Statistics in Clinical Cancer Research

Clinical Cancer Research

Adaptive Prediction Model in Prospective Molecular Signature–Based Clinical Studies

Guanghua Xiao[1], Shuangge Ma[6], John Minna[2,3,5], and Yang Xie[1,4]

Abstract

Use of molecular profiles and clinical information can help predict which treatment would give the best outcome and survival for each individual patient, and thus guide optimal therapy, which offers great promise for the future of clinical trials and practice. High prediction accuracy is essential for selecting the best treatment plan. The gold standard for evaluating the prediction models is prospective clinical studies, in which patients are enrolled sequentially. However, there is no statistical method using this sequential feature to adapt the prediction model to the current patient cohort. In this article, we propose a reweighted random forest (RWRF) model, which updates the weight of each decision tree whenever additional patient information is available, to account for the potential heterogeneity between training and testing data. A simulation study and a lung cancer example are used to show that the proposed method can adapt the prediction model to current patients' characteristics, and, therefore, can improve prediction accuracy significantly. We also show that the proposed method can identify important and consistent predictive variables. Compared with rebuilding the prediction model, the RWRF updates a well-tested model gradually, and all of the adaptive procedure/parameters used in the RWRF model are prespecified before patient recruitment, which are important practical advantages for prospective clinical studies. *Clin Cancer Res; 20(3); 531–9. ©2013 AACR.*

Introduction

The goal of molecular signature–based medicine is to use patients' molecular profiles and clinical information to predict their clinical outcomes such as survival before treatment and, thereby, select the best possible therapy, which can greatly improve the efficacy and reduce the toxicities of treatments. Recent studies have shown its feasibility and challenges. For example, the gene expression profiles have been used to predict disease prognosis (1–4) and responses to treatments (5, 6) in multiple types of cancers. The conventional strategy for developing those prediction models is to build a model (also called a predictive signature) from one dataset (called a training set), and then validate the model using one or several independent datasets (called testing or validation sets; refs. 7–10). In this circumstance, testing the model prospectively, i.e., using prospective studies as the testing dataset, is the most objective and unbiased approach. A major challenge in developing a clinically useful predictive signature is the heterogeneity between the training and testing datasets, which may be caused by different patients' cohorts and experimental procedures. In addition, the testing dataset in a prospective study is always collected after the training set, so potential batch effects associated with profiling experiments may also lead to the heterogeneity. One feature of a prospective study is that the patients are usually recruited sequentially into the study, so clinical outcomes from the earlier patients accumulate during the study. The conventional approaches use a fixed prediction model, which is built on the training data only, throughout the entire study. Intuitively, such an approach can be less efficient for patients enrolling later in the testing set, as information on patients enrolling earlier in the testing set is not used. It is desirable to have a rigorous and prespecified mechanism, so that the information accumulated from earlier patients in the study can be used to update the prediction model for subsequent patients. In this article, we develop an adaptive prediction methodology to address these heterogeneity and efficiency issues, by using the accumulated information in the testing cohort to validate and update the prediction model. There are a few related issues for molecular signature–based clinical studies, such as the sample size and designs. We acknowledge the importance of these issues, but refer to other published studies for more discussion (11–14).

Authors' Affiliations: [1]Quantitative Biomedical Research Center, Department of Clinical Sciences; Departments of [2]Internal Medicine and [3]Pharmacology; [4]Simmons Cancer Center; [5]Hamon Center for Therapeutic Oncology, University of Texas Southwestern Medical Center, Dallas, Texas; and [6]Department of Biostatistics, School of Public Health, Yale University, New Haven, Connecticut

