

Editor's note:

The special column "Statistics in Oncology Clinical Trials" is dedicated to providing state-of-the-art review or perspectives of statistical issues in oncology clinical trials. Our Chairs for the column are Dr. Daniel Sargent and Dr. Qian Shi, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA. The column is expected to convey statistical knowledge which is essential to trial design, conduct, and monitoring for a wide range of researchers in the oncology area. Through illustrations of the basic concepts, discussions of current debates and concerns in the literature, and highlights of evolutionary new developments, we are hoping to engage and strengthen the collaboration between statisticians and oncologists for conducting innovative clinical trials. Please follow the column and enjoy.

Statistics in Oncology Clinical Trials

Biomarker based clinical trial design

Richard Simon

Biometric Research Branch, National Cancer Institute, Bethesda, MD, USA

Corresponding to: Richard Simon, Chief, Biometric Research Branch, National Cancer Institute, 9609 Medical Center Dr, Rockville, MD 20892-9735, USA. Email: rsimon@nih.gov.

Abstract: The established molecular heterogeneity of human cancers and the subsequent stratification of conventional diagnostic categories require the development of new paradigms for the development of a reliable basis for predictive medicine. We review clinical trial designs for the development of new therapeutics and predictive biomarkers to inform their use. We cover designs for a wide range of settings. At one extreme is the development of a new drug with a single biomarker and strong biological evidence that marker negative patients are unlikely to benefit from the new drug. At the other extreme are phase III clinical trials involving both genome-wide discovery and internal validation of a predictive classifier that identifies the patients most likely and unlikely to benefit from the new drug.

Keywords: Predictive biomarker; clinical trial design; adaptive design; companion diagnostic; enrichment trial

Submitted Jan 07, 2014. Accepted for publication Feb 27, 2014.

doi: 10.3978/j.issn.2304-3865.2014.02.03

View this article at: <http://dx.doi.org/10.3978/j.issn.2304-3865.2014.02.03>

Introduction

Randomized clinical trials have been key for the development of a reliable evidence based medicine. Randomized trials generally evaluate a treatment relative to a control regimen for a broadly defined population of patients traditionally defined based on primary site, histologic diagnosis, stage and number of prior treatments. One limitation of randomized clinical trials is that they have also led to the over-treatment of broad populations of patients, most of whom don't benefit from the drugs and procedures shown to have statistically significant average treatment effects.

Tumors of a primary site in many cases represent a heterogeneous collection of diseases that differ with regard to the mutations that cause them and drive their invasion. The heterogeneous nature of tumors of the same primary site offers new challenges for drug development and clinical trial design. Physicians have always known that cancers of the same primary site were heterogeneous with regard to natural history and response to treatment. Today we have better tools for characterizing the tumors biologically and using this characterization in the design and analysis of clinical trials that utilize this information prospectively.

Presently, most oncology drugs are being developed

for defined molecular targets. In some cases the targets are well understood and there is a compelling biological basis for restricting development to the subset of patients whose tumors are characterized by deregulation of the drug target. For other drugs there are multiple targets and more uncertainty about how to measure whether a drug target is driving tumor invasion in an individual patient (1). It is clear that the primary analysis of the new generation of oncology clinical trials must consist of more than just treating broad patient populations and testing the null hypothesis of no average effect. But it is also clear that the tradition of post-hoc data dredging subset analysis is not an adequate basis for predictive oncology. For establishing practice standards and for drug approvals we need prospective analysis plans that provide for both preservation of the type I experiment-wise error rate and for focused predictive analyses that can be used to reliably select patients in clinical practice for use of the new regimen (2-4). The type I experiment-wise error rate is the probability of making any false positive claim (for the overall population or any subset) based on the analysis of the clinical trial. These two primary objectives involve co-development of a drug and a companion diagnostic.

