Published in final edited form as:
Clin Trials. 2010 October ; 7(5): 597–604. doi:10.1177/1740774510381285.

Prospective alpha allocation in the clarification of optimal anticoagulation through genetics (COAG) trial

Jungnam Joo[a], **Nancy L Geller**[a], **Benjamin French**[b], **Stephen E Kimmel**[b], **Yves Rosenberg**[c], and **Jonas H Ellenberg**[b]

[a]Office of Biostatistics Research, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA

[b]Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

[c]Atherothrombosis and Coronary Artery Disease Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA

Abstract

Background—The Clarification of Optimal Anticoagulation through Genetics (COAG) trial is a large, multicenter, double-blinded, randomized trial to determine whether use of a genotype-guided dosing algorithm (using clinical and genetic information) to initiate warfarin treatment will improve anticoagulation status when compared to a dosing algorithm using only clinical information.

Purpose—This article describes prospective alpha allocation and balanced alpha allocation for the design of the COAG trial.

Methods—The trial involves two possibly heterogeneous populations, which can be distinguished by the difference in warfarin dose as predicted by the two algorithms. A statistical approach is detailed, which allows an overall comparison as well as a comparison of the primary endpoint in the subgroup for which sufficiently different doses are predicted by the two algorithms. Methods of allocating alpha for these analyses are given – a prospective alpha allocation and allocating alpha so that the two analyses have equal power, which we call a `balanced alpha allocation.'

Results—We show how to include an analysis of the primary endpoint in a subgroup as a co-primary analysis. Power can be improved by incorporating the correlation between the overall and subgroup analyses in a prospective alpha allocation approach. Balanced alpha allocation for the full cohort and subgroup tests to achieve the same desired power for both of the primary analyses is discussed in detail.

Limitations—In the COAG trial, it is impractical to stratify the randomization on subgroup membership because genetic information may not be available at the time of randomization. If imbalances in the treatment arms in the subgroup are found, they will need to be addressed.

Conclusions—The design of the COAG trial assures that the subgroup in which the largest treatment difference is expected is elevated to a co-primary analysis. Incorporating the correlation between the full cohort and the subgroup analyses provides an improvement in power for the subgroup comparison, and further improvement may be achieved via a balanced alpha allocation

Author for correspondence: Jungnam Joo, Office of Biostatistics Research, National Heart, Lung and Blood Institute, National Institutes of Health, 6701 Rockledge Drive, MSC 7913, Bethesda, Maryland, USA. jooj@nhlbi.nih.gov.

approach when the parameters involved in the sample size calculation are reasonably well estimated.

Introduction

This article describes statistical approaches used in the design of a clinical trial comparing the dosing of warfarin via algorithms which use clinical and genetic information versus algorithms which use clinical information alone.

Warfarin is an oral anticoagulant (or blood thinner) commonly used to treat thrombosis (or blood clots), such as deep vein thrombosis and pulmonary embolus, and to prevent the formation of a thrombus in high-risk individuals, such as those with atrial fibrillation and mechanical heart valves. The drug is difficult to dose properly because life-threatening complications may arise from over-dosing and inadequate efficacy may result from under-dosing. The dose requirements for different patients, however, can vary over 30-fold, from 0.5 mg daily to 15 mg daily. Currently, warfarin treatment is initiated at an empiric dose, typically 5 mg, and then monitored frequently using a normalized prothrombin time, also known as the INR (International Normalized Ratio). The INR reflects the ratio of the time of a patient's blood to clot compared to a normal control subject. The therapeutic range for warfarin is narrow (typically an INR between 2 and 3), which further complicates proper warfarin management.

Many clinical and environmental factors influence warfarin response [1]. Despite knowledge of these factors, a large portion of the variability in warfarin dose required to maintain the INR in the therapeutic range remains unexplained. Recent studies of the effects of genetic polymorphisms on the pharmacokinetics and pharmacodynamics of warfarin have offered some insight. Extensive studies have focused on two genes [2–4]. The first is Cytochrome P450 2C9 (*CYP2C9*), which is related to warfarin pharmacokinetics, or how the body metabolizes warfarin. The second is Vitamin K Epoxide Reductase Complex-1 (*VKORC1*), which is related to warfarin phamacodynamics, or how warfarin works on the body. Two small trials of genetic-based dosing were recently conducted [5,6]. The results suggested that prospective genetic testing may enhance the safety, accuracy, and efficiency of warfarin dosing and demonstrated the need for larger scale studies.