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Yang Xie, Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390. Phone: 214-648-5178; Fax: 214-648-1663; E-mail: Yang.Xie@UTSouthwestern.edu

doi: 10.1158/1078-0432.CCR-13-2127

Random forest (rf) prediction model (15) is an ensemble learning method using classification trees as the base classifier. For high-dimensional data, random forest has comparable or superior performance compared with alternatives (16, 17). In this article, we introduce a reweighted random forest (RWRF) model, which gives different weights to decision trees in the prediction model. The weights are adjusted within the clinical study, using the information accumulated from earlier patients. By doing so, the prediction model becomes adapted to the patient cohort in the current study, and the prediction performance will be improved. Figure 1 illustrates the rationale of our approach. Before the study starts, a prediction model (model 1) is built on the basis of training data. When the first patient is enrolled into the study, model 1 is used for prediction. Model 1 is also applied to the second patient. Suppose that the clinical outcome (resistant or sensitive to the treatment) of the first patient is available before the third patient is enrolled. This information will be used to update model 1, by adjusting the weight of each classification tree. The classification trees that correctly predict the outcome of the first patient will have increased weights, and those that predict incorrectly will have decreased weights. As the third patient enters the study, the prediction is made on the basis of the updated model (model 2). This evaluation and updating process continues whenever new information is available throughout the entire study. The prediction is always made by the newest model, which has been updated using all available information. In this method, the adaptation refers to updating the prediction model, using available patient information to improve the prediction accuracy in the new cohort.

Recent studies (18, 19) have shown the benefit of using weighted approaches to combine classifiers. Wolpert (20) developed a stacking method using a weighted average. Pan and colleagues (21) demonstrated a more general scheme using input-dependent weights. The goal of those methods is to combine different classifiers. The proposed approach shares a similar spirit. In addition, it may significantly advance from the existing approaches by continuously adapting the prediction model to the new patient cohort. In the proposed RWRF method, the weights are updated using the newly available information, and so the prediction model is adjusted to the patients' characteristics in the new cohort.

A motivating example

Lung cancer is the leading cause of death from cancer in the United States, with a 5-year survival rate of 15% (2), and non–small cell lung cancer (NSCLC) accounts for up to 85% of lung cancer–related deaths (22). The goal of treating late-stage NSCLC with chemotherapy is to prolong patients' survival time with limited toxicities. Current first-line chemotherapy options for patients with advanced NSCLC, such as the combination of a platinum-based agent with paclitaxel, gemcitabine, vinorelbine, or docetaxel, have substantial toxicity and limited clinical efficacy (23). Gefitinib (Iressa, ZD1839; AstraZeneca) is an orally active EGF receptor (EGFR) tyrosine kinase inhibitor, and has been approved by the U.S. Food and Drug Administration (FDA) to treat advanced NSCLC. Four phase I studies have shown that gefitinib is generally well tolerated (23–26). NSCLC patients' responses to gefitinib are very diverse—some patients can completely recover from the advanced cancer but others do not respond to the treatment at all. Therefore, identifying the subgroup of patients who will respond to gefitinib has tremendous clinical benefit for NSCLC

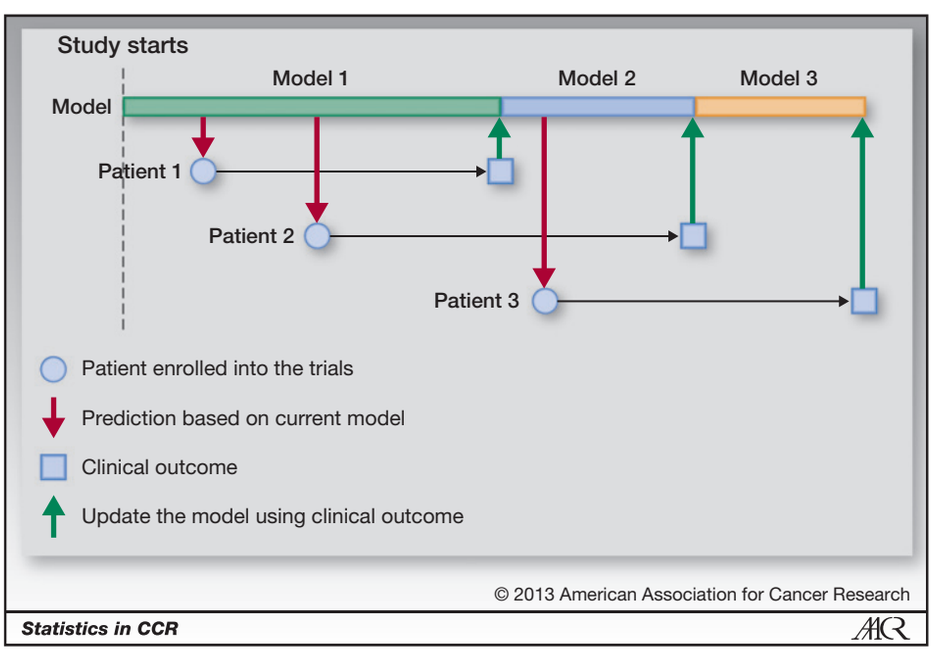

Figure 1. Illustration of adaptive prediction models for molecular signature–based clinical studies.