The ideal approach to co-development of a drug and companion diagnostic involves (I) identification of a predictive biomarker based on understanding the mechanism of action of the drug and the role of the drug target in the pathophysiology of the disease. A predictive biomarker is a biological measurement that indicates whether the patient is likely to respond to the particular drug. It is distinguished from a prognostic biomarker which may indicate the pace of progression of the underlying disease. This biological understanding should be validated and refined by pre-clinical studies and early phase clinical trials. The predictive biomarkers for successful cancer drugs have generally involved a single gene or protein rather than a multivariate classifier. Multivariate classifiers have found use as prognostic indicators that reflect a combination of the pace of the disease and the effect of standard therapy (5). They can identify which patients have such good prognosis with conservative management that they do not require more aggressive treatment. Multivariate classifiers have rarely been used as predictive biomarkers for response to specific drugs, however, because their use often reflects an incomplete understanding of the mechanism of action of the drug or the role of its molecular target; (II) development of an analytically validated test for measurement of the relevant biomarker. Analytically validated implies that the test accurately measures what it is supposed to measure,

or if there is no gold-standard measurement, that the test is reproducible and robust; (III) use of the defined test to design and analyze a new clinical trial to evaluate the effectiveness of the investigative drug and how the effectiveness relates to the biomarker value.

Phase II trials

Candidate predictive biomarkers are often evaluated in traditional phase II trials for patients with tumors of a single primary site. Puzstai and Hess (6) and Jones and Holmgren (7) have described extensions of Simon's two-stage single arm phase II design to accommodate a single binary candidate marker. These designs are focused primarily on ensuring that promising activity of the drug is not missed in cases where its activity is restricted to test-positive patients, and yet excessive numbers of patients are not required in cases where its activity is sufficiently broad that the marker is not needed. Freidlin *et al.* (8) have described a design for use with a single binary biomarker in a randomized phase II design that enables one to determine whether the drug should be developed in a phase III enrichment trial, an all-comers trial, or dropped from further development.

There are many more complicated phase II settings, where no natural cut-point of the biomarker is known in advance, or where there are multiple candidate biomarkers. The BATTLE I trial in NSCLC is an example of a phase II clinical trial in which four different tests were evaluated in the context of four different drug regimens (9). Treatment assignment among the four regimens was randomized, but the randomization weights varied as the trial went along according to which treatment had the best performance within each of the four biomarker strata using freedom from progressive disease at week 8 as the endpoint. There were two main objectives of the adaptive randomization. One was to efficiently screen four treatments in four pre-determined strata of NSCLC patients. The second objective was to provide patients with a trial in which they could feel that the design was adapting to assign them the drug regimen that was best for their form of the disease. Korn and Freidlin (10) have raised questions about the effectiveness of such response adaptive randomization designs for reducing the number of patients receiving what turns out to be a less active regimen and Simon (2) has raised questions about how efficient this design is relative to use of optimal two-stage designs for each drug-stratum combination. The I-SPY 2 phase II design being conducted in breast cancer

also uses an adaptive design with pre-specified biomarker strata and multiple treatments (11).

Phase IIa basket discovery trials

Large tumor sequencing studies (12) like the Cancer Genome Project in the UK and The Cancer Genome Atlas (TCGA) in the US have identified recurrent genomic changes in a variety of primary tumor sites. These data provide a scientific basis for treatment of individual patients based on the biological characterization of their tumors. There are, however, many challenges in moving tumor genomics to clinical oncology. These include challenges of logistics, ethics, bioinformatics, study design, regulatory, analytical assay validation and interdisciplinary collaboration. Moving genomics to therapeutics involves using drugs for new indications and dealing with uncertainties about which mutations in a given gene effect the function of the protein product, which are important for the invasive properties of the tumor and which should be considered “actionable” for administration of a drug that was developed for somewhat different mutations in a different primary site. There is much yet to learn about effective matching of drugs to genomically characterized tumors (13). Treating patients with drugs selected based on current knowledge to block the de-regulation caused by genomic alterations can, however, provide a database for improving our knowledge of how to combine tumor genomics with therapeutics. It may be much less informative to treat patients without prospective biological characterization and hope to correlate responses to post-hoc assessed genomic tumor alterations although the latter approach may be useful for trying to understand unusually good responses to standard treatments.

“Umbrella” discovery trials include patients with advanced cancer of multiple primary disease sites which are resistant to standard treatment (14). The patients have their tumor DNA sequenced and it is determined (based on a pre-specified algorithm) whether an actionable mutation is present. Actionable means that a drug is available whose range of molecular targets ‘mesh’ with the genomic alterations of the tumor in a way that suggest treatment may result in benefit for that patient. The evidence that a drug is actionable for a given mutation varies and is often based on biological or pre-clinical data or on data in a different tumor type. The rules of actionability should be prospectively defined. Basket trials have only a single drug available and attempt to discover the types of patients for

whom the drug should be developed in later phase studies. In other cases, multiple drugs are available. In some cases the trial is randomized in which outcome on drugs matched based on actionability rules are compared to outcome on drugs selected based on physicians choice without genomic characterization data. Other trials do not use a control arm.

