Advances in Clinical Trial Designs for Predictive Biomarker Discovery and Validation

Richard Simon, DSc

Corresponding author
Richard Simon, DSc
Biometric Research Branch, National Cancer Institute,
9000 Rockville Pike, Bethesda, MD 20892, USA.
E-mail: rsimon@nih.gov

Current Breast Cancer Reports 2009, **1**:216–221
Current Medicine Group LLC ISSN 1943-4588

Cancers of the same primary site are in many cases heterogeneous in molecular pathogenesis, clinical course, and treatment responsiveness. Current approaches for treatment development, evaluation, and use result in treatment of many patients with ineffective drugs and lead to the conduct of large clinical trials to identify small, average treatment benefits for heterogeneous groups of patients. New genomic and proteomic technologies provide powerful tools for the identification of patients who require systemic or aggressive treatment and the selection of those likely or unlikely to benefit from a specific regimen. In spite of the large literature on developing prognostic and predictive biomarkers and on statistical methodology for analysis of high dimensional data, there is considerable uncertainty about proper approaches for the validation of biomarker-based diagnostic tests. This article attempts to clarify these issues and provide a guide to recent publications on the design of clinical trials for evaluating the clinical utility and robustness of prognostic and predictive biomarkers.

Introduction

This article reviews the developments in clinical trial designs for development of predictive biomarkers. First, however, it is important to clarify terminology. Biomarkers are biological measurements that are used for diverse purposes. *Validation* has meaning only in the sense of "fit for purpose" and so general definitions of the term *biomarker* sometimes create confusion by mixing the standards for validation of one type of biomarker with those of others.

Biomarkers can be subdivided into measurements that are made once at baseline and those that are made repeatedly. Here, the focus is on baseline biomarkers. They are usually divided into prognostic biomarkers and predictive biomarkers.

Prognostic biomarkers are often defined as measurements made at diagnosis that provide information about patient prognosis in the absence of treatment or in the presence of standard treatment. In many prognostic marker studies, heterogeneous groups of patients are included without regard to stage or treatment. Such studies rarely result in the development of markers that are used in clinical practice [1]. Markers are not usually measured in practice unless they have utility in the sense of informing physicians to make improved treatment decisions. Perhaps the most basic problem with the design of many prognostic marker studies is their lack of focus on a medical indication. This lack of focus results in heterogeneous patient selection and an exploratory approach to data analysis.

Predictive biomarkers are measured at baseline to identify patients who are likely or unlikely to benefit from a specific treatment. Estrogen receptor (ER) overexpression is probably both a prognostic and a predictive biomarker. Patients with ER-positive tumors have longer survival in the absence of systemic therapy, making ER a prognostic marker. ER positivity is a predictive marker for benefit from anti-estrogens such as tamoxifen. ER negativity is also a predictive marker for benefit from several cytotoxic chemotherapy regimens. Human epidermal growth factor receptor 2 (HER-2) amplification is a predictive marker for benefit from trastuzumab and perhaps also from doxorubicin [2,3] and taxanes [4]. A predictive biomarker can also be used to identify patients who are poor candidates for a particular drug. For example, advanced colorectal cancer patients whose tumors have *KRAS* mutations appear to be poor candidates for treatment with epidermal growth factor receptor (EGFR) antibodies [5].

Developing Predictive Biomarkers

Predictive biomarkers based on single gene/protein measurements are attractive because they are often closely linked to the target of the drug and are thus biologically interpretable. In some cases, the target of the drug is known but it is not clear how to best measure target

inhibition or whether the target is driving tumor growth and invasion for an individual patient. In other cases, the drug may have several targets and the options for measurement will be more numerous. Sawyers [6•] has stated "One of the main barriers to further progress is identifying the biological indicators, or biomarkers, of cancer that predict who will benefit from a particular targeted therapy." If a predictive biomarker is to be co-developed with the drug, then the phase 1 and phase 2 studies should be designed to evaluate the candidate markers and assays available, to select one, and then to perform analytical validation of the assay prior to launching the phase 3 trial. Accomplishing all of this prior to initiation of a phase 3 trial can be very challenging.