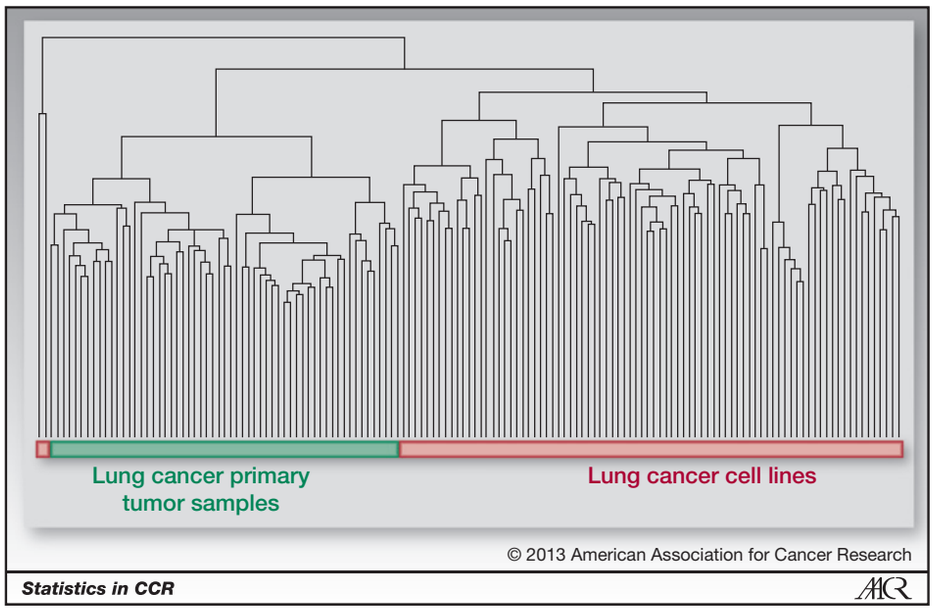

Figure 2. Hierarchical clustering after quantile–quantile normalization. Gene expressions from patients are labeled as green and those from cell lines are labeled as red. Even after normalization, the patient and cell line samples form different clusters, indicating differences between patient and cell line gene expression.

treatment. A promising approach to identifying the gefitinib-sensitive patient subgroup is to use genomic profiling to predict tumor sensitivity to gefitinib. However, it is challenging to develop gefitinib response–predictive signatures using patients' molecular profiles, as not many patients have been treated with gefitinib, and have frozen tumor samples available for molecular profiling. Alternatively, the predictive profiles can be generated using cell-line models (*in vitro*) as a "short cut" (27). For decades, human immortal cancer cell lines have constituted an accessible, easily usable set of biologic models with which to investigate cancer biology, and to explore the potential efficacy of anticancer drugs. Nowadays, cancer cell lines have become valuable sources for studying responses to new therapeutic drugs because cell line responses to any treatment can be tested in a laboratory right after the drug development process. To develop a gefitinib response gene signature, the University of Texas Lung Specialized Program of Research Excellence (SPORE) collected a large amount of data on drug response to gefitinib and gene expression for 86 NSCLC cell lines. The gefitinib response of each cell line was measured by 50% inhibition concentration (IC_{50}) values using MTS assay, and each cell line can be categorized as sensitive or resistant to gefitinib based on its IC_{50} value. The expression of approximately 43,000 probes from 86 NSCLC cell lines, along with 59 primary tumor samples from patients with NSCLC, was measured using Affymetrix U133AB GeneChips. The goal of the study is to develop a predictive signature using the cell line data, and then test the signature on the independent datasets with primary tumors from patients with lung cancer.

