Auto-adaptive Alpha Allocation: a strategy to mitigate risk on study assumptions

Yue Shentu[1],Cong Chen[2], Lei Pang[3], Robert A. Beckman[4,5]

[1] Biostatistics and Research Decision Sciences, Merck Research Laboratories (MRL), RY34-A304 126, E. Lincoln Avenue, Rahway, NJ 07065, USA, yue_shentu@merck.com

[2] Biostatistics and Research Decision Sciences, Merck Research Laboratories (MRL), UG1C-46, 375 Sumneytown Pike, Upper Gwynedd, PA 19454, USA, nicole_li@merck.com

[3] Biostatistics and Research Decision Sciences, Merck Research Laboratories (MRL), UG1C-46, 375 Sumneytown Pike, Upper Gwynedd, PA 19454, USA, cong_chen@merck.com

[4] Departments of Oncology and of Biostatistics, Bioinformatics, and Biomathematics, Lombardi Cancer Center, Georgetown University Medical Center, 3800 Reservoir Road Northwest, Washington, DC 20007, USA, eniac1915@gmail.com

[5] Center for Evolution and Cancer, Helen Diller Family Cancer Center, University of California at San Francisco, 1450 3rd Street, San Francisco, California 94305, USA, eniac1915@gmail.com

Abstract

In some clinical development programs, there are potential biomarkers with promising but uncertain predictive effect, while the probability of success in the overall population cannot be readily dismissed. It is risky to focus only on the overall population, or just the biomarker subpopulation. In 2009, Chen and Beckman proposed a Bayesian decision framework to optimize the type I error rate (alpha) allocation in a Phase III clinical study with possible predictive subset effect. The utilization of internal data in this framework is of particular interest because it provides an opportunity to mitigate the potential risk of mis-specified study assumptions using an auto-adaptive strategy. In this paper, we examine this auto-adaptive strategy in detail through extensive numerical case studies, and provide guidance on the appropriate use of partial current trial (internal) data in this data-driven optimization framework. We show that internal data can be used to inform the alpha allocation to hypothesis testing in the overall population and the subgroup. The resulting adaptive testing strategy is robust with respect to the uncertainty in the predictive subgroup effect and biomarker prevalence.

Key words:

Adaptive design; Informational design; Oncology; Targeted therapy; Multiplicity

1. Introduction

With rising interest on targeted therapy and personalized medicine, Phase III clinical trials that are designed to test the treatment effect in both the overall general population and a biomarker-identified subgroup are becoming increasingly popular. These trials are often necessary due to uncertainty in the potential predictive biomarker effect. For an experimental therapy, preliminary evidence may suggest stronger treatment effect for patients with positive biomarker expressions, but the general population may also derive some benefit regardless of the biomarker status. Such uncertainty poses a challenge at the onset of the Phase III program. Traditionally, Phase III trials focused on the overall population. In the RADIANT trial [1] that compared adjuvant erlotinib versus placebo in EGFR positive patients with completely resected stage IB–IIIA non-small cell lung cancer (NSCLC), erlotinib improved disease free survival in the EGFR del19/L858R Activating Mutation (M+) Subgroup. The result was however not statistically significant due to the pre-specified hierarchical testing that prioritized the overall population. On the other hand, some sponsors chose to limit their Phase III trials to only the biomarker-positive subpopulation in search for a targeted therapy, when the drug worked reasonably well for a larger population, or when the correct biomarker subgroup was not what was expected. Gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase (EGFR-TK), improved disease-related symptoms and induced radiographic tumor regressions in later line patients with NSCLC, but later studies surprisingly found “ no consistent associations between levels of EGFR expression and radiographic or symptomatic improvements” [2]. The correct subgroup was later found to be defined by the sequence of the EGFR receptor [3,4,5]

The risk for not being able to identify a meaningful subgroup effect and the risk of unnecessarily restricting the target population may be mitigated by targeting both the overall population and the biomarker subgroup in the same trial. The trial will start out to enroll “all-comers”, including both biomarker positive and biomarker negative patients. The two co-primary hypotheses, one for the overall population, another for the pre-determined subgroup, are pre-specified in the statistical analysis plan. The

trial will be considered positive when the null hypothesis of no treatment effect is rejected in either the overall population, or the biomarker subgroup.

Since the ongoing Phase III trial provides the best data to substantiate the biomarker effect for the experimental therapy, practitioners often turn to adaptive designs to modify the trial based on partial trial information (Wang [6]). If the subgroup identification or biomarker expression cutoff has not been finalized before the Phase III trial starts, partial data can be used to determine the subgroup; the remaining portion of the trial can be used to test for the subgroup treatment effect (Freidlin and Simon [7]). However, analytical validation of the new assay cutoff within the framework of a seamless trial may be challenging. If the subgroup has already been determined before Phase III trial starts, interim data can also be used to determine the viability of the identified subgroup. A significant literature has been developed in recent years for subpopulation enrichment in this setting. These strategies emphasize the adaptations upon determination of futility for the biomarker negative subgroup at an interim analysis. Notable examples include the work of Friede [8], Mehta [9], Brannath [10], and Jenkins [11]. The biomarker positive subpopulation would typically continue as the only target population (with possible enrichment) in this framework. An alternative scenario, where some possible therapeutic activities have also been observed in the biomarker negative group, cannot be addressed by subpopulation selection and enrichment. In practice, one might argue that the latter scenario could be more common, when the identification of a perfect biomarker remains an elusive goal.