The randomized discovery designs address two distinct questions (14). One is the testing of the null hypothesis that the policy of trying to match the drug to the genomics of the tumor is no more effective than a physicians’ choice strategy without using any tumor characterization beyond that used for standard of care. Whereas most clinical trials evaluate a single drug or regimen, the null hypothesis for multi-drug umbrella trials relates to a matching policy for a given set of drugs and biomarkers available for the study. This makes it particularly important to obtain a broad enough menu of potent inhibitors of their targets. The policy is also determined by the type of genomic characterization performed and by the “rules” for matching drug to tumor. If the matching is done by a tumor board and is not rule-based or if the rules change frequently, the pragmatic value of the clinical trial will be limited. It may also be difficult for regulatory bodies to approve use of investigational drugs for use as decided by a tumor board rather than in a more rule-based manner. Consequently, it is important that the policy of treatment-assignment by genomic characterization be transparent and that the duration of the trial be short so that the rules do not change frequently. The use of a randomized control group ensures that comparisons of progression free survival (PFS) between the matched group and the control group are not biased by differences in patient characteristics or biases in assessment of progression. The proof-of-principle embodied by the null hypothesis may be more meaningful, however, in a multi-drug trial of a single histologic category than in cases where a wide range of primary sites of disease are included.

A second objective of the randomized studies is the screening of individual drugs used in specific tumor contexts. For some primary sites a gene may be mutated sufficiently frequently for the study to provide an adequate phase II evaluation of the drug for that new indication (13). In many cases, however, the available patient numbers will not be adequate for a proper phase II evaluation. Nevertheless, the trial may serve to screen for drug-mutation matches for which there is a substantial degrees of activity. These leads must be confirmed in an expanded cohort of a follow-up trial (13). In this discovery mode, assessment of activity of a drug against tumors with a given gene mutated must

take into account the possibility that the primary site may indicate a genomic context which may modulate activity of the drug against the alteration.

The non-randomized trials are sometimes called “N of 1” trials in the sense that each patient is different and the outcome of treatment must be evaluated individually in terms of the individual characterization of his or her tumor. This nomenclature can be misleading, however. The “N of 1” approach traditionally referred to a design in which individual patients were treated sequentially for multiple courses with either a test drug or control, with the sequence of treatment or control determined by randomization. This is clearly not possible for cancer studies however. The only endpoint clearly interpretable for non-randomized studies is objective tumor response. Tumors generally do not shrink spontaneously, and so an objective tumor response can usually be attributed to the effect of the drug. Durable objective responses for patients with far advanced metastatic disease are generally rare and can be used for discovering promising ways to target molecularly characterized tumors. PFS is much less interpretable in non-randomized studies. The pace of disease can vary substantially even in advanced cases and so comparing PFS between different subsets of patients is hazardous. PFS is subject to measurement error and ascertainment bias depending on the frequency of surveillance. For a patient who has a PFS prior to entry on study of eight weeks, a PFS ratio (relative to the PFS on the previous treatment) in excess of 1.3 may only mean that progression was not declared at the first eight week follow-up of the genomic based study. This is not strong evidence of an effective treatment effect.

Phase III targeted (enrichment) designs

Designs in which eligibility is restricted to those patients considered most likely to benefit from the experimental drug are called “targeted designs” or “enrichment designs.” With an enrichment design, an analytically validated diagnostic test is used to restrict eligibility for a randomized clinical trial comparing a regimen containing a new drug to a control regimen. This approach has now been used for pivotal trials of many drugs whose molecular targets were well understood in the context of the disease. Prominent examples include trastuzumab (15), vemurafenib (16), and crizotinib (17).