We first checked whether the expression profiles from primary tumors are different from those from cell lines. The gene expression data were processed using the Robust multichip Average (RMA) approach and quantile–quantile normalization (28). All gene expression values were $\log_2$ transformed. Average values were used for the different probe sets corresponding to the same gene. Figure 2 shows the hierarchical clustering result of NSCLC lines and primary tumors of patients with lung cancer, based on gene expression profiles. It is clear that the cell lines and patients samples were separated into different clusters, indicating that differences remain even after stringent normalization. These differences between training (cell line) data and testing (patient) data need to be taken into account in prediction models. This motivated us to develop an adaptive prediction method to account for the difference between testing and training data, by gradually adjusting the prediction model using available information.

Materials and Methods

Random forest

Random forests are classification and regression methods based on growing an ensemble of many randomized classification trees. For the integrity of this article, we give a brief review of the method here. We denote x^{tr} as the input variables and y^{tr} as the outcome variables in a training dataset with size N. For the drug response example, x is the gene expression profile, and y is the drug response status ($y = 1$ for sensitive cases and $y = -1$ for resistant cases) for lung cancer cell lines. The random forest algorithm proceeds as follows: (i) Randomly select N samples from the original training set with replacement (the bootstrap samples). (ii) A tree-based classifier $f^b(x)$ is constructed using the bth random training set. In the drug response example, $f^b(x)$ is a binary function [$f^b(x) = 1$ for the predicted sensitive cases and $f^b(x) = -1$ for the predicted resistant cases]. (iii) Repeat steps (i) and (ii) for B times. (iv) The final classifier from the random forest model is determined by the

majority vote of all B trees, and the prediction is based on:

$$f_{\mathrm{rf}} = \mathrm{sign}\left(\frac{1}{B}\sum_{b=1}^{B} f^b(x)\right) \quad (1)$$

The random forest model predicts the probability that a new observation is sensitive to the treatment. The prediction was dichotomized into binary variables using 0.5 as cutoff (i.e., a new observation with a predicted probability of being sensitive greater than 0.5 was predicted as sensitive and otherwise as resistant).

RWRF for adaptive prediction

The classifiers from the conventional random forest models are constructed on the basis of the training data only. In this study, we developed a RWRF method to incorporate the information generated from earlier patients in the clinical study, to account for the potential heterogeneity between training and testing data, and, hence, improve prediction performance.

Supposing that the total number of patients in a prospective clinical study is M, in our proposed method, the prediction for the kth patient is based on a weighted average of the classification trees:

$$f_i(x_k) = \frac{1}{\sum_{b=1}^{B} w_{b,i}} \sum_{b=1}^{B} w_{b,i} f^b(x_k) \quad (2)$$

where x_k is the gene expression data of patient k, and i denotes the index set of subjects whose clinical outcomes are available when patient k is enrolled in the study ($i \leq k$). $f^b(x)$ denotes the bth classification tree, and its weight in the new prediction model is $w_{b,i}$. Here, $f_i(x_k)$ is the prediction model for patient k. In the proposed model, the weight of each individual tree is determined by its performance with the previous patients whose clinical outcomes are available. At the beginning of the study, the weights are set to be equal, i.e., $w_{1,1} = w_{2,1} = \ldots = w_{B,1} = 1$, so $f_1(x)$ is equivalent to $f_{\mathrm{rf}}(x)$, the standard random forest. When the clinical outcome of patient $i+1$ is available, the weights are adjusted according to the prediction performance of each individual tree by:

$$w_{b,i+1} = w_{b,i} e^{\alpha I[y_{i+1} = f^b(x_{i+1})]} \quad (3)$$

where $I(x)$ is an indicator function, which equals 1 for a correct prediction and 0 otherwise, and α is a positive constant that determines the learning speed. If a tree predicts the outcome of patient $i+1$ correctly, then its relative weight, $w_{b,i+1} / \sum w_{b,i+1}$ in the prediction model for later patients, $f_{i+1}(x_k)$, $k \geq i$, will increase, and vice versa. Intuitively, the model gives more weights to the trees with good prediction performance in all previous samples (both in the training and available testing datasets).