Instead of selecting one of the subgroups to move forward, Phase III all-comer trials often have to test the treatment effect hypotheses in both the overall population and the biomarker subgroup. These trials are most useful given adequate prevalence of the biomarker subgroup; otherwise partial enrichment may be required. With the simultaneous testing of two hypotheses, multiplicity control becomes an important statistical consideration. Let $H_0^{(1)}$ and $H_0^{(2)}$ be the null hypotheses of no treatment effects in the overall population and the subgroup respectively. By the closed testing procedure the family-wise error rate is controlled at α if $P\left(\text{rejection}\middle|H_0^{(1)} \cap H_0^{(2)}\right) \leq \alpha$, and $P\left(\text{rejection}\middle|H_0^{(i)}\right) \leq \alpha, i = 1 \text{ or } 2$. Different testing strategies can be used to achieve the multiplicity control. The fall-back or ordered testing strategy can prioritize testing in either the overall population or the subgroup; the Bonferroni or Simes procedure [12] allow the remaining hypothesis to be tested in case one hypothesis fails to be rejected. More recently, Spiessens and Debois [13] proposed a method to control the FWER while improving the overall study power by taking advantage of the positive correlation between the testing statistics in the subgroup and the overall population. None of these methods, however, consider what type I error rate (alpha) should be used for each hypothesis to optimize the probability of overall trial success (i.e. rejection of null hypothesis in either the overall population or the biomarker subgroup).

In 2009, Chen and Beckman proposed a method that “provides objective decision rules for Phase III assuming a candidate subpopulation biomarker is available after Phase II, but it is not fully validated” [14]. They demonstrated that this method can be used to optimize “alpha-splitting” between the hypothesis for the overall population and the subgroup. The method of Chen and Beckman shares some common element with that of Spiessens and Debois in the treatment of correlated testing statistics. The additional optimization step in [14] is comparable to selecting an alpha-spending function that maximizes the overall study power in the setting of Spiessens and Debois. In this article, we expand on the concept of [14], and seek to optimize the alpha allocation to the overall population and the subgroup hypotheses

by maximizing the Bayesian conditional expected power given external and/or partial internal data. We show that by paying a pre-determined amount of penalty on nominal type I error rate, this auto-adaptive strategy can draw upon the trial information to guide the alpha allocation between the two hypotheses. The resulting testing strategy is robust for a variety of possible configurations regarding the treatment effects in subgroup and overall population.

The article is organized as follows. Section 2 introduces the method and discusses the practical considerations of type I error control and sample size adjustment. Section 3 presents a series of case studies that demonstrate the performance of the auto-adaptive alpha allocation strategy in a hypothetical Phase III oncology trial. We conclude with discussion in Section 4, including the possible extensions for subpopulation de-selection involving multiple biomarker cutoffs.

2. Methods

Consider a confirmative Phase III clinical trial with a biomarker of interest. The trial will be considered successful if the null hypothesis of no treatment effect for either the overall population or the biomarker subgroup is rejected. Let the planned total Fisher information of the Phase III trial be $I^{(i)}$. The Fisher information measures the strength of the data in statistical inference, and is typically proportional to the number of events in a survival trial (total number of events divided by 4 in a 1:1 randomized study), or to the sample size in a trial studying a continuous or binary clinical endpoint. We shall refer to $I^{(i)}$ as the total internal information of the Phase III trial, to differentiate from $I^{(e)}$, the external information. External information may be derived from Phase II studies or other publicly available data on drugs with similar mechanism. Point estimates and associated standard errors for the Phase III clinical endpoint based on a small exploratory study, for instance, could reasonably serve as the external information for the current trial. Often, however, such external information may be based on early surrogate endpoints that may not correlate well with the actual clinical benefit.

We further assume that in both the external and internal data, the information ratio in the subgroup vs. overall population is r. Therefore, the external and internal information pertinent to the treatment effect in the subgroup is $rI^{(e)}$ and $I^{(i)}$, respectively. As in the case of external information for the treatment effects, prior/external information about the information ratio r may be available, but the reliability of such information is often questionable.

Let X_1 and X_2 be the two Z statistics for the overall population and the subgroup. The correlation between X_1 and X_2 is $\sqrt{r}$. Under the global null hypothesis, $(X_1, X_2) \sim MVN\left([0,0], \begin{pmatrix} 1 & \sqrt{r} \\ \sqrt{r} & 1 \end{pmatrix}\right)$; under an alternative hypothesis, $(X_1, X_2) \sim MVN\left(\left[\sqrt{I^{(i)}}\delta_1, \sqrt{rI^{(i)}}\delta_2\right], \begin{pmatrix} 1 & \sqrt{r} \\ \sqrt{r} & 1 \end{pmatrix}\right)$, where δ_i's are the treatment effects in question, such as the log hazard ratio in a time-to-event trial, or the mean reduction of a continuous clinical endpoint.

Denote α as the Type I error rate of the entire trial. To control the familywise error under the global null hypothesis, the Type I error rates α_1 and α_2 allocated to the overall population and subgroup hypotheses should satisfy

$$1 - \Phi_{\sqrt{r}}\left(Z_{1-\alpha_1}, Z_{1-\alpha_2}\right) = \alpha, \tag{1}$$

where $Z_{1-\alpha_1}, Z_{1-\alpha_2}$ are the standard normal quantiles. Since both α_1 and α_2 are smaller than α, the familywise type I error is strongly controlled at α. There are infinite many pairs of (α_1, α_2) that will satisfy (1), but a particular configuration of (α_1, α_2) may exist to achieve optimality as defined by an objective function of choice. In this article, we will focus on the (conditional) expected power as the objective function for the optimization.