The Clarification of Optimal Anticoagulation through Genetics (COAG) trial (clinicaltrials.gov identifier: NCT00839657) is a large, multicenter, double-blinded, randomized trial to determine whether use of genotype-guided dosing algorithms (using clinical and genetic information) to initiate warfarin treatment (the first 4–5 days of therapy) will improve anticoagulation control compared to dosing algorithms using only clinical information. Dosing on days 1–3 will be determined using a dose-initiation algorithm, and a dose-revision algorithm will be applied on day 4 and/or 5 for each treatment arm [7]. Following this dose-initiation phase, patients in both arms will have subsequent dose titrations based on the same standardized algorithm. The primary outcome of the COAG trial is the percentage of time within the therapeutic INR range (PTTR) during the first 4 weeks of therapy, which will be calculated using linear interpolation [8]. PTTR is the most widely accepted measure of improvement in warfarin management [9].

Certain assumptions underlie the design of the COAG trial. First, if a patient receives his or her long-term maintenance dose at the initiation of warfarin treatment, the time to reach the maintenance dose will be shorter, and the variability in INRs will be reduced. Second, if the maintenance dose is achieved quickly with less variability, the risk of over- and under-anticoagulation will be reduced. Based on these assumptions, the COAG trial hypothesizes that the additional use of genetic information in algorithms that predict maintenance dose

will provide initial doses closer to the maintenance dose more frequently than using clinical information alone, thereby increasing the percentage of time within the therapeutic INR range.

In this article, we describe statistical approaches to include a subgroup comparison as a co-primary analysis in the COAG trial. The subgroup of interest is those participants with large differences between the predicted initial doses calculated by two dosing algorithms. The trial uses a marker-based strategy [10] or biomarker-guided [11] design. In a typical marker-based strategy design, the subgroup of patients who are marker-positive will have greater benefit from the marker-based intervention. In the next section, we first explain an important baseline-defined enriched subgroup in the COAG study population that acts as a so-called `marker-positive group.' A prospective alpha allocation approach that formally allows for the evaluation of the treatment benefit in an enriched subgroup as a co-primary endpoint is then explained [12,13]. Sample size and power considerations of the COAG trial are then presented followed by a method that improves power by incorporating the correlation between the two tests of the primary hypotheses. Finally, we propose allocating overall alpha to the full cohort and subgroup comparisons to achieve the same power for both of the primary analyses. We call this `balanced alpha allocation.'

General statistical design considerations

An enriched subgroup in the COAG study population

In the COAG trial, we defined the enriched subgroup as those participants with large differences between the predicted initial doses calculated by two dosing algorithms. In previous studies [5], it has been shown that patients with certain genetic profiles tend to require extreme doses. For example, patients without any genetic variants of interest tend to require larger doses whereas multiple variant carriers tend to require smaller doses. The enriched subgroup was not defined directly based on the genetic profile, but rather indirectly based on the degree of treatment difference guided by the genetic profile, although dose differences are associated with, but not perfectly predicted by, the number of genetic variants. Consequently, it is expected that the benefit of using the genotype-guided dosing algorithm will be much greater for patients in the enriched subgroup because these participants will receive a more distinct intervention based on their treatment assignment compared to the remaining patients. The clinically important dose difference was chosen as 1 or more milligrams per day because a 1 mg per day change in warfarin dose from the average starting dose of 5 mg per day in many US anticoagulation clinics is sufficient to meaningfully change the INR [14–16]. Even though the genotypes and thus the difference between predicted doses using the two algorithms may not be available at the time of randomization for some patients, the dose difference is a factor fixed at the time of randomization. Therefore it defines a proper subgroup that is not based on post-randomization selection. In addition, genotype availability is independent of treatment assignment as it depends on the time of day at which the participant arrives. Some degree of dilution of treatment effect might occur due to the delayed genotyping, and this can be incorporated when calculating the sample size.

Prospective alpha allocation

In most clinical trials, subgroup analyses are typically regarded as secondary in importance. In the COAG trial, the subgroup of patients with large differences in predicted initial doses is a sub-population that may primarily drive the overall difference between the two arms. Therefore, to make a formal evaluation of the treatment benefit in this enriched subgroup, we considered a prospective alpha allocation approach [12,17–19]. Prospective alpha allocation allows a formal evaluation of the treatment effect in the subgroup as part of the

primary analysis. By denoting the treatment effect in the full cohort and the subgroup by Δ_a and Δ_s, the null and alternative hypotheses are $H_0 : \Delta_a = 0$ and $\Delta_s = 0$ versus $H_1 : \Delta_a \neq 0$ or $\Delta_s = 0$. The comparison of the two arms in the full cohort ($\Delta_a = 0$) will be tested at α_a, which is less than the overall significance level α, and the remaining $\alpha - \alpha_a = \alpha_s$ is used for testing a subgroup of the cohort ($\Delta_s = 0$). In the COAG trial, $\alpha_a = 0{:}04$ and $\alpha_s = 0.01$ are considered as in Freidlin and Simon [19] for the purpose of sample size calculation.

