Author's response to reviews

Title: Statistical design of personalized medicine interventions: The Clarification of Optimal Anticoagulation through Genetics (COAG) trial

Authors:

Benjamin French (bcfrench@upenn.edu)
Jungnam Joo (jooj@nhlbi.nih.gov)
Nancy L Geller (gellern@nhlbi.nih.gov)
Stephen E Kimmel (stevek@upenn.edu)
Yves Rosenberg (rosenbey@nhlbi.nih.gov)
Jeffrey L Anderson (jeffrey.anderson@imail.org)
Brian F Gage (bgage@dom.wustl.edu)
Julie A Johnson (johnson@cop.ufl.edu)
Jonas H Ellenberg (jellenbe@upenn.edu)

Version: 5  Date: 11 October 2010

Author's response to reviews: see over
Dear Editors,

Thank you for the opportunity to revise and resubmit our manuscript, “Statistical design of personalized medicine interventions: The Clarification of Optimal Anticoagulation through Genetics (COAG) trial” (MS 9579930874247033). The comments from the reviewers were especially helpful. We carefully considered their comments and accordingly revised the manuscript. Below we include each reviewer comment in italics font, followed by our response in standard font. Please note that all corresponding changes in the manuscript are highlighted. Also note that references embedded herein refer to references in the manuscript.

Reviewer 1

1. This is a well written manuscript on the statistical design issues in a pharmacogenetics trial.
   Thank you.

2. The authors state that they will genotype before first dosage. How will they make that happen? In clinical practice nowadays that will be quite complicated to achieve.

   Given recent technologies, same-day genotyping for warfarin is now possible in practice. In the COAG trial, clinical sites are using one of two genotyping platforms; each has a rapid turnaround time. Both platforms have been FDA approved, have high call and concordance rates, very low failure rates, and the ability to genotype the SNPs needed for the selected dosing algorithms. As discussed in the Methods section, every attempt will be made to determine a subject’s genotype prior to administration of the initial dose. However, for those subjects assigned to the genotype-guided dosing group whose genetic information is not available prior to the initial dose, it will be determined using the clinical dose-initiation algorithm. Once genetic information becomes available, the dose for these subjects will be determined using the genetic dose-initiation and dose-revision algorithms. We have included the following text in the Methods section (also see Comment 7 from Reviewer 2):

   Given recent technologies, same-day genotyping for warfarin is now possible in practice. In the COAG trial, clinical sites are using one of two genotyping platforms; each has a rapid turnaround time. Both platforms have been FDA approved, have high call and concordance rates, very low failure rates, and the ability to genotype the SNPs needed for the selected dosing algorithms.

3. The authors state that the proportion of patients that is included in the trial with the different genotypes needs to be monitored. This is because this will influence the detectable difference. I
wonder whether that would be really necessary given:

- The fact that the percentage of patients with a certain genotype in a population is known and the large sample size leads to a good pre-study prediction of the amount of subjects that will be included;
- The fact that in the analysis phase the analyses will be stratified by genotype.

Maybe the authors can add some explanation why they still think it should be monitored.

We appreciate this suggestion. As discussed in the Minimum Detectable Difference section, the assumed proportion of 0.4 who possess a single genetic variant (in either CYP2C9 or VKORC1), and hence are not expected to benefit from genotype-guided dosing, is based on the Couma-Gen trial [13] and the International Warfarin Pharmacogenetics Consortium [15]. While we believe that the assumed proportion is a reasonable estimate for this population parameter, it is only an estimate for the population from which the Couma-Gen and IWPC samples were obtained. As we discuss, uncertainty in the proportion that possess a single genetic variant implies that the minimum detectable difference used in our sample size calculations is not a known quantity, but rather an unknown quantity that depends on the genetic makeup of the subjects enrolled. In addition, our sensitivity analysis shows that power may be sensitive to the proportion of subjects with zero or multiple genetic variants in whom there is a difference between the predicted doses. Uncertainty in the population prevalence of allelic variants, uncertainty in their impact on warfarin’s efficacy, and the potentially large impact of these quantities on adequate sample size and power motivates monitoring the distribution of allelic variants during the course of the trial. We therefore modified our Conclusions to reflect our rationale (see Comment 3 from Reviewer 2). Note that the primary analysis will not be stratified by genotype, as described in the Methods section.

Reviewer 2

1. This paper describes an interesting trial designed to determine the value of genotype driven warfarin dosing relative to a dosing schedule determined by clinical factors. The objective of the manuscript is not entirely clear. In places it is written as a methods paper attempting to define approaches to trial design for clinical trials of personalized medicine using the COAG trial as the motivating example. More often it is a description of key statistical design features of the COAG trial. Either goal is worthy but the current manuscript seems to miss the mark for either. Given the nature of the trial, I would recommend the authors focus more on a detailed description of the COAG trial, including the motivation and design.