2.1. Auto-adaptive Alpha Allocation

Suppose we would like to utilize a fraction $t \in [0,1)$ of the total internal information $I^{(i)}$ in the Phase III trial to inform the alpha allocation. Let $X_{i,t}$ and $X_{i,1-t}$ be the Z-statistics based on fraction t and $1-t$ of the internal data. The Z-statistics of the final analysis are $X_i = \sqrt{t}X_{i,t} + \sqrt{1-t}X_{i,1-t}$, $i = 1,2$. Under the alternative hypotheses, a reasonable strategy is to find the alpha allocation that will maximize the conditional power given X_{1t} and X_{2t}:

$$P\left(\{X_1 > Z_{1-\alpha_1}\} \cup \{X_2 > Z_{1-\alpha_2}\} \mid X_{1t} = x_{1t}, X_{2t} = x_{2t}\right)$$

$$= 1 - \Phi_{\sqrt{r}}\left(\frac{Z_{1-\alpha_1} - \sqrt{t}x_{1,t}}{\sqrt{1-t}} - \sqrt{(1-t)I^{(i)}}\delta_1, \frac{Z_{1-\alpha_2} - \sqrt{t}x_{2,t}}{\sqrt{1-t}} - \sqrt{(1-t)rI^{(i)}}\delta_2\right), \tag{2}$$

where δ_1, δ_2 are the treatment effects in question for the overall population and the subgroup respectively.

A Bayesian auto-adaptive strategy has been proposed to account for the uncertainty in the true values of δ_1, δ_2 (Chen and Beckman, 2009). This strategy explicitly maximizes the conditional expected power

$$Q_t(\alpha_1, \alpha_2; x_{1,t}, x_{2,t}) = \int \left\{1 - \Phi_{\sqrt{r}}\left(\frac{Z_{1-\alpha_1} - \sqrt{t}x_{1,t}}{\sqrt{1-t}} - \sqrt{(1-t)I^{(i)}}\Delta_1, \frac{Z_{1-\alpha_2} - \sqrt{t}x_{2,t}}{\sqrt{1-t}} - \sqrt{(1-t)rI^{(i)}}\Delta_2\right)\right\} dG_t(\Delta_1, \Delta_2 | x_{1,t}, x_{2,t}) \tag{3}$$

under the constraint of

$$1 - \Phi_{\sqrt{r}}\left(Z_{1-\alpha_1}, Z_{1-\alpha_2}\right) = \alpha_t \tag{4}$$

where α_t is the penalized type I error rate for using a given t-fraction of internal information. The expectation is taken with respective to the posterior joint distribution $G_t(\Delta_1, \Delta_2)$ given x_{1t}, x_{2t}, and any available prior (external) information for the treatment effects.

At $t = 0$, the above strategy maximizes the probability of success, or the Bayesian expected power given the external information about the treatment effects in the overall population and the biomarker subgroup. A particularly interesting special case arises when $I^{(e)} = 0$. The alpha allocation becomes completely auto-adaptive based on a fraction of the internal data when no external data is available.

An alternative strategy to utilize partial trial data for alpha allocation is to reserve a fraction of data to estimate the treatment effects and biomarker prevalence, and use the remaining data for the hypothesis testing. In section 3, we show that this approach is sensitive to the information fraction of the partial data, and in general less powerful than the above adaptive allocation strategy.

For a pair of true treatment effect parameters (δ_1, δ_2), the unconditional power is the expectation of (2) with respect to the partial data (X_{1t}, X_{2t}), with (α_1, α_2) determined by (3) and (4). Since for every possible values of (X_{1t}, X_{2t}), a different set of (α_1, α_2) needs to be derived by an optimization procedure, the calculation of unconditional power is computationally demanding. We have made extensive use of parallel computing on a high-performance computing cluster for the numerical examples in Section 3.

2.2. Penalized Type I Error Rate

Denote $(\tilde{\alpha}_{1,t}, \tilde{\alpha}_{2,t}) = \arg\max Q(\alpha_1, \alpha_2; x_{1t}, x_{2t}, \alpha_t)$. The actual type I error under the global null hypothesis is:

$$P(\alpha_t) = \int \left[1 - \Phi_{\sqrt{r}}\left(\frac{Z_{1-\tilde{\alpha}_{1,t}} - \sqrt{t}x_{1t}}{\sqrt{1-t}}, \frac{Z_{1-\tilde{\alpha}_{2,t}} - \sqrt{t}x_{2t}}{\sqrt{1-t}}\right)\right] \times \phi_{\sqrt{r}}(x_{1t}, x_{2t})\, dx_{1t}x_{2t},$$

where $\phi_{\sqrt{r}}$ is the probability density function for a bivariate normal distribution with correlation of $\sqrt{r}$. For any $t > 0$, $P(\alpha_t) > \alpha_t$. Intuitively, this is due to the adaptive nature of the alpha allocation procedure, which attempts to assign more alpha to the hypothesis that exhibits more promising trend in the partial data. Under the null hypotheses, such a procedure will inflate type I error. If we can calculate $P(\alpha_t)$ for each value of α_t, an $\alpha_t < \alpha$ can be identified so that $P(\alpha_t) = \alpha$ (see Chen & Beckman, 2009 for a numerical procedure to calculate α_t). This can be interpreted as a penalty on the type I error rate for using internal data to inform the alpha allocation. To maintain the strong control of the familywise type I error, the auto-adaptive strategy needs to operate under the constraint (4) using the penalized type I error rate α_t.

