ORIGINAL ARTICLE

Utility of Adaptive Strategy and Adaptive Design for Biomarker-facilitated Patient Selection in Pharmacogenomic or Pharmacogenetic Clinical Development Program

*Sue-Jane Wang**

In the early to late phases of conventional clinical trials, improvement of disease status at study baseline is the anchor of an effective treatment measured by therapeutic response. These population-based clinical trials do not formally account for disease-associated marker genotype or genome-associated therapeutic response. We discuss alternative study designs in pharmacogenomic or pharmacogenetic clinical trials for genomic or genetic biomarker development, and for formally assessing the clinical utility of genomic or genetic (composite) biomarkers. A two-stage adaptive strategy from completed, ongoing or prospectively planned pharmacogenomic or pharmacogenetic clinical trials is described for development of a genomic or genetic biomarker. We present two types of adaptive design: (1) the genomic biomarker is developed external to the clinical trial, which is designed for treatment effect inference; and (2) first-stage data are used to explore a genomic biomarker, but statistical inference of treatment effect in the genomically or genetically defined biomarker subset is only performed at the second stage of the same trial. When the null hypothesis of no treatment effect in all randomized patients and the genomic patient subset are prospectively specified, we compare the statistical power between fixed and adaptive designs. We also compare the two types of adaptive design. Results from simulation studies showed that adaptive design is more powerful than fixed design for those genomic or genetic biomarkers whose clinical utility is predictive of treatment effect. Pursuit of adaptive design gains at least 20% to more than 30% genomic patient subset power when the genomic biomarker status is readily usable at study initiation, in comparison to when it is explored using the first-stage data of the same clinical trial. In exploratory studies, adaptive strategy provides wide flexibility in the process of genomic or genetic biomarker development. In contrast, an adaptive design trial that employs limited flexibility, and is an adequate and well-controlled investigation, has a greater power gain than a fixed design trial, in which the genomic biomarker is capable of predicting treatment effects that pertain only to the prespecified genomic or genetic patient subset. [*J Formos Med Assoc* 2008;107(12 Suppl): S19–S27]

Key Words: genomic biomarker, patient adaptation, personalized medicine, prospective/retrospective study, study design, treatment effect

Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, US Food and Drug Administration, Maryland, USA.

Received: August 31,2008
Revised: September 9, 2008
Accepted: October 2, 2008

***Correspondence to:** Dr Sue-Jane Wang, Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, US FDA, HFD-700, WO 21, Mail Stop Room 3562, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA.
E-mail: suejane.wang@fda.hhs.gov

Conventional clinical trials are designed around the primary objective and the primary efficacy endpoint of the study. For instance, will the addition of the experimental treatment prolong survival when compared with standard care alone in lung cancer patients? If the study concludes that the new treatment is effective, the average treatment effect, such as median survival in the above example, and its standard error obtained from the clinical trial, are used to provide a two-sided 95% confidence interval for the treatment effect, which is applicable to all randomized patients studied. This is often referred to as a one-size-fits-all approach.

Pharmacogenomic or pharmacogenetic clinical trials aim to identify effective therapeutics which are to be facilitated by the genomic or genetic characteristics of the patient (sub)population under investigation. To achieve this goal, the conventional clinical trial design that addresses only the treatment effect in all randomized patients has been critically challenged. Study designs that allow the clinical trial to address an additional study objective such as new treatment may only be effective in patient subsets characterized by a genomic or genetic profile. In this paper, we use the term genomic collectively to also represent genetic, unless the term genetic is specifically stated. We lay out the utility of an adaptive strategy for genomic biomarker development, and discuss adaptive design in pharmacogenomic clinical trials that build on two study objectives. Is there a treatment effect in all patients studied or is the effect only applicable to a subset of patients who can be characterized by the genomic biomarker[1] or genomic classifier?[2]

When genomic materials or biological specimens are a part of the data collection in clinical trials, exploration of a genomic biomarker may be pursued retrospectively from completed clinical studies or from ongoing or prospectively planned clinical trials. Here, we consider those biomarkers whose presence can influence treatment outcome, and whose status is or should have been known prior to treatment initiation. In the following section on Adaptive Strategy, I present a two-stage adaptive strategy to explore a genomic biomarker. In the section after that on Adaptive Design, I define the composite hypothesis and describe a two-stage adaptive design in pharmacogenomic or pharmacogenetic clinical trials. Adaptive designs are compared with fixed designs when the composite hypothesis is pursued.