When a sub-population more likely to benefit from the intervention can be easily identified and an accurate assay is readily available for selecting such a sub-population, a targeted design that only recruits this sub-population may be more efficient than an untargeted design [17,18]. In the COAG trial, even though the enriched subgroup is one that might have greater benefit from the genotyping, we are uncertain about the degree of treatment benefit for patients who do not belong to the subgroup. This is because an additional dose revision algorithm, based on clinical and genetic factors for one arm and clinical factors alone for the other arm, will be applied to each respective arm after the initiation dose. Moreover, the initial dose is not identical for those who do not belong to the enriched subgroup, and this allows for the possibility of some difference in baseline doses. Specially because the genotyping cost has to be spent up front before randomization, we chose to study the whole population of patients who need warfarin so that it is possible to examine the mechanisms of action for any differences observed between two treatment arms, and also between two subgroups depending on the definition of the subgroup. Our motivation is to examine both the entire population and an enriched sample within the bounds of statistical propriety for protection of type-I error, to increase our chance of fleshing out whether there is a signal with this approach. The inclusion of the whole population provides an enhancement that will allow the study to delve into any differences observed as well as allowing for post randomization secondary analyses that may benefit from redefining subgroups.

Sample size and power

The sample size calculation of the COAG trial was based on the following assumptions. First, approximately 60% of participants belong to the enriched subgroup and their mean PTTR is relatively low (61%). We denote the proportion of the patients in the enriched subgroup by p. The genotype-guided dosing algorithm will be considered efficacious if it results in a 15% relative increase in PTTR for this subgroup of patients. For the remaining 40% of participants who already have rather high mean PTTR of 73%, the genotype-guided dosing could be beneficial but we consider the benefit to be relatively minimal. To be conservative in estimating the overall treatment effect, we assumed the same PTTR for these patients regardless of their treatment assignment. The estimates of mean PTTRs and the proportion of patients in the enriched subgroup are based on the data from the Couma-Gen trial [5] and the International Warfarin Pharmacogenetics Consortium [16]. This yields an absolute difference of $\Delta_a = 5{:}49\%$ in PTTR between the two arms for the full cohort, and $\Delta_s = 9{:}15\%$ for the enriched subgroup. Further assumptions in the computation of sample size include the standard deviation of the PTTR and the expected dropout rate of participants after randomization. Based on the literature, 25% is assumed as the standard deviation [7,20,21], and a conservative estimate of 10% is assumed for the dropout rate.

A sample size of 1140 will provide 90% power to detect a 5.49% difference in the PTTRs between the two treatment arms (two-sided test) for the full cohort analysis using $\alpha_a = 0{:}04$ given 25% standard deviation, 10% dropout rate and $p = 0.6$. For the subgroup analysis to have 90% power using $\alpha_s = 0.01$, a total of 916 (550 in the subgroup) patients are required. To ensure the power of 90% for both of the primary analyses, the sample size of the study is determined by the maximum of these two sample size calculations. Thus, the full cohort analysis drives the study sample size in our study. Taking a conservative approach to protect against errors in the assumed parameters, a sample size of 1238 was chosen. This sample

size ensures at least 80% power for the full cohort analysis when the proportion of the enriched subgroup is as low as 50% or when the standard deviation is as high as 30%.

The power to detect an absolute increase of 9.15% using $\alpha_s = 0.01$ given a total study sample size of 1238 is given in Table 1. Here $p = 0.6$ and a slightly conservative estimate of 0.55 is considered. The power of the subgroup analysis is generally greater than 80%. Under the currently assumed parameters, the power of the subgroup analysis is over 97%.

Improving power considering the correlation

In a prospective alpha allocation approach, allocating α so that the sum of the type-I errors is equal to the overall type-I error is a Bonferroni-type correction, which is unnecessarily conservative when there is a positive correlation between the two tests. The correlation between the full cohort and subgroup tests can be obtained under the null hypothesis given the proportion of the subgroup p [22,23]. Then, for a fixed α_a, we can obtain α_s while controlling the overall type-I error at a desired level α.