We appreciate this suggestion. The goal of our manuscript is to provide practical guidance on the statistical design of personalized medicine interventions, using the statistical design of the COAG trial as an illustrative example. Although, as the reviewer observed, we describe the general statistical design of the COAG trial, we believe that the statistical design is sufficiently novel to provide insight into the statistical design of a personalized medicine intervention. This insight could not be provided without a thorough description of the key statistical design features of the COAG trial. Therefore, we briefly present relevant scientific background in the Introduction section, and summarize the clinical rationale for the COAG trial in the Methods section. We describe in detail randomization, determining sample size and statistical power, specifying a minimum detectable difference and level of significance, and plans for data
monitoring and final analysis in the COAG trial. Each of these aspects of the COAG statistical design provides guidance for the statistical design of a personalized medicine intervention. In addition, we provide a sensitivity analysis to show that in the COAG trial, sample size and power calculations, which are key aspects to the statistical design of any personalized medicine intervention, may be sensitive to two requisite assumptions: the distribution of relevant allelic variants in the study population; and whether the pharmacogenetic intervention is equally effective across subpopulations defined by allelic variants. We believe that our goals and presentation are consistent with the aim and scope of a Trials methodology article, more so than a Research or Study Protocol paper that focuses on the non-statistical scientific motivation and clinical design. We have therefore clarified the goals of the paper in the Introduction:

The goal of this manuscript is to provide practical guidance on the statistical design of a personalized medicine intervention that uses each subject’s genotype information in an untargeted design. The statistical design of the COAG trial serves as an illustrative example. We briefly summarize the clinical rationale and the general study design for the COAG trial. We use power and sample size calculations to illustrate the primary statistical challenge of designing a personalized therapy intervention: to accommodate the potential differential effectiveness of genotype-guided therapy across subpopulations defined by allelic variation. We provide a sensitivity analysis to quantify the extent to which power and sample size calculations may be sensitive to key assumptions required in the statistical design of a personalized medicine intervention. We conclude with general recommendations for the statistical design of personalized medicine interventions.

2. A thorough discussion of the rationale for selecting this design is needed. In COAG, all eligible patients are randomized to either genotyped-guided or clinically-guided warfarin dosing. While reasonable at first glance, the justification is less obvious when it becomes clear that treatment effect may be null for ~40% of patients, a genotypically-defined subgroup. The authors account for this null effect in the design by acknowledging that this will dilute the overall observed treatment effect and therefore inflate the sample sizes to assure power to detect a much smaller overall difference between groups. At first, I thought the problem was one of effect modification with this subgroup (possibly) experiencing a different treatment effect, a problem many trials must consider. However, in this case, the intervention itself apparently does not differ for this large subset (“subjects who possess a single genetic variant . . . would not benefit from clinical-guided dosing because previous data suggest that the genotype-guided algorithm will predict essentially the same dose as the clinical-guided algorithm” p. 6). Thus it appears not to be true effect modification but rather no effective intervention in this substantial subgroup. A fuller discussion of the pros and cons of this design relative to a targeted one, randomizing only patients for whom genotype guided dosing differs, is needed. The arguments for and against an inclusive versus a targeted design are quite different when one is expecting differential intervention effects across subgroups. The sample sizes (and hence cost) here should be very different. There are no obvious logistical concerns that make an inclusive design necessary. The statement that the inclusive design allows the results to be more generalizable does not seem appropriate; if these design assumptions are correct, the most valid interpretation will be specific to the genetically defined subgroups.

We agree that the pros and cons of an untargeted versus a targeted design deserve discussion.
Although we have powered the study under an assumption of no benefit in a genetic subgroup, we do not know for certain that this subgroup will not derive some benefit from the intervention. The assumption that those who possess a single genetic variant will not benefit from genotype-guided dosing, while suggested by the Couma-Gen trial [13], was not supported in another clinical trial in which all patients benefited from dosing based on CYP2C9 [16]. The doses calculated under the two algorithms are not exactly the same because the day 4 dose-revision algorithm includes the subject’s INR; this may lead to an imbalance overall. Therefore, if we excluded subjects who may not benefit from genotype-guided dosing based on our assumptions, we would be unable to actually evaluate our assumptions in our observed data. In addition, all subjects must be genotyped to know which genetic group they are in, so that much of the cost is incurred in screening.

In the Discussion section, we contrast the untargeted design of the COAG trial with a targeted design that would only enroll subjects anticipated to benefit from genotype-guided dosing. As we discuss, targeted designs, such as those to evaluate a novel cancer therapy, generally exclude subjects who are expected to be unresponsive to therapy based on their genetic characteristics. However, in studies designed to select a drug dose, “unresponsive” subjects are generally eligible, because these subjects would still be responsive to the drug, even though their genetic information may not provide increased accuracy over clinical information to determine their therapeutic dose.