Adaptive Strategy

There are many types of study design in pharmacogenomic or pharmacogenetic clinical trials.[3] One distinction among the possible designs is the genomic objective to investigate treatment effect in a genomic patient (sub)population either in an exploratory framework or in a statistical inferential setting. The genomic patient population may be the study population or a subpopulation of a clinical trial. The genomic objective dictates the study design. In addition, the availability of the well-defined genomic biomarker and the acceptability of the diagnostic assay entail the appropriateness of the study design in pharmacogenomic or pharmacogenetic clinical trials.

Figure 1 contrasts the timing of genomic data capture and the usage of genomic (composite) biomarker information in pharmacogenomic or pharmacogenetic clinical trials.

Development of genomic biomarkers

In clinical studies where the genomic objective is exploratory, or clinical development of a genomic (composite) biomarker is in progress, a two-stage adaptive strategy can add to the design flexibility efficiently. This is applicable to retrospective, ongoing, and prospective studies.

As shown in Figure 2, a two-stage adaptive strategy for genomic biomarker development consists of systematically training the genomic data for exploration and discovery in stage 1, and preliminarily validating the clinical utility of the trained genomic biomarker in stage 2.[4,5] There is abundant statistical and bioinformatics literature for stage 1 and 2 development.[2,6,7]

Once a genomic biomarker is developed, approaches to validation such as leave-one-out,

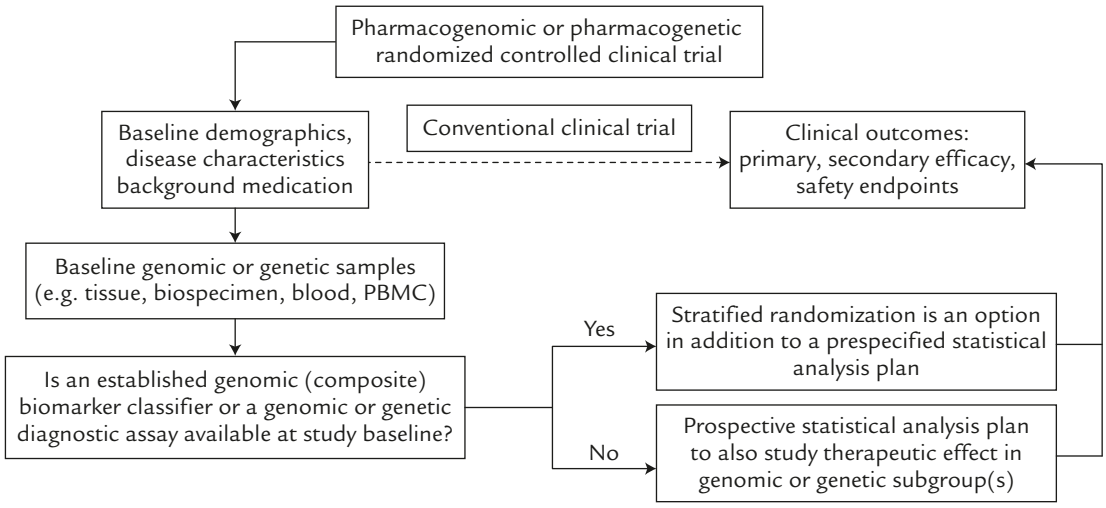
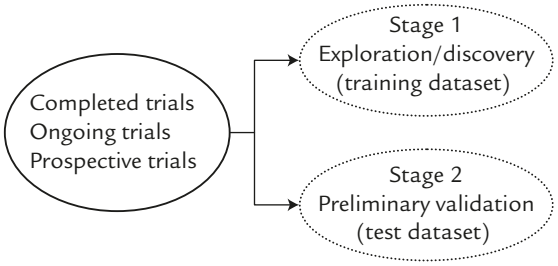