Several authors have studied the efficiency of the ‘targeted’ approach relative to the standard approach of randomizing all patients without using the biomarker test

at all (18-22). The efficiency of the enrichment design depends on the prevalence of test positive patients and on the effectiveness of the new treatment in test negative patients. When fewer than half of the patients are test positive and the new treatment is relatively ineffective in test negative patients, the number of randomized patients required for an enrichment design is dramatically smaller than the number of randomized patients required for a standard design. For example, if the treatment is completely ineffective in test negative patients, then the ratio of number of patients required for randomization in the enrichment design relative to the number required for the standard design is approximately $1/\gamma^2$ where γ denotes the proportion of patients who are test positive (2). The treatment may have some effectiveness for test negative patients either because the assay is imperfect for measuring deregulation of the putative molecular target or because the drug has off-target anti-tumor effects. Even if the new treatment is half as effective in test negative patients as in test positive patients, however, the randomization ratio is approximately $4/(\gamma+1)^2$. This equals about 2.56 when $\gamma=0.25$, i.e., 25% of the patients are test positive, indicating that the enrichment design reduces the number of required patients to randomize by a factor of 2.56.

The enrichment design was very effective for the development of trastuzumab even though the test was imperfect and has subsequently been improved. Simon and Maitournam (18-20) also compared the enrichment design to the standard design with regard to the number of screened patients. The methods of sample size planning for the design of enrichment trials available on line at <http://brb.nci.nih.gov>; the web-based programs are available for binary and survival/disease-free survival endpoints. The planning takes into account the performance characteristics of the tests and specificity of the treatment effects. The programs provide comparisons to standard non-enrichment designs based on the number of randomized patients required and the number of patients needed for screening to obtain the required number of randomized patients.

The enrichment design is appropriate for contexts where there is a strong biological basis for believing that test negative patients will not benefit from the new drug. In such cases, including test negative patients may raise ethical concerns and may confuse the interpretation of the clinical trial.

Phase III biomarker stratified design

When a predictive classifier has been developed but there

is not compelling biological or phase II data that test negative patients do not benefit from the new treatment, it is generally best to include both classifier positive and classifier negative in the phase III clinical trials comparing the new treatment to the control regimen. In this case it is essential that an analysis plan be pre-defined in the protocol for how the predictive classifier will be used in the analysis. The analysis plan will generally define the testing strategy for evaluating the new treatment in the test positive patients, the test negative patients and overall. The testing strategy must preserve the overall type I error of the trial and the trial must be sized to provide adequate statistical power for these tests. It is not sufficient to just stratify, i.e. balance, the randomization with regard to the classifier without specifying a complete analysis plan. The main value of “stratifying” (i.e., balancing) the randomization is that it assures that only patients with adequate test results will enter the trial. Pre-stratification of the randomization is not necessary for the validity of inferences to be made about treatment effects within the test positive or test negative subsets. If an analytically validated test is not available at the start of the trial but will be available by the time of analysis, then it may be preferable not to pre-stratify the randomization process. Several primary analysis plans have been described (23-25) and a web based tool for sample size planning for some of these analysis plans is available at <http://brb.nci.nih.gov>.

If one has moderate strength evidence that the treatment, if effective at all, is likely to be more effective in the test positive cases, one might first compare treatment versus control in test positive patients using a threshold of significance of 5%. Only if the treatment versus control comparison is significant at the 5% level in test positive patients, will the new treatment be compared to the control among test negative patients, again using a threshold of statistical significance of 5%. This sequential approach controls the overall type I error at 5%. To have 90% power in the test positive patients for detecting a 50% reduction in hazard for the new treatment versus control at a two-sided 5% significance level requires about 88 events of test positive patients. If at the time of analysis the event rates in the test positive and test negative strata are about equal, then when there are 88 events in the test positive patients, there will be about $88(1-\gamma)/\gamma$ events in the test negative patients, where γ denotes the proportion of test positive patients. If 25% of the patients are test positive, then there will be approximately 264 events in test negative patients. This will provide approximately 90% power for detecting a

33% reduction in hazard at a two-sided significance level of 5%. In this case, the trial will not be delayed compared to the enrichment design, but a large number of test negative patients will be randomized, treated and followed on the study rather than excluded as for the enrichment design.

