed statistical tests control the type I error rate accurately and the calculated sample sizes maintain the appropriate power. The sample size calculations require some numerical methods for integration and solving equations. The author developed Fortran programs to implement the sample size formulas, which are available upon request.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Limit of $I(0)$

Since $n^{-1} \sum_{i=1}^n$ uniformly converges to $G(t)S_{00}(t)$ under the global null hypothesis, as $n \rightarrow \infty$, we have

$$\frac{\sum_{j=1}^n Y_j(t)Z_j^{\otimes 2}}{\sum_{j=1}^n Y_j(t)} \rightarrow A_1 = \begin{pmatrix} p_1 & p_{11} & p_{11} \\ p_{11} & q_1 & p_{11} \\ p_{11} & p_{11} & p_{11} \end{pmatrix}$$

and

$$\frac{(\sum_{j=1}^n Y_j(t)Z_j)^{\otimes 2}}{(\sum_{j=1}^n Y_j(t))^2} \rightarrow A_2 = \begin{pmatrix} p_1^2 & p_1q_1 & p_1p_{11} \\ p_1q_1 & q_1^2 & q_1p_{11} \\ p_1p_{11} & q_1p_{11} & p_{11}^2 \end{pmatrix}$$

uniformly. Hence,

$$I(0) = \sum_{i=1}^n \int_0^\infty \left[\frac{\sum_{j=1}^n Y_j(t)Z_j^{\otimes 2}}{\sum_{j=1}^n Y_j(t)} - \frac{\{\sum_{j=1}^n Y_j(t)Z_j\}^{\otimes 2}}{\{\sum_{j=1}^n Y_j(t)\}^2} \right] dN_i(t)$$

converges to AD , where $D = \sum_{i=1}^n \int_0^\infty dN_i(t)$ is the number of events and

$$A = A_1 - A_2 = \begin{pmatrix} p_0p_1 & p_{11} - p_1q_1 & p_0p_{11} \\ p_{11} - p_1q_1 & q_0q_1 & q_0p_{11} \\ p_0p_{11} & q_0p_{11} & p_{11}(1 - p_{11}) \end{pmatrix}$$

Appendix B. Derivation of Jung and Chow's Formula

Let $p_k = n_k/n$ denote the prevalence of group $k(= 1, 2)$. To derive a sample size formula, we must derive the limiting distribution of the test statistic $W(\Delta_0)/\sigma_n(\Delta_0)$ under H_1 . The survivor function of the $U(b, a + b)$ censoring distribution is given as

$$G(t) = \begin{cases} 1 & \text{if } t \leq b \\ -t/b + (a + b)/a & \text{if } b < t \leq a + b \\ 0 & \text{if } t > a + b \end{cases}$$

Please note that under H_1 , $n_1^{-1}Y_1(t)$ and $n_2^{-1}Y_2(t)$ uniformly converge to $G(t)S_1(t)$ and $G(t)S_{21}(t)$, respectively.

Since

$$W(\Delta_0) = \Delta_0 \int_0^\infty \frac{Y_2(t)}{Y_1(t) + \Delta_0 Y_2(t)} dN_1(t) - \int_0^\infty \frac{Y_1(t)}{Y_1(t) + \Delta_0 Y_2(t)} dN_2(t)$$

its variance under H_1 is

$$\hat{\sigma}_1^2 = \Delta_0^2 \int_0^\infty \frac{Y_1(t)Y_2^2(t)}{\{Y_1(t) + \Delta_0 Y_2(t)\}^2} d\Lambda_1(t) + \int_0^\infty \frac{Y_1^2(t)Y_2(t)}{\{Y_1(t) + \Delta_0 Y_2(t)\}^2} d\Lambda_{21}(t)$$