The term *classifier* refers to a test that translates biomarker measurements to a set of predicted categories. For a predictive classifier, the categories often refer to patients most likely to respond to the new regimen and those less likely to respond. A biomarker based on a measurement involving a single gene or protein can be converted to a classifier by introducing one or more cut-points, depending on how many categories are desired. Classifiers can also be defined by introducing cut-points to the summary score, which combines the expression levels of many genes [7].

Gene expression profiling

Many algorithms have been used effectively with DNA microarray data for predicting a binary outcome. Dudoit et al. [8] compared algorithms using several publicly available data sets. The simplest methods, such as diagonal linear discriminant analysis and nearest neighbor classification, generally performed as well or better than the more complex methods.

A gene expression–based classifier may involve measurement of the expression of many genes, but it is a discrete indicator of two or more classes and can be used for selecting or stratifying patients in a clinical trial just like a classifier based on a single gene or protein. A gene expression classifier is not just a set of genes, however, and investigators who develop prognostic or predictive gene expression–based signatures should publish how the genes were weighted and what cut-points they used to develop risk groups, not just the list of their genes [9•].

Potti et al. [10] and Bennefoi et al. [11] reported the development of predictive biomarkers of response to standard chemotherapeutic agents using human tumor cell lines. Coombes et al. [12] and Baggerly et al. [13] were unable to confirm those findings. There is substantial interest in using cancer cell lines to develop single gene predictive biomarkers for molecularly targeted drugs [14].

van't-Veer and Bernards [15] indicated that gene expression–profiling studies have been more successful for developing prognostic markers than for predicting responses to particular therapies. They offer several reasons for this, including that the latter is a more difficult challenge and that sufficient numbers of tumor samples are rarely available for patients with metastatic disease who have received a specific therapy. They point out that in developing a predictive signature of benefit from a specific adjuvant regimen, samples from patients on a randomized clinical trial comparing the regimen with a control are necessary. Correlating gene expression levels to disease-free survivals for patients who have received a specific regimen does not ensure that the marker is predictive and not just prognostic.

Sample size for marker development

The problem of having a sufficient number of responders is most severe in attempting to develop a de novo gene expression–based predictive classifier based on whole genome profiling. Pusztai et al. [16] described a simulation experiment attempting to discover HER-2 overexpression as a predictive biomarker of response to trastuzumab based on data from a phase 2 trial and concluded that the likelihood of successful discovery was small. They recommend using candidate predictors based on the mechanism of action of the drug. They propose a tandem two-step phase 2 trial design for use with a single prespecified candidate predictor. During the first stage, unselected patients are treated. If insufficient responses are seen, the trial remains open to marker-positive patients only until there are sufficient numbers of them for separate analysis. This approach could be generalized for use with several candidate predictive biomarkers. If enough responses were not observed during the initial unselected stage, then accrual would remain open to patients who were positive for any of the candidate predictors.

The simulations by Pusztai et al. [16] were based on synthesizing phase 2 trials containing 60 patients (45 patients with normal HER-2 [including no responders] and 15 patients with HER-2 amplifications [including 5 responders]). Dobbin and Simon [17] studied sample size requirements for development of binary predictors based on de novo gene expression profiling. They recommended that at least 20 patients per group (responders and nonresponders) be included [17,18].

Phase 3 Clinical Trial Designs for Evaluating New Treatments and Predictive Biomarkers

A phase 3 therapeutic clinical trial should evaluate a new treatment with regard to a measure of patient benefit for a defined target population [19]. The role of a predictive biomarker is in the definition of the target patient population for whom the treatment is evaluated. For a defined population, the evaluation involves comparing outcomes for the patients treated with the new regimen to outcomes of patients in the control group. A predictive biomarker is a marker for which the treatment versus control difference (ie, treatment effect) differs between marker-positive and marker-negative patients. Comparing disease-free survival for marker-positive and marker-negative patients treated with the new regimen is not part of the evaluation of a predictive biomarker.