The penalized Type I error rate at an information fraction t depends on the information ratio r between the subgroup and the overall population, and the ratio between the external/prior information $I^{(e)}$ and the total current trial information $I^{(i)}$. It also depends on the family of distributions to which $g(\Delta_1, \Delta_2)$ belongs. Figure 1 illustrates α_t at different ratios of $I^{(e)}:I^{(i)}$, assuming $r = 0.5$, and $g(\Delta_1, \Delta_2)$ takes the shape of a bivariate normal distribution. In all cases, α_t decreases when more internal data are used to inform the alpha allocation as expected. Comparing different ratios of prior versus current trial information $I^{(e)}:I^{(i)}$, the most stringent penalty is taken when there is no external information (i.e. $I^{(e)}:I^{(i)} = 0:80$). With more external information available (i.e. $I^{(e)}:I^{(i)} = 5:80$ or $15:80$), the optimization depends less on the internal data, and the type I error rate is less penalized.

The auto-adaptive strategy pays a penalty on the type I error rate for the flexibility of allocating alpha based on external and/or partial internal data. Consequently the overall study power under the auto-adaptive strategy will always be lower than the maximal power that can be achieved if the true treatment effects are known and no alpha penalty is taken. Such a tradeoff is commonly seen in adaptive trials. An example is the seamless design that picks the winners from the dose-ranging stage of the study: if the final analysis of best treatment vs. control comparison includes data from the first stage, appropriate adjustment to the critical values or weighting of the data from two stages are needed to control the type I error level.

.

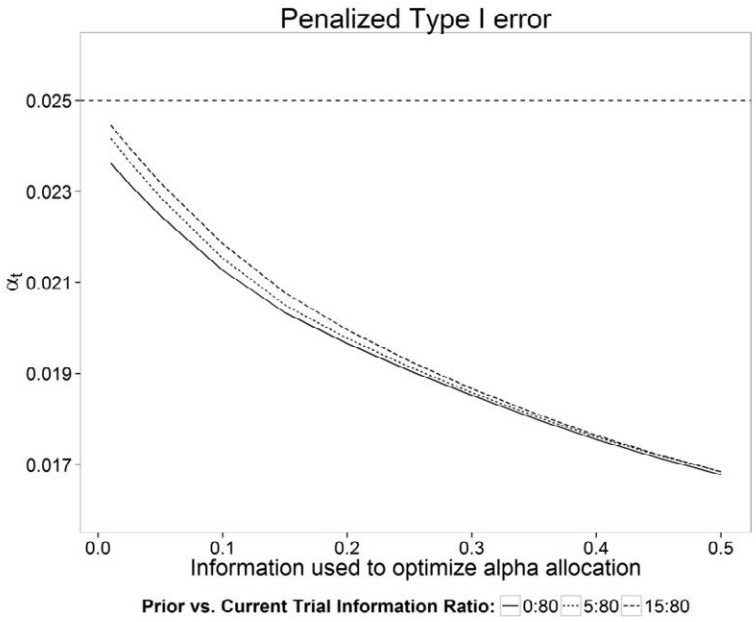

Figure 1: Nominal Type I error for intersection hypothesis $H_0^{(1)} \cap H_0^{(2)}$ after using a fraction of trial data to inform the alpha allocations. Information ratio between the subgroup and the overall population is $r = 0.5$.

3. Case Studies

Consider a Phase III oncology trial with two arms: the experimental treatment, and the control. Patients are randomized to the two treatment arms according to 1:1 allocation ratio. Assume the trial is designed to provide $I^{(i)} = 80$ units of information for the overall population. The information ratio between the biomarker subgroup and the overall population is expected to be $r = 0.5$, therefore the biomarker subgroup may have 40 units of information. If the true treatment effect parameter $\delta_1 = 0.36$, this study has 88% power to reject the null hypothesis of no treatment effect in the overall population ($\alpha = 0.025$ is allocated to the overall hypothesis). For an oncology trial with terminal events as the clinical endpoint, this is equivalent to detecting a hazard ratio of 0.7 in a study with 320 events (e.g. overall survival or progression-free survival) between two treatment arms.

The optimal alpha allocation depends on the treatment effect in the biomarker subgroup and the overall population, as well as the prevalence of the biomarker. For example, if the true treatment effect parameters $\delta_1 = 0.36$, $\delta_2 = 0.60$, and the information ratio is $r = 0.5$, the power can be maximized by allocating $\alpha_1 = 0.0033$ to the overall population, and $\alpha_2 = 0.0236$ to the subgroup; on the other hand, if $\delta_1 = \delta_2 = 0.36$, the maximal power will be achieved by allocating all 0.025 to α_1. In practice, there will

always be considerable uncertainty about these treatment effects. An auto-adaptive strategy can estimate the unknown treatment effect using data from the same trial that supports the formal hypothesis testing.

3.1 Auto-adaptation Using Current Trial Information

While mature Phase II data would provide the best external information for the planning of Phase III trials, in some cases Phase III trials may be directly triggered by promising Phase I results on surrogate endpoints. This is often seen in therapeutic areas with urgent needs for new life-saving treatments. Better response in a surrogate endpoint from a subset may prompt interest in this subgroup, but a sensible statistical strategy for the Phase III trial should be informed by direct evidence on the clinical endpoint. In this section, we will assess the auto-adaptive strategy to allocate alpha based on partial internal data alone in the Phase III trial.

Since the effect size from the subgroup relative to the overall population is unknown, we cannot optimize the alpha allocation based on external data. In practice two ordered testing strategies are often used: 1) allocating all alpha to the overall population, and step-down to the testing of hypothesis in the subgroup if the trial is positive in the overall population; 2) allocating all alpha to the subgroup, and test the hypothesis in the overall population if and only if the trial is positive in the subgroup. Both strategies maintain the strong control of familywise type I error.