Consider Z_a and Z_s, the test statistics for the overall and the subgroup analyses, respectively. Let X_{ij} be the PTTR for subgroup i ($i = s, s^c$) and individual j ($j = 1, \ldots, n_{iX}$) in the genotype-guided dosing algorithm arm, and Y_{ij} be the same for clinical-guided dosing algorithm arm. Denote the standard deviation of X_{ij} (and Y_{ij}) as σ_{ix} (and σ_{iy}) for $i = s, s^c$. Note that $i = s$ represents the enriched subgroup defined by ≥ 1 mg difference between the predicted initial doses from the two algorithms. Denote the number of patients in subgroup i for each arm as n_{iX} and n_{iY}. The total study sample size is $n_X + n_Y$ where $n_X = \Sigma_i n_{iX}$ and $n_Y = \Sigma_i n_{iY}$. Using $p_X = n_{sX}/n_X$ and $p_Y = n_{sY}/n_Y$, we can then obtain the correlation (ρ) between Z_a and Z_s under the null hypothesis using:

$$Z_a = \frac{\sqrt{n_X n_Y}}{\sqrt{\left(n_X \sigma_{ay}^2 + n_Y \sigma_{ax}^2\right)}} \left(\bar{X}_{..} - \bar{Y}_{..}\right)$$

$$Z_s = \frac{\sqrt{n_{sX} n_{sY}}}{\sqrt{\left(n_{sX} \sigma_{sy}^2 + n_{sY} \sigma_{sx}^2\right)}} \left(\bar{X}_{s.} - \bar{Y}_{s.}\right),$$

where, $\bar{X}_{..} = \Sigma_i \Sigma_{j=1}^{n_{iX}} X_{ij}/n_x$, $\bar{Y}_{..} = \Sigma_i \Sigma_{j=1}^{n_{iY}} Y_{ij}/n_y$, $\bar{X}_{s.} = \Sigma_{j=1}^{n_{sX}} X_{sj}/n_{sX}$, $\bar{Y}_{s.} = \Sigma_{j=1}^{n_{sY}} Y_{sj}/n_{sY}$, and σ_{ax}, and σ_{ay} are the standard deviation of PTTR in the full cohort for the genotype-guided dosing algorithm arm and the clinical-guided dosing algorithm. This correlation can be shown to be $\sqrt{p \cdot \gamma}$ where γ is the ratio of the variances of the PTTR in the enriched subgroup and the full cohort under the null hypothesis and p is the proportion of the subgroup in the population (see Appendix).

The overall type-I error of testing both the full cohort and subgroup is given by:

$$P_{H_0}\left(|Z_a| > z_{1-\alpha_a/2} \quad \text{or} \quad |Z_s| > z_{1-\alpha_s/2}\right),$$

which can be further simplified to:

$$\alpha_a + \alpha_s - P_{H_0}\left(|Z_a| > z_{1-\alpha_a/2} \quad \text{and} \quad |Z_s| > z_{1-\alpha_s/2}\right).$$

Because we know the correlation between two tests under the null hypothesis, we can find α_s for the subgroup given α_a so that this overall type-I error can be controlled at $\alpha = 0.05$.

Table 2 presents α_s for the subgroup given $\alpha_a = 0.04$ and 0.03 and $p = 0.6$ and 0.55. We considered $\gamma = 1$ to 1.5 by increments of 0.1 because previous studies [5] suggest the variability of the primary endpoint for the enriched subgroup may be higher ($\gamma \geq 1$). For example, the observed γ from the Couma-Gen trial [5] was about 1.2. When the standard deviations of the enriched subgroup of patients and the overall samples are the same ($\gamma = 1$), using the correlation can increase α_s noticeably for a fixed α_a. For example when p is 0.6, we can use α_s as high as 0.02 instead of 0.01 when $\alpha_a = 0.04$ and 0.0315 when $\alpha_a = 0.03$ without inflating the overall type-I error. When $\gamma \neq 1$, because the correlation is an increasing function of γ, it increases as the variance of the PTTR in the subgroup increases compared to that of the full cohort samples. Based on this observation, it should be noted that the approach based on equal variance ($\gamma = 1$) is most conservative when the variance in the subgroup is large ($\gamma \geq 1$) because it uses the smallest correlation to produce α_s.