Figure 1. Diagram of timing of genomic data capture and usage of genomic (composite) biomarker information in pharmacogenomic or pharmacogenetic clinical trials. The dashed line represents conventional randomized controlled clinical trial. The solid lines with arrows represent the chronologic timeline of a genomic drug trial. PBMC = peripheral blood mononuclear cells.



- Prespecify the statistical methodologies to systematically explore and develop a genomic or genetic biomarker classifier, e.g. using 60% of accrued patients for stage 1 exploration/discovery—the training set
- Cross-validate the potential classifier effect using the remaining data (to be) collected in stage 2

Figure 2. Development of a genomic (composite) biomarker two-stage adaptive/flexible strategy for exploration.

k-folds cross-validation or bootstrap methods are often used to assess the prediction accuracy. However, prediction error has been shown to be too liberal using cross-validation approaches than using independent validation.[8] Validation performed from a separate prospectively planned study of the same patient population contains rigorous objective criteria for estimation of the prediction error or prediction accuracy. During the development process of a genomic (composite) biomarker, studies that employ a two-stage adaptive strategy are considered exploratory, mostly to generate a genomic hypothesis for later phase study design consideration.

Preliminary clinical utility assessment

When a genomic biomarker is developed using microarray technology or genotyping approach, clear, commercially available diagnostics may be required. Examples are AmpliChip[9] and MammaPrint®.[10] These diagnostics are used to classify patients into subsets. The clinical utility of the genomic biomarker may be prognostic of disease state or drug effect, or predictive of treatment outcome to be tested in adequate and well-controlled pharmacogenomic or pharmacogenetic clinical trials.

For preliminary clinical utility assessment, cross-validation performed in stage 2 provides estimated prediction accuracy, including sensitivity and specificity of the genomic biomarker. The positive and negative predictive values of the diagnostics can also be estimated if the pharmacogenomics trial is a two-arm, placebo-controlled clinical trial designed to explore patient subsets which are to be ruled in for efficacy or ruled out because of serious adverse events. When previously completed or ongoing trial data are used to develop genomic biomarkers, the collected genomic biological specimen may be used to assess the preliminary clinical utility of an experimental diagnostics.

The major statistical issue here is the data quality of the stored genomic samples. Some genomic samples yield unknown genomic biomarker status, which is a type of missing data problem.

A compounding issue is that the genomic objective is not a primary objective of these trials. As a consequence, only optional consent is needed for the exploratory genomic objective. The genomic samples so collected are convenience samples of a double-blind randomized controlled trial. These characteristics can easily be lost for a randomized comparison of a treatment effect within a genomic subset. The clinical event following treatment intervention serves as the clinical truth, and the diagnostic assay using biological specimens to classify patients serves as the test result. Prevalence of positive genomic biomarkers can be estimated using only data from placebo patients.

Adaptive Design

An adaptive design consists of three components: (1) a prospectively planned modification of one or more specified design elements; (2) a modification performed in a specified manner based on an interim analysis of data from subjects in the study; and (3) interim analyses at prespecified time points, performed either fully blinded or unblinded, and with or without a formal statistical hypothesis test.