In the situation where one has more limited confidence in the predictive marker, the marker can still be effectively used for a “fall-back” analysis. Simon and Wang (25) proposed an analysis plan in which the new treatment group is first compared to the control group overall. If that difference is not significant at a reduced significance level (such as 0.03), then the new treatment is compared to the control group just for test positive patients. The latter comparison uses a threshold of significance of 0.02, or whatever portion of the traditional 0.05 not used by the initial test. Wang *et al.* have shown that the power of this approach can be improved by taking into account the correlation between the overall significance test and the significance test comparing treatment groups in the subset of test positive patients (26). So if, for example a significance threshold of 0.03 has been used for the overall test, the significance threshold for used for the subset can be somewhat greater than 0.02 and still have the overall chance of a false positive claim of any type limited to 5%. Real world experience with stratification and enrichment designs are described by Freidlin *et al.* (27) and by Mandrekar and Sargent (28).

Karuri and Simon (29) introduced a phase III design for the setting of a single binary biomarker stratification design in which futility monitoring of the test negative patients is performed based on a joint prior joint distribution for the treatment effects in test negative and test positive patients. The prior distribution enables the trialist to represent the prior evidence that the treatment effect will be reduced for test negative patients and use that information in monitoring the clinical trial. Although the formulation is Bayesian, the rejection region based on posterior probability is calibrated so that type I errors satisfy the usual frequentist requirements. The Karuri and Simon approach to interim monitoring permits earlier termination of accrual of marker negative patients than with traditional futility analysis methods.

Hong and Simon developed a run-in design which permits a pharmacodynamic, immunologic, or intermediate response endpoint measured after a short run-in period on the new treatment to be used as the predictive biomarker (30). Simon *et al.* (31) described a prospective-retrospective approach to using archived tumor specimens for a focused re-analysis

of a randomized phase III trial with regard to a predictive biomarker. The approach requires that archived specimens be available on most patients, and that an analysis plan focused on a single marker be developed prior to performing the blinded assays. This approach was used in establishing that a K-RAS mutation was a negative predictive biomarker for response of colorectal cancer patients to anti-EGFR antibodies.

Phase III adaptive

Jiang *et al.* (32) reported on a “Biomarker Adaptive Threshold Design” for situations where a biomarker is available at the start of the trial, but a cut-point for converting the value to a binary classifier is not established. Tumor specimens are collected from all patients at entry, but the value of the biomarker is not used as an eligibility criteria. The analysis plan does not stipulate that the assay for measuring the index needs to be performed in real time. At the final analysis Jiang *et al.* (32) determine the optimal threshold for the biomarker; that is, the threshold that identifies the subset of patients for whom the treatment effect is maximum, using a pre-specified metric. The null distribution of the treatment effect in the optimally selected subset was determined by repeating the analysis after permuting the treatment and control labels a thousand or more times. This permutation analysis automatically adjusted for the fact that a full range of thresholds were evaluated and automatically adjusts for the correlation of the treatment effects among nested subsets. Jiang *et al.* also described a method of obtaining confidence intervals for the optimal threshold using bootstrap re-sampling. Since the treatment is presumed effective only for patients with biomarker above the threshold, the confidence coefficient associated with a given biomarker value x can be interpreted as the probability that a patient with marker value x benefits from the new treatment.

The adaptive threshold design described above enables one to conduct the phase III clinical trial without pre-specifying the cut-point for the biomarker. It provides for a valid statistical significance test that has good statistical power against alternative hypotheses that the treatment effect is limited to patients with biomarker values above some unknown level, and it provides a confidence interval for estimation of the cut-point. These analyses are, however, performed at the end of the trial and accrual during the trial is not restricted by biomarker value. Several authors have studied adaptive enrichment designs in which eligibility

criteria change adaptively during the clinical trial based on interim outcome results. Wang *et al.* (33), Rosenblum and Van der Laan (34), and Karuri and Simon (29) consider the case of two strata, e.g., a biomarker positive stratum and a biomarker negative stratum, and adaptively determine when to terminate accrual in the biomarker negative stratum. Follmann (35) considers the case where there are multiple disjoint strata in the population of initially eligible patients and one can adaptively drop each stratum from accrual. Wang *et al.* (33) and Simon and Simon (36), studied more general models for eligibility modification based on multiple candidate biomarkers. The Simon and Simon (36) model was very general and developed statistical significance tests which remain valid even if outcome distributions change during the trial in a manner that depends on the eligibility modifications. Such tests are very robust for use in phase III clinical trials. Simon and Simon (36) illustrated this framework in the setting of adaptive threshold enrichment of a single biomarker.