From these calculations, we can see that the correlation of the two tests depends on p and γ which are not available until the trial ends. Therefore, usually the most conservative method (Bonferroni orŠidák's approach [24] for exact control of type-I error) would be used when designing a study to conservatively estimate the power of the subgroup, and the correlation can be incorporated when analyzing the data, which is the method the COAG trial will use. If the correlation is incorporated, the power for the subgroup analysis in the COAG trial will be higher than what is presented in Table 1.

Balanced alpha allocation

There are many aspects of study design that need to be considered in allocating alpha between the total and subgroup comparisons. These will include: the desired treatment difference to be detected for the total and subgroup comparisons; any differential in standard deviations of PTTR expected between the total and subgroups; and differential in power desired for the total and subgroup comparisons. Under the currently assumed parameters, the power of the enriched subgroup comparison is much higher than that of the full cohort comparison as seen in Table 1. This suggests that we may use α_a larger than 0.04 (thus smaller study sample size and smaller α_s for the subgroup), and still maintain the desired powers for both tests. This led us to examine an alpha allocation approach given the current parameters in the COAG trial to achieve the same power for both of the primary analyses. We call this `balanced alpha allocation.' Other considerations or points of view may dictate the decision of how to allocate overall alpha, and the approach described below can be generalized to account for those considerations.

Let Δ_a and Δ_s be the absolute differences in PTTRs between the two arms in the full cohort and subgroup, respectively, and let σ_a and σ_s be the standard deviations of PTTR in the full cohort and the subgroup. Let n and p be the sample size and the proportion of the subgroup in the population. Moreover, denote the dropout rate as d and $\gamma = \sigma_s^2/\sigma_a^2$. Then we have the following two equations for the sample size and power calculations of the full cohort and subgroup tests:

$$\frac{\Delta_a (1-d)}{\sqrt{2\sigma_a^2/n}} = z_{1-\alpha_a/2} + z_{1-\beta}$$

$$\frac{\Delta_s(1-d)}{\sqrt{2\sigma_s^2/(np)}} = z_{1-\alpha_s/2} + z_{1-\beta},$$

where z_a is the (a × 100)-th percentile of the standard normal distribution, and β is the type-II error rate.

By dividing these two equations, we have the following equation:

$$\frac{1}{\sqrt{p}}\frac{\Delta_a}{\Delta_s}\gamma = \frac{z_{1-\alpha_a/2} + z_{1-\beta}}{z_{1-\alpha_s/2} + z_{1-\beta}}.$$

Because α_s is a function of other variables in the equation (α_a, p, and γ), this equation can be solved in terms of α_a by setting β=0.1 (to have 90% power for both tests for example) for fixed values of Δ_a, p, and γ. Then, the corresponding sample size can be obtained.

Detecting smaller than 5.49% PTTR difference was determined by the COAG investigators to be of inconsequential clinical benefit. Therefore, we fixed the overall treatment effect Δ_a as 5.49%. Also to avoid having too small α_s, we considered Δ_s of not only 9.15% but also a slightly smaller value of 8.15%. When α_s is too small, the study sample size will also become small because large alpha will be used for the full cohort analysis. As for p our current best estimate is 0.6, but we also considered p = 0.55 so that the power of the subgroup analysis is protected even if the proportion of the subgroup turns out to be smaller than we expected. Note that no treatment effect for the patients who do not belong to the enriched subgroup is assumed only to conservatively estimate the overall treatment effect. However, in obtaining the balanced alpha, and thus the sample size, possible positive treatment effect for these patients are considered, and this is the reason why for varying p and Δ_s, a fixed value of Δ_a is used.

The results considering γ = 1 to 1.3 with increments of 0.1 are presented in Table 3. These calculations considered the correlation between two tests when obtaining α_s for each α_a. Under the currently assumed parameters (Δ_a = 5.49%, Δ_s = 9.15%, and p = 0.6) the balanced α_a is larger than 0.047 for γ = 1 to 1.3, and even with smaller p = 0.55, α_a is still greater than 0.042. For situations where the treatment effect in the subgroup compared to that in the remaining samples is smaller than what was anticipated (e.g., due to the possible positive treatment effect in the remaining patients), we also considered Δ_s = 8.15%. In this case, α_a ranges from 0.035 to 0.045 for p = 0.6 and from 0.028 to 0.041 for p = 0.55.