Genomic biomarker or diagnostic assay available prior to treatment assignment

Depending on the genomic study objective, there are several design options.[3] Ideally, there should be a developed genomic biomarker and an acceptable diagnostic assay available at study initiation to classify patients' biomarker status (Figures 1 and 2). In such cases, a two-stage adaptive designed pharmacogenomic or pharmacogenetic trial can be devised. The basic setup considered in this paper is that, at trial initiation, all patients are randomized, preferably stratifying by their biomarker status to ensure randomization balance within each genomic patient subset. Statistical inference of a treatment effect relative to its comparator is formally assessed in all patients randomized and in the patient subset defined by the presence of the genomic biomarker. We refer to this as the composite hypothesis. It is worthwhile noting that assessment of differential treatment effects between biomarker-positive and -negative patients is only of secondary interest, which is also known as the interaction hypothesis.[3]

The composite hypothesis is specified *a priori*. The adaptive features at the end of stage 1 interim analysis may be to conduct: (1) a futility or safety assessment of the patient subset with absence of the genomic biomarker; (2) an interim predictive power assessment of treatment effect in the patient subset with the genomic biomarker; (3) an increase in total sample size; and (4) an increase in sample size in the to-be-enriched genomic biomarker patient subset. One can also build in a formal interim analysis for early efficacy decision as in adaptive group sequential design. However, we generally discourage such adaptation, especially if an interim analysis is performed early when the sample size is small.

When the composite hypothesis is the primary study objective, a study where no design element is adapted is a fixed design, within which a prespecified multiplicity adjustment method is in place, e.g. Hochberg procedure. Compared with the fixed design, the attractive utility of adaptive design in terms of gaining the study power to detect a treatment effect in the genomic subset is when the genomic biomarker is predictive of treatment effect, *viz.* those solid subset power curves for $1 = (\Delta > 0, \Delta g+ = 0.4, \Delta g- = 0)$ denoted by AD-1 (adaptive design 1) and $2 = (\Delta = 0, \Delta g+ = 0.4, \Delta g- < 0)$ denoted by AD-2 versus those dashed subset power curves for FD-1 (fixed design 1) and FD-2 depicted in Figure 3. With $1 = (\Delta > 0, \Delta g+ = 0.4, \Delta g- = 0)$, the true standardized effect size is 0.4 in the genomic-biomarker-positive subset and no treatment effect in the biomarker-negative subset ($\Delta g- = 0$). Similarly, with $2 = (\Delta = 0, \Delta g+ = 0.4, \Delta g- < 0)$, in the extreme situation, the null overall treatment effect is the result of the treatment effect being offset because it benefits one type of genomically defined patient subset, but is futile to the complementary genomic subset. In contrast, when the genomic biomarker is prognostic-predictive of treatment effect, *viz.* there appears to be differential

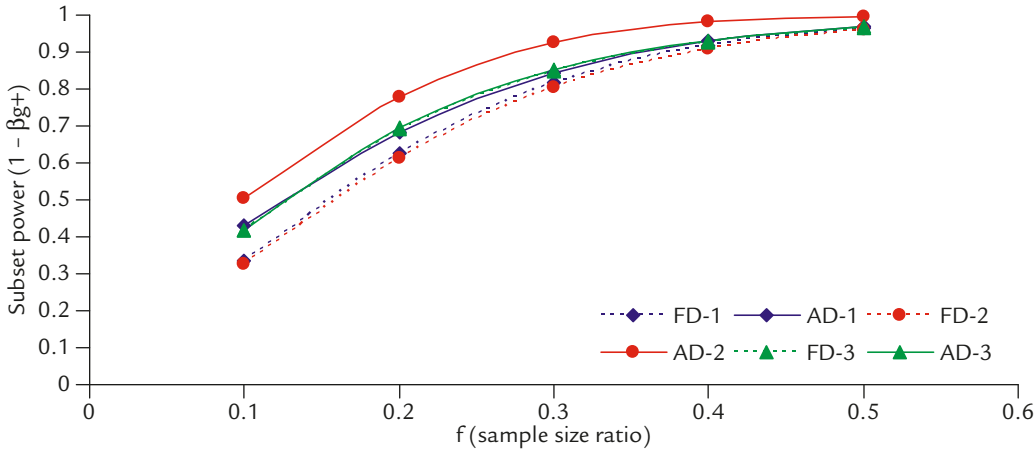


Figure 3. Power comparison for Δg+ with the Hochberg method (1 = (Δ, 0.4, 0); 2 = (0, 0.4, Δg– < 0); 3 = (0.2, 0.4, Δg–)). Δ, Δg+ and Δg– are the standardized effect sizes for all randomized patients, the genomic biomarker positive patients, and the genomic biomarker negative patients. Sample size ratio is the ratio of sample size in genomic biomarker positive patients over all randomized patients. FD = fixed design; AD = adaptive design. 1 and 2 refer to the genomic biomarker that is predictive of treatment effect, 3 refers to prognostic prediction of treatment effect.