There are many challenges in using candidate predictive biomarkers in the design of phase 3 trials for evaluating treatments. In some cases, there may be so much biological and phase 2 evidence that marker-negative patients do not benefit from the new treatment that it would not be appropriate to include such patients in the phase 3 trial. In other cases, it will be appropriate to include marker-negative patients, but the challenge is to limit the number of such patients or to design the clinical trial in a manner that supports claims for the overall population if it turns out that the candidate marker is not useful but that the treatment is effective. It is also important to be able to design the phase 3 trial without including vastly more patients than would have been necessary without using the biomarker. As Jorgensen [20] points out: "If personalized medicine is to have a real breakthrough there needs to be incentives for those who are going to do the research and development work – the pharmaceutical and diagnostic companies." The following sections review a variety of clinical trial designs that have been proposed for phase 3 trials of new drugs and predictive biomarkers.

Marker strategy design

The marker strategy design is sometimes considered for evaluating the medical utility of a predictive marker for informing the use of approved chemotherapy [21,22]. With this design, patients are randomized to be tested or not. For those who are not tested, their treatment is determined based on stage and standard clinical prognostic factors and practice standards. For those patients randomized to be tested, the results of the test can be used in conjunction with stage and standard prognostic factors to inform treatment decisions. Although the marker strategy design is regarded by some as a gold standard, it is often inefficient because many patients may receive the same treatment regardless of the group to which they are randomized [23]. In order to have reasonable statistical power to detect differences in outcome among the two randomization groups as a whole, a very large number of patients may have to be randomized. This inefficiency is particularly problematic for prognostic markers for identifying low-risk patients for whom chemotherapy may be withheld because the prospective study is a therapeutic equivalence trial involving detection of a small treatment effect.

The defects in the marker strategy design can be avoided by performing the test in all patients and only randomizing patients for whom the treatment assignment is influenced by marker result. This was the approach used in the Microarray in Node-Negative Disease May Avoid Chemotherapy (MINDACT) trial for evaluating a 70-gene signature for guiding the use of standard chemotherapy in women with node-negative breast cancer [24].

Enrichment designs

With an enrichment design, a diagnostic test is used to restrict eligibility for a randomized clinical trial comparing a regimen containing a new drug with a control regimen. This approach played a crucial role in the development of trastuzumab. Patients with metastatic breast cancer whose tumors expressed HER-2 in an immunohistochemistry test were eligible for randomization. Because the drug has little effect for test-negative patients and because about 75% of patients are negative, a standard clinical trial randomizing all comers would require an enormous sample size to detect the diluted treatment effect. Pusztai et al. [25] describe simulations that illustrate this dilution effect.

Simon and Maitournam [26–28] studied the efficiency of this approach relative to the standard approach of randomizing all patients without using the test at all. They found that the efficiency of the enrichment design depended 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 often substantially smaller than the number of randomized patients required for a standard design. Zhao and Simon [29] have made the methods of sample size planning for the design of enrichment trials available online. The Internet-based programs are available for binary and survival/disease-free survival end points.

The enrichment design is appropriate for contexts where there is such a strong biological basis for believing that test-negative patients will not benefit from the new drug that including them in the study would raise ethical concerns. The enrichment design does not provide data on the effectiveness of the new treatment compared with control for test-negative patients. Consequently, unless there are phase 2 data on the clinical validity of the test for predicting response or compelling biological evidence that the new drug is not effective in test-negative patients, the enrichment design may not be adequate to support approval of the test.

Including test-positive and test-negative patients

When a predictive classifier has been developed but there are not compelling biological or phase 2 data that test-negative patients do not benefit from the new treatment, it is generally best to include both classifier-positive and classifier-negative patients in the phase 3 clinical trials comparing the new treatment with the control regimen. In this case, it is essential that an analysis plan be predefined in the protocol for how the predictive classifier will be used in the analysis. It is not sufficient to just stratify (ie, balance) the randomization with regard to the classifier without specifying a complete analysis plan. In fact, the main importance of stratifying the randomization is that it assures that only patients with adequate test results will enter the trial.