Because the correlation is usually not known until the trial ends, we also considered the balanced alpha allocation without considering the correlation so that the most conservative estimate of α_s would be used in designing the study. In this case, α_s becomes smaller for a fixed α_a and this will produce the balanced α_a smaller than that calculated considering the correlation. The results for Δ_a = 5.49% and Δ_s = 9.15 % for p = 0.6 and 0.55 without considering the correlation are presented in Table 4. Under the currently assumed parameters, the balanced α_a is 0.042 when γ = 1.2 (estimate from the Couma-Gen Trial [5]), and if p becomes smaller (0.55), the balanced alpha becomes 0.038. Our choice α_a = 0.04 provides some protection against lower than expected p (Table 4, p = 0.55 and γ as high as 1.1), and also against lower than expected Δ_s compared to Δ_a if we incorporate correlation (Table 3, Δ_s = 8.15%, p = 0.6, and γ = 1.1).

Discussion

In the COAG trial, a prospective alpha allocation approach [12,13,19] was considered to formally include the comparison of the treatment benefit in an enriched subgroup, that is, the subgroup of patients with large differences between the predicted initial doses from the two dosing algorithms. These are the patients who are expected to have greater benefit from genotype-guided dosing.

To ensure balanced sample size of the enriched subgroup between the two arms, stratification based on subgroup membership might be considered [10,19]. In the COAG trial, even with the best effort, the predicted dose difference may not be observed for some patients until after the randomization because of the possible delay in genotyping. This prevents us from considering stratification based on this factor. However, as previously noted, although the subgroup information may not be available when the patient is randomized, whether or not the patient is in the subgroup is a factor fixed at the time of randomization, and thus the subgroup is a proper (baseline) subgroup. For the same reason, the exact α_s considering the correlation could not be specified in advance, as it depends on the proportion of patients in the subgroup and the ratio of the standard deviations between two groups. Therefore, the power calculation of the subgroup was based on the most conservative approach using $\alpha_s = \alpha - \alpha_a$ to ensure adequate power for the subgroup analysis. The correlation will be incorporated when analyzing data at the end of the study to improve power.

Similarly, the balanced alpha allocation depends not only on p and γ but also on the ratio of Δ_a and Δ_s, and again, applying the result from the section `Balanced alpha allocation' when designing a study is difficult. In the COAG trial, 0.04 is used as α_a because the investigators felt that too much uncertainty would be involved in obtaining the balanced alpha. We have shown, however, that the balanced α_a was fluctuating around 0.04 considering slight conservative estimates of p, γ, and the ratio of Δ_a and Δ_s.

An interim analysis might be useful when implementing this approach in practice. For example, pooled estimates of p and γ, and also the ratio of two treatment effects α_a and Δ_s using, for example, the first half of the data can be used to calculate the balanced α_a. Then the study can be adaptively redesigned based on this value, allowing only an increase in sample size. Even though only the ratio of the treatment effect is used, this method is likely to inflate type I error, and further research on adjustment of the type-I error is needed.

When two tests are considered to have great importance as in the COAG trial, there are alternative testing strategies including those based on a closed testing approach. For example, if the importance of the subgroup analysis outweighs that of the full cohort analysis, one may consider testing the subgroup at 0.05 using the full type-I error, and test the full cohort at 0.05 if the subgroup test is positive. The advantage of this method is that the choice of type-I errors for two tests does not depend on many design parameters and can be determined in a straightforward manner. The shortcoming of this approach is that the chance of detecting significant treatment benefit in the full cohort depends on the success of the subgroup analysis. As previously discussed, even though we expect patients in the enriched subgroup to have greater benefit from the genotype-guided dosing, the degree of treatment benefit in the remaining patients is uncertain due to the definition of the subgroup in the COAG trial and also the dose-revision algorithm after the initial doses. Therefore, in the COAG trial, rather than using sequential testing approach, we considered both tests at the same time.

Acknowledgments

The research of the authors (Benjamin French, Stephen E Kimmel, Jonas H Ellenberg) is supported under a contract from the National Heart, Lung and Blood Institute.