treatment effects between biomarker-positive and -negative patient subsets with treatment differences on an order of magnitude. In such cases, there is essentially no genomic subset power gain with the adaptive design approach (see AD-3 versus FD-3 in Figure 3).[11]

Genomic materials collected, but biomarker status not known prior to treatment assignment

When exploration of clinical utility and development of a genomic biomarker is prospectively planned using stage 1 data, there is clearly no established genomic biomarker or acceptable diagnostic assay at study initiation. Let us assume that one is willing to accept that emerging scientific findings external to the clinical trial independently validate the prediction accuracy of a genomic biomarker preliminarily developed within the trial and before its completion. In this case, a two-stage adaptive design may be pursued.[12] The adaptive signature design by Freidlin and Simon[12] uses interim stage 1 data to identify existence of a sensitive genomic patient subset.

Wang et al[13] studied the performance characteristics of a two-stage adaptive design when the genomic biomarker was readily available at study baseline versus when it was only available at the end of stage 1 interim analysis. Using a one-sided equal α-split of 0.0125 for each of the two prospectively specified hypotheses,[12] simulation studies[13] have shown that, when the genomic biomarker is predictive of clinical outcome, an adaptive design with available genomic biomarker status at study baseline yields at least 20% to more than 30% subset power improvement, compared to adaptive design without available genomic biomarker status at trial initiation (Figure 4).[13] This is roughly a 15–30% subset power gain for prognostic utility (Figure 5A in Wang et al,[13] which is reproduced as Figure 5 in this article).

Discussion

In previously completed controlled clinical trials where genomic biological specimens were collected prior to trial initiation, we rationalized the feasibility of genomic biomarker exploration using a two-stage adaptive strategy. The exploration could be driven by the completed studies succeeding in demonstrating a treatment effect. However, redefining a responsive genomic patient subpopulation may become necessary because of the changing definition of disease or syndrome in the clinical community, or de-selecting an unsafe genomic patient subpopulation may become ethically plausible because of irreversible drug-related