To evaluate the role of each variable (i.e., the importance of expression of a single gene) in the prediction models, we define $c_{j,i}$ as the contribution of variable (gene) j in the prediction model i as:

$$c_{j,i} = \sum_{b=1}^{B} w_{b,i} q_{j,b} \quad (4)$$

where $q_{j,b}$ is the frequency of variable j appearing in the bth classification tree, and $w_{b,i}$ is the relative weight of the bth classification tree in the prediction model i. In the RWRF model, the contributions of variables change as the study goes on, and we define an adaptive score (AS) for gene j as

$$\mathrm{AS}_j = \frac{c_{j,M}}{c_{j,1}} \quad (5)$$

where $c_{j,M}$ is the contribution of gene j in the final model, and $c_{j,1}$ is the contribution of gene j in the initial model.

Simulation studies

We used simulation studies to evaluate the performance of the proposed RWRF method. We simulated 100 samples in the training set, and 200 patients enrolled in the study sequentially as the testing set. The outcome variable of the simulation study is sensitive or resistant to treatment, and there are 50 variables. Among those variables, 10 are "good" predictors, which are defined as the genes whose expressions are associated with the outcomes in both the training set and the testing set. These genes are real biomarkers. Another 10 are "unstable" predictors, which were simulated to be associated with the outcomes only in the training set, but not in the testing set. Such unstable predictors are commonly seen in gene signature studies. For example, assume a gene whose expression level is associated with patient outcomes and it is identified as a predictor using the training set. However, the measurement of its expression level in the testing set is very noisy due to some technical problems, so it is not a good predictor in the testing set. In this study, we proposed the adaptive prediction model to minimize the impact of the unstable predictors on the prediction performance. The variables are summarized in Supplementary Table S1. If a variable is a predictor (with a check mark in Supplementary Table S1), then it was simulated from a normal distribution N(0,1) for sensitive and N(1,1) for resistant cases. Otherwise, it was simulated from N(μ,1), where μ is a random variable from Uniform(0,1) distribution, for both sensitive and resistant cases. In this study, the variables $X1,\ldots,X10$ are good predictors, $X11,\ldots,X20$ are unstable predictors, and $X21, \ldots X50$ are noise predictors (i.e., not associated with outcome in the training set).

For simplicity, in the simulation, we assumed that the patients' clinical outcomes would be available immediately after the prediction and treatment, which is close to real practice when the response time is short and accrual rate is low. In simulation studies, we built the random forest using training data and predicted the cases in the testing data one by one, using the proposed RWRF approach. The simulations were repeated 1,000 times, and the prediction accuracy is the mean accuracy across the 1,000 simulations.

Figure 3A shows the prediction accuracy versus the number of patients enrolled in the study. The prediction accuracy increases as the study goes on. At the beginning of the study, the overall prediction accuracy is 0.82, and the accuracy increases as the prediction model gradually adapts to the testing data. The improvement is fast at the beginning