The purpose of the pivotal trial is to evaluate the new treatment in the subsets determined by the prespecified classifier. The purpose is not to modify or optimize the classifier. If the classifier is a composite gene expression–based

classifier, the purpose of the design is not to re-examine the contributions of each gene. Simon [30••,31] described several specific primary analysis strategies, and sample size planning for all of these analysis plans are available online [29]. For example, if one has very limited a priori confidence in the predictive marker, one can use it for a "fall-back" analysis. Simon and Wang [23] proposed an analysis plan in which the new treatment group is first compared with 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 with 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. [32] 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.

Adaptively modifying types of patients accrued

Wang et al. [32] proposed a phase 3 design comparing a new treatment with a control that starts with accruing both test-positive and test-negative patients. An interim analysis is performed evaluating the new treatment in the test-negative patients. If the observed efficacy for the control group exceeds that for the new treatment group and the difference exceeds a futility boundary, then accrual of test-negative patients terminates and accrual of additional test-positive patients is substituted for the unaccrued test-negative patients until the originally planned total sample size is reached. Wang et al. [32] show computer simulations that indicate this design has greater statistical power than nonadaptive approaches, but their design involves many more test-positive patients and may require much longer trial duration.

Liu et al. [33] proposed a two-stage accrual design in which only marker-positive patients are accrued during the first stage. At the end of the first stage, an interim analysis is performed comparing outcome for the new treatment versus control for the marker-positive patients. If the results are not promising for the new treatment, then accrual stops and no treatment benefit is claimed. If the results are promising for the marker-positive patients at the end of the first stage, then accrual continues for marker-positive patients and it also commences for marker-negative patients in the second stage.

Adaptive threshold design

Jiang et al. [34] reported on a "biomarker adaptive threshold design" for situations where a predictive index is available at the start of the trial but a cut-point for converting the index to a binary classifier is not established. With their design, tumor specimens are collected from all patients at entry, but the value of the predictive index is not used as an eligibility criterion. Analysis begins with comparing outcomes for all patients receiving the new treatment with those for all control patients. If this difference in outcomes is significant at a prespecified significance level of α_1, then the new treatment is considered effective for the eligible population as a whole. Otherwise, a second-stage test is performed using a significance threshold of $\alpha_2 = 0.05 - \alpha_1$. The second-stage test involves finding the cut-point b* for the predictive index, which leads to the largest treatment versus control treatment effect when restricted to patients with predictive index above b*. Statistical significance is determined by randomly permuting the labels of which patients are in the new treatment group and which are controls, and then determining the maximized cut-point restricted treatment effect for the permuted data. This is done for thousands of random permutations. A significance threshold of α_2 is used for this second stage of analysis. Jiang et al. [34] also describe construction of a confidence interval for the optimal cut-point b* using a bootstrap resampling approach.

Adaptively determining the marker

For co-development of a new drug and companion diagnostic, it is best to have the candidate diagnostic completely specified and analytically validated prior to its use in the pivotal clinical trials. This is difficult, however, and in some cases is not feasible. The approach of Jiang et al. [34] can be generalized to the setting where one has several candidate predictive classifiers (eg, B_1, $B_{2...}$, B_K). If S(k) denotes the log-likelihood of treatment effect for patients positive for biomarker B_k and k* denotes the biomarker for which S(k) is maximum, then the statistical significance of S(k*) is determined by permuting the treatment group labels of the patients and then re-evaluating the treatment effects within the positive subsets of the K binary classifiers. Using bootstrap resampling, one can evaluate the proportion of the times that each patient is included in the positive subset of the selected biomarker and obtain a confidence interval for the treatment effect in the selected subset.

Freidlin and Simon [35] proposed a design for a phase 3 trial that can be used when no predictive classifier is available at the start of the trial. The analysis plan is in two parts. At the conclusion of the trial, the new treatment is compared with the control overall using a threshold of significance of α_1, which is somewhat less than the total 0.05. A finding of statistical significance at that level is taken as support of a claim that the treatment is broadly effective. If the overall treatment effect is not significant at the α_1 level, then a second analysis takes place. The patients are divided into a training set and testing set. The data for patients in the training set is used to define a single subset of patients who appear to benefit from the new treatment compared with the control. When that classifier has been developed on the training set, the new treatment is compared with the control for classifier-positive patients in the test set. The comparison of new treatment with control for the subset is restricted to patients in the test set in order to preserve the principle of separating the data used to develop a classifier from the data used to test treatment effects in subsets defined by that classifier. The comparison of treatment with control

for the subset uses a threshold of significance of 0.05 - α_1 in order to assure that the overall chance of a false-positive conclusion is no greater than 0.05. These thresholds can be sharpened using the methods of Song and Chi [36].