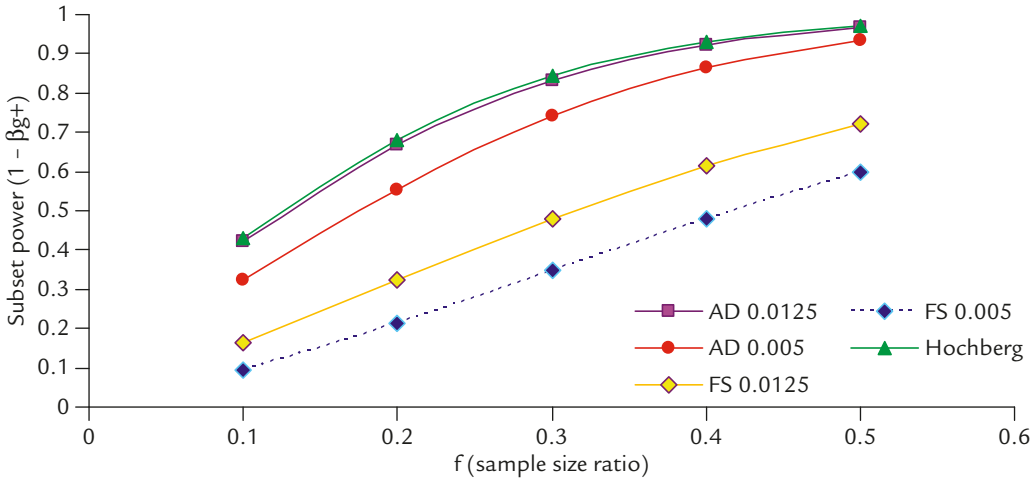


Figure 4. Power comparison for Δg+ under adaptive designs with the standardized effect sizes for all randomized patients, the genomic biomarker positive patients, and the genomic biomarker negative patients: Δ (a function of Δg+, Δg– and the sample size ratio), Δg+=0.4, Δg–=0. Sample size ratio is the ratio of sample size in genomic biomarker positive patients over all randomized patients. AD=adaptive design with alpha allocation by Wang et al;[13] FS=adaptive design by Freidlin and Simon;[12] 0.0125 refers to a one-sided 1.25% type I error rate (equal allocation) for testing H_0: Δg+=0 and H_0: Δ=0 each; 0.005 refers to a one-sided 0.5% type I error rate for testing H_0: Δg+=0 and the remaining one-sided 2% type I error rate for testing H_0: Δ=0; Hochberg refers to testing the two hypotheses following the adaptive design using the Hochberg procedure by Wang et al.[13]

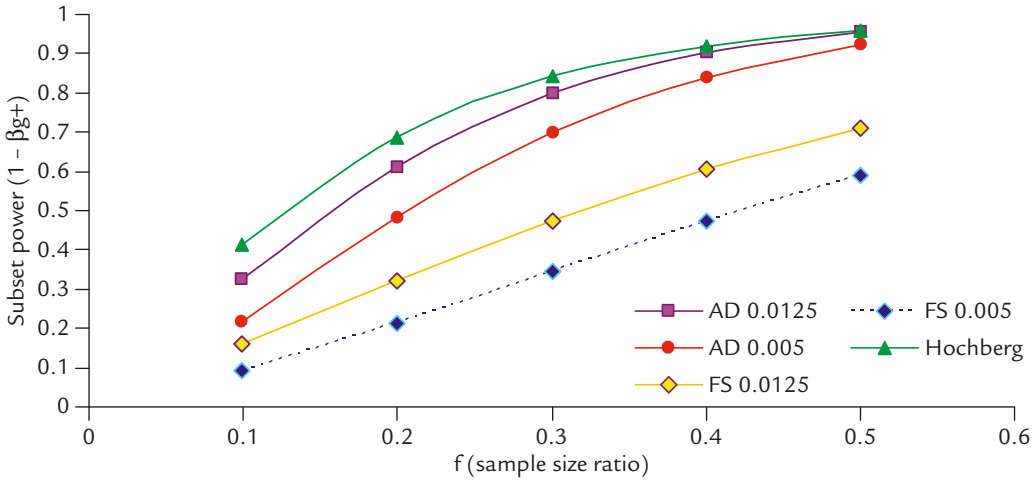


Figure 5. Power comparison for Δg+ under adaptive designs with the standardized effect sizes for all randomized patients, the genomic biomarker positive patients, and the genomic biomarker negative patients: Δ=0.2, Δg+=0.4, Δg– (a function of Δ, Δg+, and the sample size ratio). Sample size ratio is the ratio of sample size in genomic biomarker positive patients over all randomized patients. AD=adaptive design with alpha allocation by Wang et al;[13] FS=adaptive design by Freidlin and Simon;[12] 0.0125 refers to a one-sided 1.25% type I error rate (equal allocation) for testing H_0: Δg+=0 and H_0: Δ=0 each; 0.005 refers to a one-sided 0.5% type I error rate for testing H_0: Δg+=0 and the remaining one-sided 2% type I error rate for testing H_0: Δ=0; Hochberg refers to testing the two hypotheses following the adaptive design using the Hochberg procedure by Wang et al.[13]

adverse events.[14] The exploration could also be driven by studies failing to statistically demonstrate treatment effects based on all the patients studied.