Figure 2 illustrates the performance of the alpha allocation strategies when the biomarker subgroup has significant stronger treatment effect ($\delta_2 = 0.60$) compared to the overall population ($\delta_1 = 0.36$). In this case, allocating more alpha to the biomarker subgroup hypothesis improves the overall study power. The auto-adaptive allocation strategy correctly identifies the trend, and the overall study power is significantly improved compared to the strategy that prioritizes the overall population. It is interesting to note that the information fraction to be used to inform the auto-adaptive allocation has minor impact on the overall study power. In contrast, if the partial data used to inform alpha allocation are excluded from the hypothesis tests, the power decreases quickly as more data are used to estimate the treatment effects and biomarker prevalence, as shown by the dot-dash line in Figure 2.

In Figure 2 the biomarker-focused strategy has the highest power. Such strategy, however, carries significant risk. Figure 3 presents the two scenarios in which a biomarker-focused strategy has low probability of success:

a. Prevalence of biomarker-expressed subset is 50% as expected, but biomarker subgroup does not have improved treatment effect over the overall population.
b. Prevalence of biomarker-expressed subset is 20%, lower than the expected 50%, although biomarker subgroup does have significantly better treatment effect than the overall population

In both Figure 3a and 3b, the auto-adaptive strategy appropriately adjusts the alpha allocation according to the treatment effect and the actual biomarker prevalence. The overall power is significantly higher than the biomarker-focused strategy. Similar to Figure 2, the information fraction used for such auto-adaptation does not significantly impact the overall study power.

The auto-adaptive strategy does not have the highest power in either Figure 2 or 3. This is not surprising considering the type I error penalty paid for using partial trial data. The non-adaptive strategy may achieve better power in a particular case, but performs poorly in other cases. The auto-adaptive strategy

will always be reasonably close to the highest power one can achieve in any given scenario. The global robustness of the auto-adaptive strategy is a desirable property.

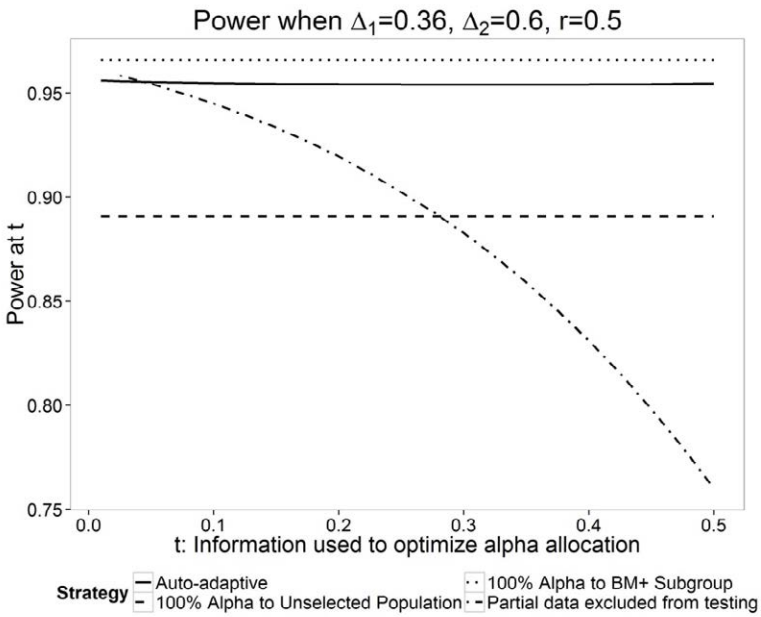

Figure 2: Comparison of overall study power when biomarker subgroup exhibits better treatment effect compared to overall population ($\delta_2 = 0.6$ vs $\delta_1 = 0.36$), and the prevalence of biomarker positive patients is 50% of overall population.

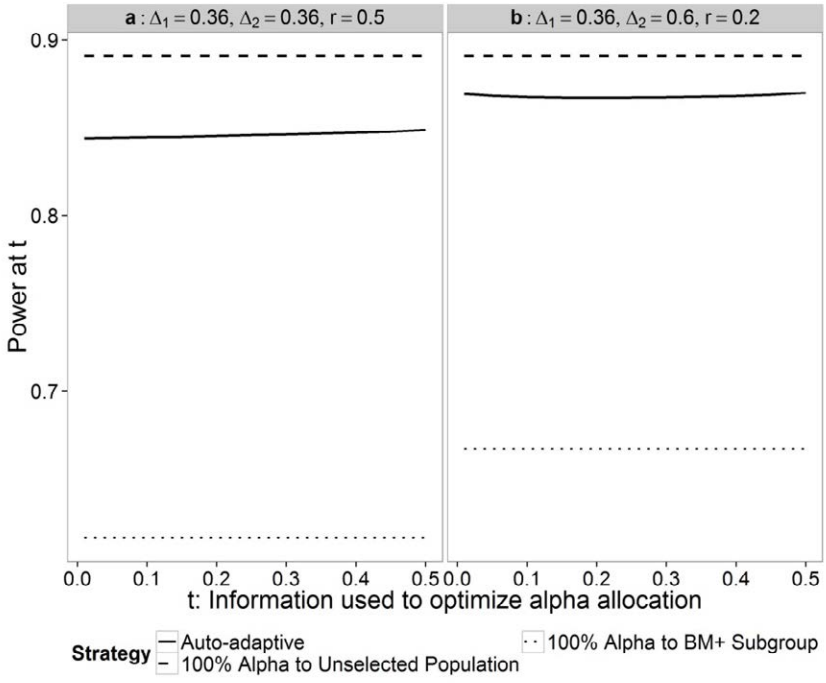

Figure 3: Comparison of overall study power when a) biomarker does not correlate with treatment effect ($\delta_1 = \delta_2 = 0.36$), and b) biomarker subgroup has better treatment effect ($\delta_1 = 0.36$ vs. $\delta_2 = 0.6$), but biomarker prevalence is lower than expected (20% vs. 50%).