Friedlin and Simon [35] proposed the adaptive signature design in the context of de novo development of multivariate gene expression–based classifiers. The approach can be used more broadly, however. For example, it could be used when several candidate gene expression signatures are available at the outset and it is not clear which ones to include in the final statistical testing plan. It could also be used with classifiers based on a single gene but several candidate tests for measuring expression or deregulation of that gene. In these settings with a few candidate classifiers, a smaller training set may suffice instead of the 50/50 split used by Freidlin and Simon [35].

Conclusions

Advances in cancer genomics and biotechnology are providing increased opportunities for development of more effective therapeutics and predictive biomarkers to inform their use. These opportunities have enormous potential benefits for patients and for containing health care costs. Co-development of drugs and companion diagnostics adds complexity to the development process, however. Traditional post hoc correlative science paradigms do not provide an adequate basis for reliable predictive medicine. New paradigms are required for separating biomarker development from therapeutic evaluation. New clinical trial designs are required that incorporate prospective analysis plans that provide flexibility in identifying the appropriate target population in a manner that preserves overall false-positive error rates. Such analysis plans must be constructed to provide information about the specificity of treatment effects without requiring sample sizes so large as to discourage development of predictive biomarkers or to require physicians to expose large numbers of patients to drugs from which they are not expected to receive benefit. This article has attempted to clarify some areas of confusion about predictive biomarkers and their development and to provide a review of recently developed clinical trial designs.

Disclosure

No potential conflict of interest relevant to this article was reported.

References and Recommended Reading

Papers of particular interest, published recently, have been highlighted as:

- • Of importance
- •• Of major importance

1. Pusztai L: **Perspectives and challenges of clinical pharmacogenomics in cancer.** *Pharmacogenomics* 2004, **5:**451–454.
2. Hayes DF: **Prognostic and predictive factors revisited.** *Breast* 2005, **14:**493–499.
3. Gennari A, Sormani MP, Pronzato P, et al.: **HER2 status and efficacy of adjuvant anthracyclines in early breast cancer: a pooled analysis of randomized clinical trials.** *J Natl Cancer Inst* 2008, **100:**14–20.
4. Hayes DF, Thor AD, Dressler LG, et al.: **HER2 and response to paclitaxel in node-positive breast cancer.** *N Engl J Med* 2007, **357:**1496–1506.
5. Amado RG, Wolf M, Peeters M, et al.: **Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer.** *J Clin Oncol* 2008, **26:**1626–1634.
6. • Sawyers CL: **The cancer biomarker problem.** *Nature* 2008, **452:**548–552.

This is an insightful review of the importance and challenges in developing predictive biomarkers

7. van't-Veer LJ, Paik S, Hayes DF: **Gene expression profiling of breast cancer: a new tumor marker.** *J Clin Oncol* 2005, **23:**1631–1635.
8. Dudoit S, Fridlyand J, Speed TP: **Comparison of discrimination methods for the classification of tumors using gene expression data.** *J Am Stat Assoc* 2002, **97:**77–87.
9. • Dupuy A, Simon R: **Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting.** *J Natl Cancer Inst* 2007, **99:**147–157.

This is a review of the problems found in published studies relating tumor gene expression profiles to response or survival. It includes a list of do's and don'ts for use by authors, reviewers, and readers.