Designs such as the “adaptive signature design” have been developed for adaptive multivariate classifier development and internal validation based on high dimensional genomic tumor characterization (37). This design employs a “learn and confirm” structure in which a portion of the patients are used to select the biomarker hypothesis, i.e., to develop an “indication classifier” which identifies the target population of patients in which the test treatment is most likely to be effective, and to use the remainder of the patients to test the treatment effect in that subset. The adaptive signature design does not modify eligibility criteria. It is adaptive in the sense that the treatment effect is tested in a single subset determined based on the clinical trial data but in a manner that separates classifier development from testing of treatment effect. Since the adaptive signature design does not use the patients on which the classifier was developed for the testing of the treatment effect, it thus avoids the inflation of type I error described by Wang *et al.* (38) for other approaches. Scher *et al.* described the use of the adaptive signature design for planning a pivotal trial in advanced prostate cancer (39). The key principle of the adaptive signature approach is to replace multiple significance testing based subset analysis with development and internal validation of a single “indication classifier” that informs treatment selection for individual patients based on their entire vector of covariate values.

The adaptive signature design approach is very general with regard to the methodology applied to the training set for identifying the single candidate subset in which

treatment effect will be tested in the validation set. Many methods of predictive classifier development can be developed using the training set. It is important to recognize, however, that one is not developing a prognostic classifier. The classifier is used to classify patients as likely to benefit from the new treatment. Matsui *et al.* (40) used their model to predict a continuous score reflecting the expected benefit for the new treatment relative to the control rather than just classifying patients into one of two subsets. Gu *et al.* (41) have developed a two-step strategy for developing a model for predicting outcome as a function of treatment and selected biomarkers. The biomarkers are selected using a group lasso approach in which the main effects of a biomarker are grouped with the interactions of that marker with treatments and can be used with two or more treatments.

Freidlin *et al.* (42) described further extensions of the adaptive signature approach. They use cross-validation to replace simple splitting of the trial into a training set and test set in order to increase the statistical power.

Conclusions

Recognition of the molecular heterogeneity of human diseases such as cancers of a primary site and the tools for characterizing this heterogeneity presents new opportunities for the development of more effective treatments and challenges for the design and analysis of clinical trials. In oncology, treatment of broad populations with regimens that do not benefit most patients is less economically sustainable with expensive molecularly targeted therapeutics and less likely to be successful. The established molecular heterogeneity of human diseases requires the development of new approaches to use randomized clinical trials to provide a reliable basis predictive medicine. This paper has attempted to review here some prospective designs for the co-development of new therapeutics with companion diagnostics.

Acknowledgements

Disclosure: The author declares no conflict of interest.

References

1. Sawyers CL. The cancer biomarker problem. *Nature* 2008;452:548-52.
2. Simon RM, eds. *Genomic clinical trials and predictive medicine*. Cambridge, UK: Cambridge University Press, 2013.
3. Simon RM. An agenda for clinical trials: clinical trials in the genomic era. *Clin Trials* 2004;1:468-70.
4. Simon R. New challenges for 21st century clinical trials. *Clin Trials* 2007;4:167-9; discussion 173-7.
5. Mamounas EP, Tang G, Fisher B, et al. Association between the 21-gene recurrence score assay and risk of locoregional recurrence in node-negative, estrogen receptor-positive breast cancer: results from NSABP B-14 and NSABP B-20. *J Clin Oncol* 2010;28:1677-83.
6. Puzstai L, Hess KR. Clinical trial design for microarray predictive marker discovery and assessment. *Ann Oncol* 2004;15:1731-7.
7. Jones CL, Holmgren E. An adaptive Simon Two-Stage Design for Phase 2 studies of targeted therapies. *Contemp Clin Trials* 2007;28:654-61.
8. Freidlin B, McShane LM, Polley MY, et al. Randomized phase II trial designs with biomarkers. *J Clin Oncol* 2012;30:3304-9.
9. Kim ES, Herbst RS, Wistuba II, et al. The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov* 2011;1:44-53.
10. Korn EL, Freidlin B. Outcome--adaptive randomization: is it useful? *J Clin Oncol* 2011;29:771-6.
11. Barker AD, Sigman CC, Kelloff GJ, et al. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther* 2009;86:97-100.
12. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009;458:719-24.
13. Simon R, Roychowdhury S. Implementing personalized cancer genomics in clinical trials. *Nat Rev Drug Discov* 2013;12:358-69.
14. Simon R, Polley E. Clinical trials for precision oncology using next generation sequencing. *Personalized Medicine* 2013;10:485-95.
15. Shak S. Overview of the trastuzumab (Herceptin) anti-HER2 monoclonal antibody clinical program in HER2-overexpressing metastatic breast cancer. Herceptin Multinational Investigator Study Group. *Semin Oncol* 1999;26:71-7.
16. Chapman PB, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* 2011;364:2507-16.
17. Shaw AT, Yeap BY, Solomon BJ, et al. Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: a retrospective analysis. *Lancet Oncol* 2011;12:1004-12.