3.2 Auto-adaptation vs. Optimization Using External Information

In other cases, some direct or indirect evidence on the Phase III clinical endpoint is available from Phase II trials or reports on drugs in the same therapeutic class. Such information can be used to form an informative prior for the treatment effects in question. In addition to the ordered testing strategy as discussed in Section 3.1, we may also optimize the alpha allocation based on the prior distribution. For example, the special case of (2) when $t = 0$ can be used to maximize the expected power based on the prior distribution $G_0(\Delta_1, \Delta_2)$, and the estimated information ratio $\hat{r}$ between the subgroup and the overall population.

$$Q_0(\alpha_1, \alpha_2) = \int \left\{1 - \Phi_{\sqrt{\hat{r}}}\left(Z_{1-\alpha_1} - \sqrt{I^{(i)}}\Delta_1,\ Z_{1-\alpha_2} - \sqrt{\hat{r}I^{(i)}}\Delta_2\right)\right\} dG_0(\Delta_1, \Delta_2) \tag{5}$$

For ease of exposition, we assume that the Z test statistics for (Δ_1, Δ_2) based on the external data are:

$$\begin{pmatrix} X_1^{(e)} \\ X_2^{(e)} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \Delta_1\sqrt{I^{(e)}} \\ \Delta_2\sqrt{\hat{r}I^{(e)}} \end{pmatrix},\ \begin{pmatrix} 1 & \sqrt{\hat{r}} \\ \sqrt{\hat{r}} & 1 \end{pmatrix}\right),$$

which in turn gives the prior distribution of (Δ_1, Δ_2) for the Phase III trial:

$$\begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix} \Big| X_1^{(e)}, X_2^{(e)} \sim MVN\left(\begin{bmatrix} X_1^{(e)}/\sqrt{I^{(e)}} \\ X_2^{(e)}/\sqrt{\hat{r}I^{(e)}} \end{bmatrix},\ \begin{bmatrix} 1/I^{(e)} & 1/I^{(e)} \\ 1/I^{(e)} & 1/(\hat{r}I^{(e)}) \end{bmatrix}\right) \tag{6}$$

The utility of auto-adaptive strategy in the presence of prior information depends on the strength and accuracy of the external data. In general, when the external information on the key trial parameters is reliable, it is sufficient to use an informative prior to guide the alpha allocation to the two hypotheses of interest. The potential gain in power by using additional internal information for the alpha allocation has to be balanced against the alpha penalty. Nevertheless, in some cases where the prior is weak and its point estimates are potentially biased, the auto-adaptive alpha allocation could help mitigate the risk.

In Figure 4, we consider a study where the overall population treatment effect $\Delta_1 = 0.36$, the subgroup treatment effect $\Delta_2 = 0.60$, and the information ratio $r = 0.3$ between the subgroup and the overall population. The overall study power at different levels of overall nominal alpha level is illustrated, corresponding to the cases where 0%, 10% and 20% of current trial data are used to allocate alpha by optimization of (3). The overall study power that can be achieved depends on the alpha allocated to the overall population (which in turn determines the alpha for the subgroup hypothesis based on (4)). The total nominal alpha available for the allocation decreases as the fraction of trial information used for alpha allocation increases.

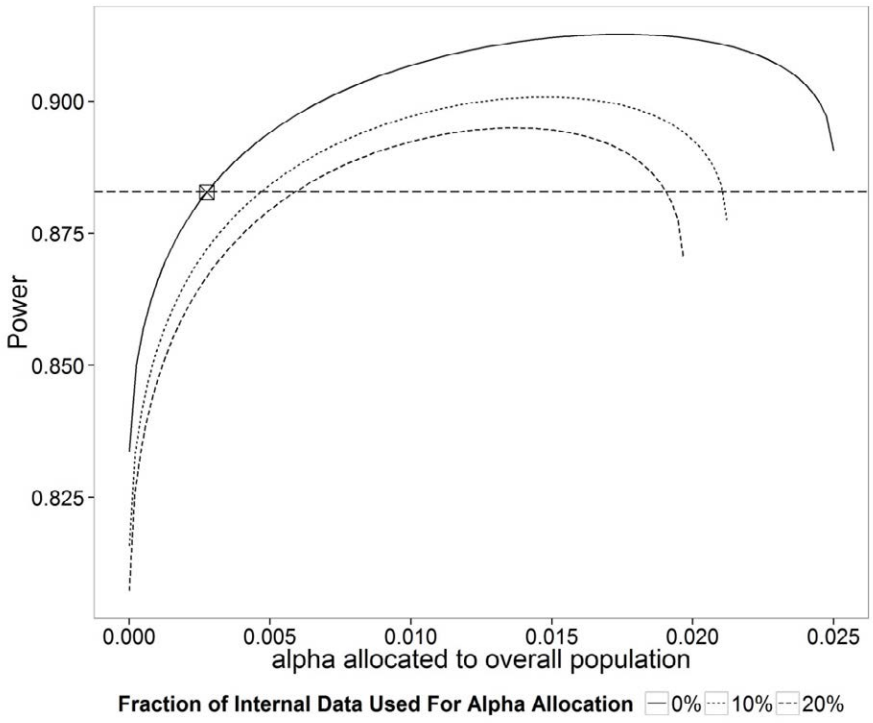

Figure 4: Comparison of overall study power achieved by allocating different levels of alpha to the overall population, when 0%, 10% and 20% of current trial data are used to inform the alpha allocation to overall population hypothesis and the subgroup hypothesis. The point marked on the solid line is the alpha allocation calculated based on a prior distribution that indicates strong subgroup effect.