We believe that it is appropriate to consider the generalizability of the COAG trial to the general population. If the COAG trial, or any personalized medicine intervention for that matter, indicates that genotype-guided dosing provides increased efficacy compared to dosing determined without regard to genotypes, then it motivates consideration of the policy question of whether all patients prescribed the drug should be genotyped to predict the drug’s efficacy. Such a policy would be implemented for the entire population of patients.

Given these additional insights raised by the reviewer, we have added the following paragraph to the Discussion section:

For the COAG trial, we favored including all participants, regardless of their genetic variants. First, the assumption that those who possess a single genetic variant will not benefit from genotype-guided dosing, while suggested by the Couma-Gen trial [13], was not supported in another clinical trial in which all patients benefited from dosing based on CYP2C9 [16]. Therefore, if we excluded subjects who may not benefit from genotype-guided dosing, we would be unable to evaluate our assumptions. Second, all subjects are genotyped prior to randomization, so that much of the cost is already incurred in screening. Third, including subjects potentially unresponsive to genotype-guided dosing allows the results of the trial to be more generalizable. That is, if the COAG trial indicates that genotype-guided dosing provides increased efficacy compared to clinical-guided dosing, then it motivates consideration of the policy question of whether all patients prescribed warfarin should be genotyped to predict the drug’s efficacy.

3. *I don’t understand the distinction the authors (are trying to make p. 5) between this “personalized medicine intervention” and a traditional clinical trial. In many traditional trials*
there are covariates thought to be effect modifiers. Usually one does not know in advance the prevalence of these covariates and one must guard against an unfavorable representation of them in the recruited population by either increasing the sample size or imposing some constraints on the distribution. How is the distribution of allelic variants different?

As the reviewer suggests above, the statistical issue we discuss is not simply the presence of effect modification by a fixed and known covariate, but rather that there may be no effective intervention in one or more subpopulations defined by their allelic variants. Consider, for example, gender. It may be well known that an intervention is more effective among males, and the distribution of males in the population is also well known, easily measured, and easily monitored during the study. In contrast, the distribution of allelic variants in the study population and whether the pharmacogenetic intervention is equally effective across subpopulations defined by allelic variants is currently less well known given the novelty of pharmacogenetics. For example, the assumed proportion of 0.4 who possess a single genetic variant in either CYP2C9 or VKORC1, and hence are not expected to benefit from genotype-guided dosing, is only an estimate for the population from which the Couma-Gen and IWPC samples were obtained. In addition, the assumption that those who possess a single genetic variant will not benefit from genotype-guided dosing, while suggested by the Couma-Gen trial, was not supported in another clinical trial in which all patients benefited from dosing based on CYP2C9 [16]. Given this increased uncertainly in the distribution of allelic variants and uncertainty in the effectiveness of the intervention across subpopulations, greater care is required in the design of a personalized medicine to ensure that the study will have adequate power to detect a clinically relevant minimum detectable difference. It was not our intent to make a strong distinction between “traditional” clinical trials and personalized medicine interventions. Rather, our goal is to discuss specific issues that deserve careful attention in the statistical design of a personalized medicine intervention. Therefore, we have modified the paragraph in the Methods section to which the reviewer referred:

A critical element of the statistical design of a randomized clinical trial is to determine a sample size so that a statistical test has adequate power to detect a clinically relevant difference in the primary outcome between treatment groups. The parameters considered in the estimation of sample size include: a minimum detectable difference in the primary outcome between groups; an assumed level of significance for the statistical test of the primary outcome; a measure of variability for the primary outcome in the study population; and the percentage of subjects, if any, expected to drop out of the trial. The sample size parameters for a personalized medicine intervention require additional considerations: the distribution of relevant allelic variants in the study population; and whether the intervention is equally effective across subpopulations defined by allelic variants (e.g., if patients with particular genotypes are not expected to benefit from genotype-guided drug therapy, as we illustrate for warfarin). Due to uncertainly in the distribution of allelic variants and uncertainty in the effectiveness of the intervention across subpopulations, careful attention is required in the design of a personalized medicine to ensure that the study will have adequate power to detect a clinically relevant minimum detectable difference.

We have also modified our Conclusion (also see Comment 3 from Reviewer 1):
In summary, we found that sample size and power calculations may be sensitive to key assumptions required in the design of a personalized medicine intervention: the distribution of relevant allelic variants in the study population; and whether the pharmacogenetic intervention is equally effective across subpopulations defined by allelic variants. Given the novelty of pharmacogenetic research, we recommend that these assumptions be monitored during the conduct of a personalized medicine intervention.