Therefore, $n^{-1}\hat{\sigma}_1^2$ converges to

$$\sigma_1^2 = p_1 p_2 \int_0^\infty \frac{G(t)S_1(t)S_{21}(t)\{p_2 \lambda_1 \Delta_0^2 S_{21}(t) + p_1 \lambda_{21} S_1(t)\}}{\{p_1 S_1(t) + p_2 \Delta_0 S_{21}(t)\}^2} dt$$

On the other hand, under H_1 , $n^{-1}\sigma_n(\Delta_0)$ converges to

$$\sigma_0^2 = \Delta_0 p_1 p_2 \int_0^\infty \frac{G(t)S_1(t)S_{21}(t)\{p_1 \lambda_1 S_1(t) + p_2 \lambda_{21} S_{21}(t)\}}{\{p_1 S_1(t) + p_2 \Delta_0 S_{21}(t)\}^2} dt$$

Also, the expected value of $n^{-1}W(\Delta_0)$ under H_1 is given as

$$\omega = p_1 p_2 \int_0^\infty \frac{G(t)S_1(t)S_{21}(t)\{\lambda_1 \Delta_0 - \lambda_{21}\}}{p_1 S_1(t) + p_2 \Delta_0 S_{21}(t)} dt$$

Please note that we derive σ_0^2 , σ_1^2 , and ω using the exact asymptotic results under H_1 , while Jung (2018) derives them approximately under the nearby alternative hypothesis.

Hence, under H_1 , $W(\Delta_0)$ is approximately normal with mean $n\omega$ and variance $n\sigma_1^2$, so that given n , the power of the test statistic with 1-sided α is

$$\begin{aligned} 1 - \beta &\approx P\left(\frac{W(\Delta_0)}{\sigma_0 \sqrt{n}} > z_{1-\alpha} | H_1\right) \\ &= P\left(\frac{W(\Delta_0) - n\omega}{\sigma_1 \sqrt{n}} \times \frac{\sigma_1}{\sigma_0} + \frac{\omega \sqrt{n}}{\sigma_0} > z_{1-\alpha} | H_1\right) \\ &= \Phi\left(\frac{\sigma_0}{\sigma_1} z_{1-\alpha} - \frac{\omega \sqrt{n}}{\sigma_1}\right) \end{aligned}$$

By solving this equation with respect to n , we obtain the sample size required for a power of $1 - \beta$ as

$$n = \frac{(\sigma_0 z_{1-\alpha} + \sigma_1 z_{1-\beta})^2}{\omega^2}$$

References

- Freidlin, B.; McShane, L.M.; Korn, E.L. Randomized Clinical Trials with Biomarkers: Design Issues. *J. Natl. Cancer Inst.* **2010**, *102*, 152–160. [[CrossRef](#)] [[PubMed](#)]
- Freidlin, B.; McShane, L.M.; Polley, M.Y.C.; Korn, E.L. Randomized Phase II Trial Designs with Biomarkers. *J. Clin. Oncol.* **2012**, *30*, 3304–3309. [[CrossRef](#)] [[PubMed](#)]
- Mandrekar, S.J.; Sargent, D.J. Clinical Trial Designs for Predictive Biomarker Validation: Theoretical Considerations and Practical Challenges. *J. Clin. Oncol.* **2009**, *27*, 4027–4034. [[CrossRef](#)] [[PubMed](#)]
- Jung, S.H. Phase II Cancer Clinical Trials for Biomarker-Guided Treatments. *J. Biopharm. Stat.* **2018**, *28*, 256–263. [[CrossRef](#)] [[PubMed](#)]
- Ozasa, H.; Oguri, T.; Uemura, T.; Miyazaki, M.; Maeno, K.; Sato, S.; Ueda, R. Significance of Thymidylate Synthase for Resistance to Pemetrexed in Lung Cancer. *Cancer Sci.* **2010**, *101*, 161–166. [[CrossRef](#)] [[PubMed](#)]
- Sun, J.M.; Ahn, J.S.; Jung, S.H.; Sun, J.; Ha, S.Y.; Han, J.; Park, K.; Ahn, M.J. Pemetrexed plus Cisplatin versus Gemcitabine plus Cisplatin according to Thymidylate Synthase Expression in Nonsquamous Non-Small-Cell Lung Cancer: A Biomarker-Stratified Randomized Phase II Trial. *J. Clin. Oncol.* **2015**, *33*, 2450–2456. [[CrossRef](#)] [[PubMed](#)]
- Cox, D.R. Regression Models and Life Tables (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1972**, *34*, 187–220.
- Fleming, T.R.; Harrington, D.P. *Counting Processes and Survival Analysis*; Wiley: New York, NY, USA, 1991.
- Sun, J.M.; Han, J.; Ahn, J.S.; Park, K.; Ahn, M.J. Significance of Thymidylate Synthase and Thyroid Transcription Factor 1 Expression in Patients with Nonsquamous Non-Small Cell Lung Cancer Treated with Pemetrexed-Based Chemotherapy. *J. Thorac. Oncol.* **2011**, *6*, 1392–1399. [[CrossRef](#)] [[PubMed](#)]