Considering a scenario where a possible prior for the treatment effects is

$$MVN\left(\begin{pmatrix} 0.22\sqrt{I^{(e)}} \\ 0.69\sqrt{\hat{r}I^{(e)}} \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{\hat{r}} \\ \sqrt{\hat{r}} & 1 \end{pmatrix}\right),$$

where $\hat{r} = 0.5$, and $I^{(e)} = 15$. This corresponds to a realistic case where the external information prior to the Phase III trial points to a more significant subgroup treatment effect, and a higher prevalence for the subgroup identifying biomarker in the general patient population.

Using this misleading prior, the optimization for (5) would allocate alpha of 0.002 to the overall population, and 0.0239 to the subgroup. The power achieved by this alpha allocation is 88%, as marked by the horizontal dashed line in Figure 4. In contrast, the power achieved by the auto-adaptive alpha allocation using 10% or 20% of the current trial data is approximately 90%. Since the adaptive procedure depends on the partial data $x_{1,t}, x_{2,t}$, the alpha allocated to the overall population cannot be illustrated in Figure 4 by a stationary point. Note that the power achieved by adaptive alpha allocation using 20% of trial information in this example is 90%, while the maximal power from any fixed alpha allocation resulting from auto-adaptive allocation, after penalizing alpha for the auto-adaptation, is below 90%. A possible explanation for this phenomenon is that the power curves in Figure 4 are generated by assuming the auto-adaptive procedure had generated a fixed alpha allocation unrelated to the observed $x_{1,t}, x_{2,t}$, while the proposed procedure maximizes the expected conditional power based on the observed partial data. Although they are under the same alpha constraint by $\alpha_t|t = 0.2$, the adaptive procedure optimizes

the alpha allocation for each specific incidence of the partial data, which also contributes to the final test statistics for the entire trial.

As another example, suppose the overall population treatment effect $\delta_1 = 0.22$, the subgroup treatment effect $\delta_2 = 0.79$, and the information ratio $r = 0.2$ between the subgroup and the overall population. The subgroup effect is highly significant, but the prevalence of the subgroup in the final data is only 20%, i.e. $r = 0.2$. The prior distribution for the treatment effect, on the other hand, could be:

$$MVN\left(\begin{pmatrix} 0.43\sqrt{I^{(e)}} \\ 0.51\sqrt{\hat{r}I^{(e)}} \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{\hat{r}} \\ \sqrt{\hat{r}} & 1 \end{pmatrix}\right),$$

where $\hat{r} = 0.5$, and $I^{(e)} = 15$. These external data support the prior assumption of a moderate subgroup effect, and 50% prevalence for the biomarker.

Using this prior, the optimization for (5) would allocate alpha of 0.018 to the overall population, and 0.009 to the subgroup. The power achieved by this alpha allocation is 83%, marked by the horizontal dashed line in Figure 5. In contrast, the power achieved by the auto-adaptive alpha allocation using 10% or 20% of the current trial data is approximately 86%. Again, the power achieved by the auto-adaptive procedure is close to or even exceeds the maximal power achieved by fixed alpha allocation under the same nominal alpha constraint.

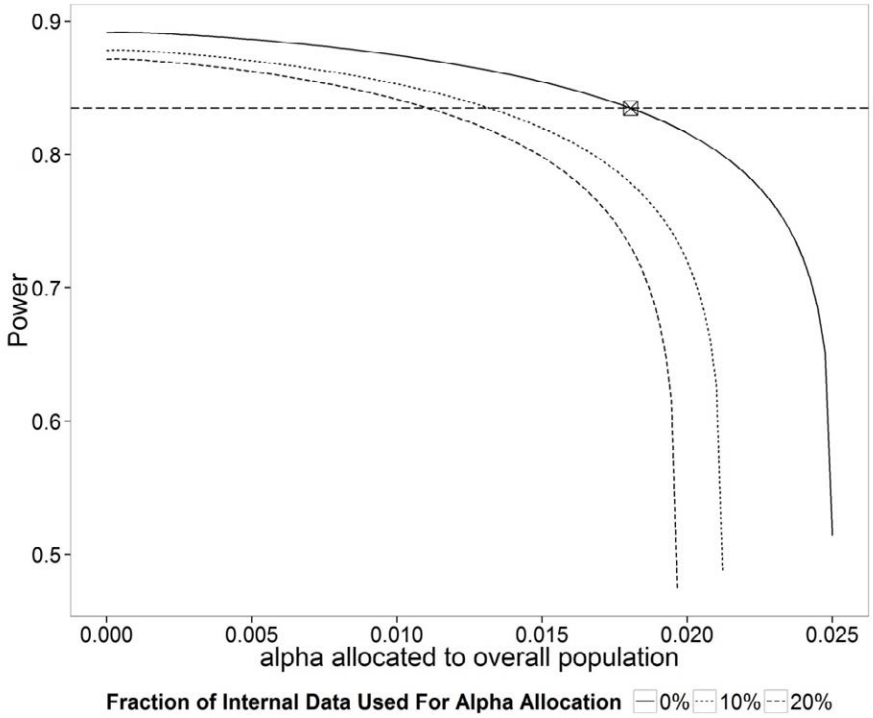

Figure 5: Comparison of overall study power achieved by allocating different levels of alpha to the overall population, when 0%, 10% and 20% of current trial data are used to inform the alpha allocation to overall population hypothesis and the subgroup hypothesis. The point marked on the solid line is the alpha allocation calculated based on a prior distribution that indicates moderate subgroup effect and higher prevalence.