4. I was not able to reproduce these power calculations using some validated calculators. I get somewhat smaller required samples sizes for all of the settings I checked. (E.g., For overall test alpha = 0.04, beta = .2, delta = 5.49, SD = 20, the required sample size is 446 instead of the 550 shown in Table 1), Perhaps I misinterpreted some of the parameters—is the SD for the total sample or within strata (one genetic variant vs 0/multiple)? Were the weights (w_i) assumed to be fixed or random? Please check the calculations.

We appreciate the reviewer for verifying our calculations. We believe that the reviewer neglected to adjust the calculated sample size to accommodate our assumed drop-out rate of 10%. In our calculations we adopted the approach as suggested by Lachin (1981) to adjust the sample size for drop out by dividing by (1 – dropout rate)^2. To clarify, we have edited the second paragraph on page 8 to read:

We also assumed that 10% of subjects would drop out before reaching the primary endpoint and increased the sample size by dividing the calculated sample size by the square of one minus the dropout rate [22].

We have also added the drop-out rate to the first paragraph of the Results section.

5. There are some smaller points that would be worth amplifying in a fuller trial design paper. Probably the most important issue is describing the minimal detectable effect that is worth finding. Specifically, what does a 15% relative difference in PTTR translate into, in terms of clinical endpoints (morbidity, hospital admissions, etc.)

We appreciate this suggestion. There is consensus in the clinical literature that a 15% percent difference in PTTR is clinically relevant. We have added the following reference in the Minimum Detectable Difference section:


6. This is to be a double-blinded trial for the first 4 weeks of the program. How will double-blinding be implemented? Will a dosing schedule that varies from what is clinically recommended cause automatic unblinding?

As discussed in the Methods section, double-blinding will be achieved because neither the study investigators, clinicians, or subjects will be aware of whether the warfarin dose is determined by genotype-guided or clinical-guided dosing. The “dosing schedules” do not differ between the
two treatment groups—both treatment groups will receive their warfarin on a standard-of-care schedule. Rather, how the dose was determined (based on either genotype plus clinical information versus only clinical information) is the only factor that differs between the two groups, as explained in the Methods section. To clarify, we have added the following sentence to the Methods section:

All subjects will receive their warfarin on a standard-of-care schedule.

7. Further, since the authors indicate that there may be cross-overs in the trial (because genetic information may not available at the time of initial dosing) won’t this also introduce both unblinding and further dilution of the observable difference. Are there estimates of how often this is expected to happen?

The availability of genetic information prior to administration of the initial dose will not introduce unblinding. Because the “dosing schedule” does not differ between the two treatment groups, but rather how the dose was calculated, no unblinding will result based on the availability of genetic information.

Yes, the availability of genetic information prior to administration of the initial dose may dilute the observable difference. As discussed in the Methods section, for those subjects assigned to the genotype-guided dosing group whose genetic information is not available prior to the initial dose, the initial dose will be determined using the clinical dose-initiation algorithm. This could result in a further dilution of the observable difference. However, the genotype-guided dose-initiation algorithm on day 1 only uses information on VKORC1 (not CYP2C9) [8]. Therefore, we expect the dose differences on day 1 to be small relative to the dose differences after the first day, and thus any dilution of the observable difference to be small. Note that the genotype-guided dose-initiation algorithm on day 2, as well as the genotype-guided dose-revision algorithm on day 4/5 [9], uses information from both VKORC1 and CYP2C9, so that the availability of genetic information by day 2 will allow the full use of the subject’s genetic information to determine their dose for days 2 through 5. We fully expect genotype information to be available on almost all subjects within 24 hours and certainly by the time of the dose revision calculations on days 4 and 5. We have included the following text in the Methods section (also see Comment 2 from Reviewer 1):

The genotype-guided dose-initiation algorithm on day 1 only uses information on VKORC1 (not CYP2C9) [8]. Therefore, we expect the dose differences on day 1 to be small relative to the dose differences after the first day. The genotype-guided dose-initiation algorithm on day 2, as well as the genotype-guided dose-revision algorithm on days 4 and 5 [9], uses information from both VKORC1 and CYP2C9, so that the availability of genetic information by day 2 will allow the full use of the subject’s genetic information to determine their dose for days 2 through 5. We fully expect genotype information to be available on almost all subjects within 24 hours, and certainly by the time of the dose revision calculations on days 4 and 5.

8. Some description of the dosing schedules for the two arms by strata would also be helpful.
As discussed above, the “dosing schedules” do not differ between the two arms. Readers interested in details regarding the dose-initiation and dose-revision algorithms used to calculate the dose may consult [8] and [9].

Thank you for considering our revision.

Sincerely,

Benjamin French
Department of Biostatistics and Epidemiology
University of Pennsylvania School of Medicine