Appendix: The correlation between Za and Zs

Rewrite Z_a and Z_s as

$$Z_a = \frac{\sqrt{n_X n_Y}}{\sqrt{\left(n_X \sigma_{ay}^2 + n_Y \sigma_{ax}^2\right)}} \times \left(\frac{\sum_{j=1}^{n_{sX}} X_{sj} + \sum_{j=1}^{n_{s^cX}} X_{s^cj}}{n_X} - \frac{\sum_{j=1}^{n_{sY}} Y_{sj} + \sum_{j=1}^{n_{s^cY}} Y_{s^cj}}{n_Y} \right).$$

$$Z_s = \frac{\sqrt{n_{sX} n_{sY}}}{\sqrt{\left(n_{sX} \sigma_{sy}^2 + n_{sY} \sigma_{sx}^2\right)}} \left(\frac{\sum_{j=1}^{n_{sX}} X_{sj}}{n_{sX}} - \frac{\sum_{j=1}^{n_{sY}} Y_{sj}}{n_{sY}} \right).$$

Then, the correlation between Z_a and Z_s under the null hypothesis ($\sigma_a^2 = \sigma_{ax}^2 = \sigma_{ay}^2$, $\sigma_s^2 = \sigma_{sx}^2 = \sigma_{sy}^2$) can be obtained as

$$\mathrm{Cov}(Z_a, Z_s) = \frac{\sqrt{n_X n_Y}}{\sqrt{(n_X + n_Y)\,\sigma_a^2}} \frac{p\sqrt{n_X n_Y}}{\sqrt{p\,(n_X + n_Y)\,\sigma_s^2}} \times \left(\frac{\sigma_s^2}{n_X} + \frac{\sigma_s^2}{n_Y} \right) = \sqrt{p \frac{\sigma_s^2}{\sigma_a^2}} = \sqrt{p\gamma}.$$

References

1. Ansell JE. Oral anticoagulant therapy – 50 years later. Arch Intern Med. 1993; 153:586–96. [PubMed: 8439222]
2. Higashi MK, Veenstra DL, Kondo LM, et al. Association between *CYP2C9* genetic variants and anticoagulation-related outcomes during warfarin therapy. JAMA. 2002; 287:1690–98. [PubMed: 11926893]
3. Rieder MJ, Reiner AP, Gage BF, et al. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. N Engl J Med. 2005; 352:3385–93.
4. Sconce EA, Khan TI, Wynne HA, et al. The impact of *CYP2C9* and *VKORC1* genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. Blood. 2005; 106:2329–33. [PubMed: 15947090]
5. Anderson JL, Horne BD, Stevens SM, et al. Randomized trial of genotype-guided versus standard warfarin dosing in patients initiating oral anticoagulation. Circulation. 2007; 116:2563–70. [PubMed: 17989110]
6. Caraco Y, Blotnick S, Muszkat M. *CYP2C9* Genotype-guided warfarin prescribing enhances the efficacy and safety of anticoagulation: a prospective randomized controlled study. Clin Pharmacol Ther. 2008; 83:460–70. [PubMed: 17851566]
7. Lenzini PA, Grice GR, Milligan PE, et al. Laboratory and clinical outcomes of pharmacogenetic versus clinical protocols for warfarin intiation in orthopedic patients. J Thromb Haemost. 2008; 6:1655–62. [PubMed: 18662264]

8. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E. A method to determine the optimal intensity of oral anticoagulant therapy. Thromb Haemost. 1993; 69:236–39. [PubMed: 8470047]
9. Verhovsek M, Motlagh B, Crowther MA, et al. Quality of anticoagulation and use of warfarin interacting medications in long-term care: a chart review. BMC Geriatr. 2008; 8:13. [PubMed: 18598364]
10. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. J Clin Oncol. 2005; 23:2020–27. [PubMed: 15774793]
11. Wang SJ. Biomarker as a classifier in pharmacogenomics clinical trials: a tribute to 30th anniversary of PSI. Pharm Stat. 2007; 6:283–96. [PubMed: 17957727]
12. Moyé LA, Deswal A. Trials within trials: confirmatory subgroup analyses in controlled clinical experiments. Control Clin Trials. 2001; 22:605–19. [PubMed: 11738119]
13. Simon R, Wang SJ. Use of genomic signatures in therapeutics development in oncology and other diseases. Pharmacogenomics J. 2006; 6:166–73. [PubMed: 16415922]
14. Gage BF, Eby C, Milligan PE. Use of pharmacogenetics and clinical factors to predict the maintenance dose of warfarin. Thromb Haemost. 2004; 91:87–94. [PubMed: 14691573]
15. Spandorfer JM, Merli GJ. Outpatient anticoagulation issues for the primary care physician. Med Clin North Am. 1996; 80:475–91. [PubMed: 8614182]
16. The International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. N Engl J Med. 2009; 360:753–64. [PubMed: 19228618]
17. Maitournam A, Simon R. On the efficiency of targeted clinical trials. Statist Med. 2005; 24:329–39.
18. Suman VJ, Dueck A, Sargent DJ. Clinical trials of novel and targeted therapies: endpoints, trial design, and analysis. Cancer Invest. 2008; 26:439–44. [PubMed: 18568764]
19. Freidlin B, Simon R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. Clin Cancer Res. 2005; 11:7872–78. [PubMed: 16278411]
20. Hillman MA, Wilke RA, Yale SH. A prospective, randomized pilot trial of model-based warfarin dose initiation using CYP2C9 genotype and clinical data. Clin Med Res. 2005; 3:137–45. [PubMed: 16160068]
21. Hirsh J, Fuster V. Guide to anticoagulant therapy. part 2: oral anticoagulants. American heart association. Circulation. 1994; 89:1469–80. [PubMed: 8124830]
22. Wang SJ, O'Neill RT, Hung HMJ. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. Pharm Stat. 2007; 6:227–44. [PubMed: 17688238]
23. Alosh M, Huque MF. A flexible strategy for testing subgroups and overall population. Stat Med. 2009; 28:3–23. [PubMed: 18985704]
24. Šidák Z. On multivariate normal probabilities of rectangles: their dependence on correlations. Ann Math Statist. 1968; 39:1425–34.