Note for the two scenarios we considered, the conventional approach would likely allocate all alpha to the subgroup or the overall population completely based on the prior data, which would lead to further power loss when the actual treatment effect deviates from our prior assumption. The adaptive allocation strategy could ensure that reasonable power is maintained.

4. Discussion

We have demonstrated through a series of examples that the auto-adaptive alpha allocation based on the optimization of a Bayesian conditional power function can mitigate the risk of incorrect assumptions on critical trial parameters. Notwithstanding the penalty on nominal type I error one has to pay for using partial trial information, the power remains stable as the hypothesis testing strategy is adaptive to the observed data. The benefit of auto-adaptation is particularly evident when the predictive effect of a biomarker or the prevalence of the biomarker is not well understood.

Although the auto-adaptation may occur at an interim analysis, the methodology presented in Section 2 does not stipulate when and how the partial data should be chosen. In theory, a representative subset of the final data upon conclusion of the Phase III trial can be selected. This is potentially helpful to alleviate the temporal bias that is sometimes observed in survival trials when sicker patients tend to enroll early, or when standard of care progresses over the duration of the trial. In addition, the methodology discussed in Section 2 can even accommodate the use of the full trial data for $G_t(\Delta_1, \Delta_2 | x_{1,t}, x_{2,t})$ at $t = 1$, where the penalized nominal alpha level is 1.5%. The use of full data to inform alpha allocation may seem controversial, but a parallel might be drawn with the Hochberg step-up method. In both approaches, the testing strategies are adaptive to the test statistics, or the corresponding p-values. The adaptive alpha allocation in this case simply makes use of the distributional assumptions for the test statistics in the subgroup and overall population.

In this article we have only discussed the alpha allocation for the hypotheses in the overall population and a pre-specified subgroup. In practice a Phase III trial may consider more than one tentative cut-point for the biomarker expressions; the choice of target subgroup (e.g. patients with any positive biomarker expressions, or patients with very strong expressions in the biomarker) may become a data-driven decision. Whether partial data from the current trial can be used to inform both the cut-point decision and the subsequent alpha allocation is further discussed in [15].

References

[1] Kelly K, Altorki NK, Eberhardt WE, et al: A randomized, double-blind phase 3 trial of adjuvant erlotinib versus placebo following complete tumor resection with or without adjuvant chemotherapy in patients with stage IB-IIIA EGFR positive (IHC/FISH) non-small cell cancer: RADIANT results. ASCO Annual Meeting. Abstract 7501. Presented June 2, 2014.

[2] Tsao, M.S. et al. (2005), “Erlotinib in Lung Cancer—Molecular and Clinical Predictors of Outcome,” New England Journal of Medicine, 353, 133–144. 431

[3] Yang, C. H., Yu, C. J., Shih, J. Y., Chang, Y. C., Hu, F. C., Tsai, M. C., ... & Yang, P. C. (2008). Specific EGFR mutations predict treatment outcome of stage IIIB/IV patients with chemotherapy-naive non–small-cell lung cancer receiving first-line gefitinib monotherapy. Journal of Clinical Oncology, 26(16), 2745-2753.

[4] Sequist, L. V., Martins, R. G., Spigel, D., Grunberg, S. M., Spira, A., Jänne, P. A., ... & Lynch, T. J. (2008). First-line gefitinib in patients with advanced non–small-cell lung cancer harboring somatic EGFR mutations. Journal of Clinical Oncology, 26(15), 2442-2449.

[5] Paez, J. G., Jänne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., ... & Meyerson, M. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science, 304(5676), 1497-1500.

[6] Wang, S., O'Neill, R. T., & Hung, H. M. J. (2007). Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. Pharmaceutical Statistics, 6, 227–244.

[7] Freidlin, B., & Simon, R. (2005). Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. Clinical Cancer Research : An Official Journal of the American Association for Cancer Research, 11(21), 7872–8.

[8] Friede, T., Parsons, N., & Stallard, N. (2012). A conditional error function approach for subgroup selection in adaptive clinical trials. Statistics in Medicine, 31, 4309–20.

[9] Mehta, C., Schäfer, H., Daniel, H., & Irle, S. (2014). Biomarker driven population enrichment for adaptive oncology trials with time to event endpoints. Statistics in Medicine, 33(26), 4515–31.

[10] Brannath, W., Zuber, E., Branson, M., Bretz, F., Gallo, P., Posch, M., & Racine-poon, A. (2009). Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. Statistics in Medicine, 28, 1445–1463.

[11] Jenkins, M., Stone, A., & Jennison, C. (2011). An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharmaceutical Statistics, 10(4), 347–56.

[12] Simes, R. J. (1986). An improved bonferroni procedure for multiple tests of significance. Biometrika, 73(3), 751–754.

[13] Spiessens B, DeboisM. Adjusted significance levels for subgroup analysis in clinical trials. Contemporary Clinical Trials 2010; 31:647–656.

[14] Chen, C., & Beckman, R. A. (2009). Hypothesis Testing in a Confirmatory Phase III Trial With a Possible Subset Effect. Statistics in Biopharmaceutical Research, 1(4), 431–440.

[15] Chen, C., Li, N., Shentu Y., Pang, L., Beckman, R. A. (2015) Informational Design of Confirmatory Phase III Trials for Expedited Development of Personalized Medicines. Unpublished manuscript