Development of a genomic biomarker is exploratory in its process. The critical issues are not necessarily the multiplicity that occurs in the selection of gene features or prediction models, or the strong control of falsely identifying baseline genomic biomarkers for patient classification. The main clinical question is whether there is a useful clinical utility of the developed genomic biomarker

or an added clinical utility over existing baseline clinical indicators. Therefore, a two-stage adaptive strategy provides flexibility when it comes to investigating several plausible statistical methods, selecting individual genes/single nucleotide polymorphisms and several plausible prediction algorithms and developing a genomic (composite) biomarker. Through an iterative process searching for a genomic biomarker, ultimately, the genomic biomarker that best informs the clinical utility of drug response or treatment effect will be brought forward for genomic inference of a treatment effect in the drug development process.

It has been argued that DNA is highly stable physically and biologically. Thus, DNA genotyping data will be the same regardless of when the DNA samples are acquired, if these samples are properly stored and maintained. Consequently, if the genotyping laboratory is blinded to the clinical outcome data and the treatment assignment, some have argued that designing a prospective pharmacogenomic study to test the clinical composite hypothesis using available unblinded clinical data is possible from previously completed drug trials, assuming that the study sample size is sufficiently powered for the genomic hypothesis. This is known as a prospective/retrospective study design.[14]

The scientific validity of a prospective/retrospective study design depends on the intended clinical utility of the established genomic biomarker and an acceptable diagnostic assay for patient classification.[14] The prospectiveness of the prospective/retrospective study design in terms of prespecified modification of design elements in a two-stage adaptive design trial might be acceptable using previous successfully completed clinical trials for the "new" composite hypothesis, but not in failed trials that no longer have an objective type I error definition and which should be considered hypothesis generation at best.

However, the scientific validity of the prospective/retrospective study design has been a subject of controversy. It is challenging when the intended clinical utility is "predictive of drug effect at group level", and the two-stage adaptive design[13] gives a large power gain over the conventional one-size-fits-all design or fixed design with prespecified multiplicity adjustment of the composite hypothesis. One cannot rule out the possibility of iterative discovery from retrospective sources, such as previously completed clinical trials, being used in hypothesis generation. The multiplicity that results from iterative discovery may be of concern. The validity of the two-stage adaptive design can also be challenged because of the unavailability of a homegrown assay or an unacceptable diagnostic assay that lacks performance characteristics expressed by sensitivity, specificity, accuracy, positive predictive value, negative predictive value, and receiver operating characteristic curve. The fundamental issues are the completeness of ascertaining biological specimens in all randomized patients, whether the specimen quality allows accurate classification of all randomized patients, the amount of missing data on genomic biomarker classification status caused by specimen quality, and the convenience genomic samples due to optional biospecimen collection. Thus, the issue of the level of scientific rigor and the various issues of bias of prospective/retrospective investigations that arise from the design, analysis, conduct and interpretation of a completed clinical trial cannot be ignored.

When the specificity and sensitivity add up to 100% or positive and negative predictive values add up to 100%, it is likely that a genomic patient subset will be randomly selected, which shows no clinical utility of the genomic biomarker or the diagnostic assay. When the diagnostic performances deviate from random classification, there are always trade-offs between sensitivity and specificity, and similarly, between positive and negative predictive values. In general, high sensitivity is favored for early, less expensive, noninvasive tests, and high specificity is favored for later stages when more expensive and invasive tests are performed. Alternatively, predictive value may be used for diagnostic performance. The trade-offs for the predictive value of a genomic biomarker may be more serious for its predictive clinical utility than for its prognostic clinical utility. In the case of MammaPrint®, which has been approved for its prognostic clinical utility, a reasonably high

negative predictive value may be acceptable for detecting no metastatic breast cancer within 5 years. For rare clinical events, a high negative predictive value is easily achievable and the positive predictive value is naturally low, e.g. MammaPrint® was approved on the basis of a high negative predictive value of 0.95 (95% confidence interval of 0.91–0.99), but a very low positive predictive value (0.22, with 95% confidence interval of 0.16–0.28) for metastatic breast cancer within 5 years.[15]

Studies with the genomic objective that a genomic biomarker is predictive of treatment effect should only be prospectively planned and designed, and should not be a secondary attempt from within a previously completed clinical trial. This is because one predicts a future event, and not a future event in the past tense. In addition, the design of the original trials might add exploratory genomic objectives at best when there was a lack of emerging scientific evidence. Thus, the original trial would be unlikely to accommodate the DNA analysis in its formal primary study objective or its key secondary objective, from which these completed studies cannot possibly make themselves prospective for collection of legitimate future events.