Table 1

Power estimates for the subgroup analysis using $\alpha_s = 0.01$ and 10% dropout rate given the total sample size of 1238. *p* is the proportion of patients in the subgroup, and σ is the common standard deviation of PTTR

p	σ = 20%	σ = 25%	σ = 30%
0.6	99.9%	97.2%	87.8%
0.55	99.7%	95.7%	84.3%

Table 2

The significance level (α_a) for the subgroup analysis given $\alpha_a = 0.04$ and 0.03 and $p = 0.6$ and 0.55 for $\gamma = 1 - 1.5$ considering the null correlation between two tests, α_s^* is the significance level for the subgroup analysis without considering the correlation

$a_a\left(a_s^*\right)$	p	γ	α_s	$a_a\left(a_s^*\right)$	p	γ	α_s
0.04	0.6	1	0.02	0.03	0.6	1	0.0315
(0.01)		1.1	0.0222	(0.02)		1.1	0.0333
		1.2	0.0245			1.2	0.0355
		1.3	0.275			1.3	0.0380
		1.4	0.0313			1.4	0.0409
		1.5	0.0363			1.5	0.0444
	0.55	1	0.0186		0.55	1	0.0300
		1.1	0.0202			1.1	0.0316
		1.2	0.022			1.2	0.0333
		1.3	0.0242			1.3	0.0353
		1.4	0.0269			1.4	0.0375
		1.5	0.0302			1.5	0.0401

Table 3

Balanced α_a for the full cohort analysis and the sample size under various values of Δ_a, Δ_s, and γ. Total n is the total sample size required when the overall standard deviation σ_a is 25%. The correlation between two tests are incorporated when obtaining α_s given α_a

Δ_a	Δ_s	γ	$p = 0.6$		$p = 0.55$	
			Balanced α_a (α_s)	Total n	Balanced α_a (α_s)	Total n
5.49%	9.15%	1	0.0492 (0.0036)	1082	0.0479 (0.006)	1088
		1.1	0.0486 (0.0064)	1086	0.0465 (0.01)	1098
		1.2	0.0478 (0.0104)	1090	0.0446 (0.015)	1110
		1.3	0.0468 (0.0155)	1096	0.0423 (0.0209)	1124
5.49%	8.15%	1	0.0450 (0.0125)	1106	0.0410 (0.0173)	1134
		1.1	0.0423 (0.0188)	1124	0.0369 (0.0242)	1164
		1.2	0.0391 (0.0258)	1146	0.0324 (0.0310)	1299
		1.3	0.0354 (0.033)	1174	0.0277 (0.0374)	1244

Table 4

Balanced α_a for the full cohort analysis and the sample size under various values of p and γ. Total n is the total sample size required when $\alpha_s = 0.05 - \alpha_a$ and the overall standard deviation is 25%

Δ_a	Δ_s	γ	$p = 0.6$		$p = 0.55$	
			Balanced α_a (α_s)	Total n	Balanced α_a (α_s)	Total n
5.49%	9.15%	1	0.0467 (0.0033)	1096	0.0446 (0.0054)	1110
		1.1	0.0444 (0.0056)	1110	0.0415 (0.0085)	1130
		1.2	0.0415 (0.0085)	1130	0.0379 (0.0121)	1156
		1.3	0.0382 (0.0118)	1154	0.0339 (0.0161)	1186