10. Yateman, N.A.; Skene, A.M. Sample Size for Proportional Hazards Survival Studies with Arbitrary Patient Entry and Loss to Follow-up Distributions. *Stat. Med.* **1992**, *11*, 1103–1113. [[CrossRef](#)] [[PubMed](#)]
11. Maitournam, A.; Simon, R. On the Efficiency of Targeted Clinical Trials. *Stat. Med.* **2005**, *24*, 329–339. [[CrossRef](#)] [[PubMed](#)]
12. Aalen, O.O. Nonparametric Inference for a Family of Counting Processes. *Ann. Stat.* **1978**, *6*, 701–726. [[CrossRef](#)]
13. Nelson, W. Hazard Plotting for Incomplete Failure Data. *J. Qual. Technol.* **1969**, *1*, 27–52. [[CrossRef](#)]
14. Jung, S.H.; Kang, S.J.; McCall, L.; Blumenstein, B. Sample Size Computation for Noninferiority Log-Rank Test. *J. Biopharm. Stat.* **2005**, *15*, 957–967. [[CrossRef](#)] [[PubMed](#)]
15. Peto, R.; Peto, J. Asymptotically Efficient Rank Invariant Test Procedures (with discussion). *J. R. Stat. Soc. Ser. A Stat. Soc.* **1972**, *135*, 185–206. [[CrossRef](#)]
16. Jung, S.H.; Chow, S.C. On Sample Size Calculation for Comparing Survival Curves under General Hypotheses Testing. *J. Biopharm. Stat.* **2012**, *22*, 485–495. [[CrossRef](#)] [[PubMed](#)]
17. Straus, D.J.; Jung, S.H.; Pitcher, B.; Kostakoglu, L.; Grecula, J.; His, E.; Schöder, H.; Popplewell, L.; Chang, J.; Moskowitz, C.; et al. CALBG 50604: Risk-Adapted Treatment of Non-Bulky Early Stage Hodgkin Lymphoma based on Interim PET. *Blood* **2018**, *132*, 1013–1021. [[CrossRef](#)] [[PubMed](#)]
18. Jung, S.H. *Randomized Phase II Cancer Clinical Trials*; Chapman & Hall/CRC: New York, NY, USA, 2013.
19. Simon, R. Optimal Two-Stage Designs for Phase II Clinical Trials. *Control. Clin. Trials* **1989**, *10*, 1–10. [[CrossRef](#)]
20. Jung, S.H.; Sargent, D.J. Randomized Phase II Clinical Trials. *J. Biopharm. Stat.* **2014**, *24*, 802–816. [[CrossRef](#)] [[PubMed](#)]
21. Simon, R.; Maitournam, A. Evaluating the Efficiency of Targeted Designs for Randomized Clinical Trials. *Clin. Cancer Res.* **2004**, *10*, 6759–6763. [[CrossRef](#)] [[PubMed](#)]