Genomic technology[4,5,9,10] promises a revolution in therapeutic discovery and clinical investigation that aims to provide more precise diagnosis and prognosis, a better understanding of drug action, and the ability to better define therapeutic strategies. Pharmacogenomic clinical trials provide the link between target agent and target (sub)population, by way of enriching genomic patient population or adaptively enriching genomic patient subpopulation. To achieve these goals, one can think ahead in early phase drug development to seek an influential genomic biomarker that supports or discovers the drug mechanism of action or identifies drug targets. Adaptive approaches are feasible. The successful treatment of individual patients based on a predictive genomic biomarker that provides added clinical value and clinical utility can be viewed as a substantial first step toward personalized medicine.

We have introduced and discussed two-stage adaptive strategy for genomic biomarker development and two-stage adaptive design to preidentify genomic or genetically prone patients who are expected to respond (better) to therapy, in the sense of treatment risk/benefit balance, than patients without these genomic or genetic characteristics in pharmacogenomic and pharmacogenetic clinical trials.

Acknowledgments

The author thanks Dr Chin-Fu Hsiao of the National Health Research Institutes, Taiwan, for the invitation to make this contribution. This research work was supported by the regulatory science research grants #02-06, #04-06, #05-2, #05-14, #08-48 awarded by the Center for Drug Evaluation and Research, US Food and Drug Administration. The research views expressed in this article are the author's professional views and are not necessarily those of the US Food and Drug Administration.

References

1. Wang SJ. Genomic biomarker derived therapeutic effect in pharmacogenomics clinical trials: a biostatistics view of personalized medicine. *Taiwan Clinical Trials* 2006;4:57–66.
2. Simon R, Wang SJ. Use of genomic signatures in therapeutics development in oncology and other diseases. *Pharmacogenomics J* 2006;6:166–73.
3. Wang SJ. Biomarker as a classifier in pharmacogenomics clinical trials: a tribute to 30th anniversary of PSI. *Pharm Stat* 2007;6:283–96.
4. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817–26.
5. Wang H, Thomas DC, Pe'er I, Stram DO. Optimal two-stage genotyping designs for genome-wide association scans. *Genet Epidemiol* 2006;30:356–68.
6. Inza I, Larranaga P, Blanco R, Cerrolaza AJ. Filter versus wrapper gene selection approaches in DNA microarray domains. *Artif Intell Med* 2004;31:91–103.
7. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005;21:3301–7.
8. Tsai CA, Chen CH, Lee TC, et al. Gene selection for sample classification in microarray experiments. *DNA Cell Biol* 2004;23:607–14.
9. Jain KK. Applications of AmpliChip CYP450. *Mol Diagn* 2005;9:119–27.

10. Glas AM, Floore A, Delahaye LJ, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006;7:278.
11. Wang SJ, Hung HMJ, O'Neill RT. Stage-wise planning for clinical trials from Phase II to Phase III. *Proceedings of the American Statistical Association, Biopharmaceutical Section* [CD-ROM] Alexandria, VA: American Statistical Association, 2007.
12. 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;11:7872–8.
13. 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.
14. Wang SJ, Cohen N, Katz DA, et al. Retrospective validation of genomic biomarkers—what are the questions, challenges and strategies for developing useful relationships to clinical outcomes—workshop summary. *Pharmacogenomics J* 2006;6:82–8.
15. Agendia MammaPrint. *510(k) Substantial Equivalence Determination Decision Summary.* Available from: http://www.fda.gov/cdrh/reviews/K062694.pdf