18. Simon R, Maitournam A. Evaluating the efficiency of targeted designs for randomized clinical trials. *Clin Cancer Res* 2004;10:6759-63.
19. Simon R, Maitournam A. Evaluating the efficiency of targeted designs for randomized clinical trials: supplement and correction. *Clin Cancer Res* 2006;12:3229.
20. Maitournam A, Simon R. On the efficiency of targeted clinical trials. *Stat Med* 2005;24:329-39.
21. Hoering A, Leblanc M, Crowley JJ. Randomized phase III clinical trial designs for targeted agents. *Clin Cancer Res* 2008;14:4358-67.
22. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *J Clin Oncol* 2009;27:4027-34.
23. Simon R. The use of genomics in clinical trial design. *Clin Cancer Res* 2008;14:5984-93.
24. Freidlin B, Sun Z, Gray R, et al. Phase III clinical trials that integrate treatment and biomarker evaluation. *J Clin Oncol* 2013;31:3158-61.
25. Simon R, Wang SJ. Use of genomic signatures in therapeutics development in oncology and other diseases. *Pharmacogenomics J* 2006;6:166-73.
26. Wang SJ, O'Neill RT, Hung HM. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharm Stat* 2007;6:227-44.
27. Freidlin B, McShane LM, Korn EL. Randomized clinical trials with biomarkers: design issues. *J Natl Cancer Inst* 2010;102:152-60.
28. Mandrekar SJ, Sargent DJ. Predictive biomarker validation in practice: lessons from real trials. *Clin Trials* 2010;7:567-73.
29. Karuri SW, Simon R. A two-stage Bayesian design for co-development of new drugs and companion diagnostics. *Stat Med* 2012;31:901-14.
30. Hong F, Simon R. Run-in phase III trial design with pharmacodynamics predictive biomarkers. *J Natl Cancer Inst* 2013;105:1628-33.
31. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 2009;101:1446-52.
32. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J Natl Cancer Inst* 2007;99:1036-43.
33. Wang SJ, Hung HM, O'Neill RT. Adaptive patient enrichment designs in therapeutic trials. *Biom J* 2009;51:358-74.
34. Rosenblum M, Van der Laan MJ. Optimizing randomized trial designs to distinguish which subpopulations benefit from treatment. *Biometrika* 2011;98:845-60.
35. Follmann D. Adaptively changing subgroup proportions in clinical trials. *Statistic Sinca* 1997;7:1085-102.
36. Simon N, Simon R. Adaptive enrichment designs for clinical trials. *Biostatistics* 2013;14:613-25.
37. 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.
38. Wang SJ, James Hung HM, O'Neill RT. Impacts on type I error rate with inappropriate use of learn and confirm in confirmatory adaptive design trials. *Biom J* 2010;52:798-810.
39. Scher HI, Nasso SF, Rubin EH, et al. Adaptive clinical trial designs for simultaneous testing of matched diagnostics and therapeutics. *Clin Cancer Res* 2011;17:6634-40.
40. Matsui S, Simon R, Qu P, et al. Developing and validating continuous genomic signatures in randomized clinical trials for predictive medicine. *Clin Cancer Res* 2012;18:6065-73.
41. Gu X, Yin G, Lee JJ. Bayesian two-step Lasso strategy for biomarker selection in personalized medicine development for time-to-event endpoints. *Contemp Clin Trials* 2013;36:642-50.
42. Freidlin B, Jiang W, Simon R. The cross-validated adaptive signature design. *Clin Cancer Res* 2010;16:691-8.

Cite this article as: Simon R. Biomarker based clinical trial design. *Chin Clin Oncol* 2014;3(3):39. doi: 10.3978/j.issn.2304-3865.2014.02.03