10. Potti A, Dressman HK, Bild A, et al.: **Genomic signatures to guide the use of chemotherapeutics.** *Nat Med* 2006, **12:**1294–1300.
11. Bennefoi H, Potti A, Delorenzi M, et al.: **Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG00-01 clinical trial.** *Lancet Oncol* 2007, **8:**1071–1078.
12. Coombes KR, Wang J, Baggerly KA: **Microarrays: retracing steps.** *Nat Med* 2007, **13:**1276–1277.
13. Baggerly K, Coombes K, Neeley E: **Run batch effects potentially compromise the usefulness of genomic signatures for ovarian cancer.** *J Clin Oncol* 2008, **26:**1186–1187.
14. Smollen G, Sordella R, Muir B, et al.: **Amplification of MET may identify a subset of cancers with extreme sensitivity to the selective tyrosine kinase inhibitor PHA-665752.** *Proc Natl Acad Sci U S A* 2006, **103:**2316–2321.
15. van't-Veer L, Bernards R: **Enabling personalized cancer medicine through analysis of gene-expression patterns.** *Nature* 2008, **452:**564–570.
16. Pusztai L, Anderson K, Hess KR: **Pharmacogenomic predictor discovery in phase II clinical trials for breast cancer.** *Clin Cancer Res* 2007, **13:**6080–6086.
17. Dobbin K, Simon R: **Sample size planning for developing classifiers using high dimensional DNA expression data.** *Biostatistics* 2007, **8:**101–117.
18. Dobbin KK, Zhao Y, Simon RM: **How large a training set is needed to develop a classifier for microarray data?** *Clin Cancer Res* 2008, **14:**108–114.
19. Simon R: **Randomized clinical trials: principles and obstacles.** *Cancer* 1994, **74:**2614–2619.
20. Jorgensen JT: **From blockbuster medicine to personalized medicine.** *Personalized Med* 2008, **5:**55–63.
21. Pusztai L, Hess KR: **Clinical trial design for microarray predictive marker discovery and assessment.** *Ann Oncol* 2004, **15:**1731–1737.
22. Sargent D, Allegra C: **Issues in clinical trial design for tumor marker studies.** *Semin Oncol* 2002, **3:**222–230.
23. Simon R, Wang SJ: **Use of genomic signatures in therapeutics development.** *Pharmacogenomics J* 2006, **6:**1667–1673.
24. Bogaerts J, Cardoso F, Buyse M, et al.: **Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial.** *Nat Clin Practice: Oncol* 2006, **3:**540–551.

25. Pusztai L, Broglio K, Andre F, et al.: **Effect of molecular disease subsets on disease-free survival in randomized adjuvant chemotherapy trials for estrogen-receptor positive breast cancer.** *J Clin Oncol* 2008, **26:**4679–4683.
26. Simon R, Maitournam A: **Evaluating the efficiency of targeted designs for randomized clinical trials.** *Clin Cancer Res* 2005, **10:**6759–6763.
27. Simon R, Maitournam A: **Evaluating the efficiency of targeted designs for randomized clinical trials: supplement and correction.** *Clin Cancer Res* 2006, **12:**3229.
28. Maitournam A, Simon R: **On the efficiency of targeted clinical trials.** *Stat Med* 2005, **24:**329–339.
29. Biometric Research Branch: **Division of Cancer Treatment and Diagnosis.** Available at http://brb.nci.nih.gov. Accessed July 17, 2009.
30.•• Simon R: **Using genomics in clinical trial design.** *Clin Cancer Res* 2008, **14:**5984–5993.
This is an in-depth discussion of the special features of designing randomized clinical trials for evaluating new treatments and putative predictive biomarkers. It provides specific primary analysis plans and sample size considerations.
31. Simon R: **Designs and adaptive analysis plans for pivotal clinical trials of therapeutics and companion diagnostics.** *Expert Rev Mol Diag* 2008, **2:**721–729.
32. Wang SJ, O'Neill RT, Hung HM: **Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset.** *Pharm Stat* 2007, **6:**227–244.
33. Liu A, Li Q, Yu KF, Yuan VW: **A threshold sample-enrichment approach in a clinical trial with heterogeneous subpopulations.** *Stat Med* 2009 (in press).
34. 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:**1036–1043.
35. 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:**7872–7878.
36. Song Y, Chi GY: **A method for testing a prespecified subgroup in clinical trials.** *Stat Med* 2007, **26:**3535–3549.