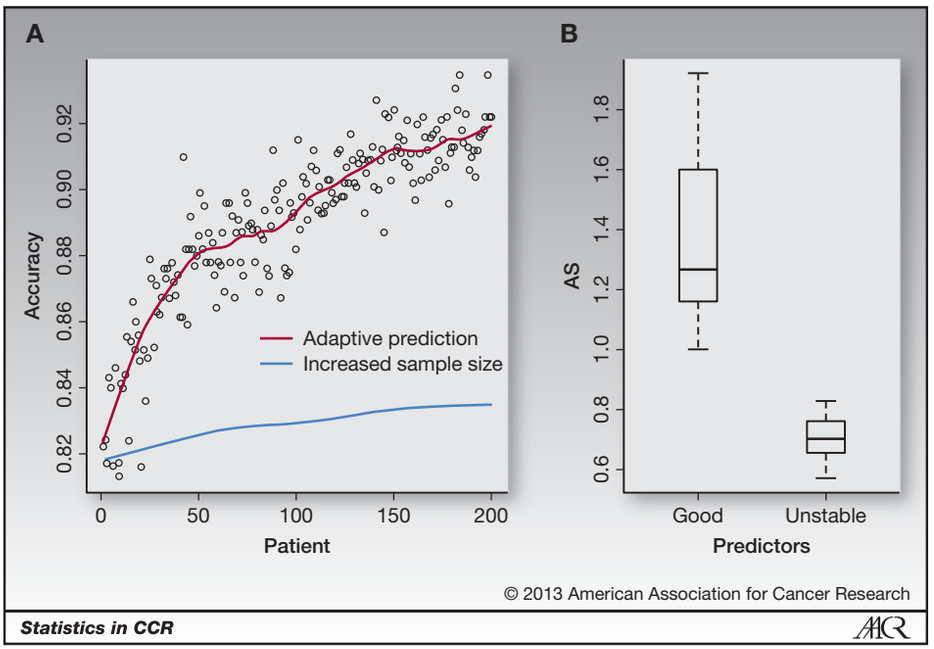

Figure 3. A, prediction accuracy using RWRF model at different stages of the studies. The accuracy for patient *k* is the mean accuracy across 1,000 simulations for the *k*th patient enrolled in the study. The circles are the mean accuracy for the *k*th patient, and the red line is the smoothed Lowess curve for accuracy of the RWRF model. Blue line represents the learning curve (i.e., prediction accuracy increases with the number of training samples increases). For the learning curve, the *x*-axis is the number increased in the training set. For example, the starting accuracy is 0.82, which corresponds with the number of training samples, i.e., 100 (the original sample size); as *x* (the number of additional training samples besides the original 100 training samples) is 50 (i.e., the total size of training set increases to 100 + 50 = 150), the accuracy is 0.83; as *x* increases to 200 (the total size of the training set is 100 + 200), the accuracy increases to 0.84. B, boxplots for the adaptive scores for good predictors and unstable predictors in the prediction models. At the end of the study, the contributions from the good predictors increased, whereas those from unstable predictors decreased.

of the study and slows down as the model is close to being fully adapted. The overall accuracy at the end of the study is 0.92, which is significantly increased in comparison with the accuracy at the beginning. Figure 3A also compared the performance of RWRF with the learning curve (i.e., prediction accuracy increases as the number of training samples increases) and shows that the accuracy improvement from RWRF is much faster than the learning curve.

We also used the receiver operating characteristic (ROC) curves to summarize the average prediction performance in 1,000 simulations at different time points and compared the performance at the beginning and the end of the study. The predicted probability score was used to determine the ROC curves. Sensitivity is the proportion of sensitive cases that were predicted correctly by the prediction model, and specificity is the proportion of resistant cases that were predicted correctly by the model. Supplementary Fig. S1 shows the ROC curves for standard random forest (black), the RWRF at the beginning of the study (red), and at the end of the study (blue). At the beginning, the RWRF has equal weights for the classification trees, so it has similar performance to the traditional random forest. The prediction performance at the end of the study [with area under the curve (AUC) = 0.94] improves significantly (P value < 0.001) beyond the traditional random forest (AUC = 0.91) as the prediction model fully adapts to the testing data set.

For variable *j*, we calculated the adaptive score (defined in equation 5) to check the contribution changes between the initial model and the final model. Figure 3B shows the adaptive score boxplots for both the good and the unstable predictors. As expected, it shows that the contributions from good predictors increased, and those from unstable predictors decreased, as the prediction model adapted to the testing data set. It indicates that the model adapted to the new data by increasing the contributions from the good predictors and decreasing those from unstable predictors.

Analysis of lung cancer data

Now return to our motivating example, in which the cell line gene expression and drug sensitivity data were used to predict the clinical response to gefitinib treatment in patients with lung cancer. In this study, we applied the RWRF model to account for the differences between cell line (training) and patient (testing) data. A training set of 86 cell lines was used to build the prediction model, and its performance was evaluated using an independent testing set of 59 patients with lung cancer. Our goal was to develop a prediction model that could work in prospective clinical studies, in which patients are enrolled sequentially and the samples in the training and testing sets are different because they came from different types of samples. As the clinical outcome of each earlier patient was available, the weights of the classification trees with correct prediction increased, and

the weights of the trees with incorrect prediction decreased. By adjusting the weights, the random forest built from training data can gradually adapt to testing data. Figure 4 illustrates the flowchart of the adaptive prediction model in the clinical setting.

In the current study, the patients' information was retrospectively collected. To simulate the prospective studies, we randomly assigned an enrollment date to each patient and used this information to test the proposed method. Previous studies (29, 30) have demonstrated that a patient with an EGFR mutation is likely to response to gefitinib. Because the responses of patients were not currently available, we used the status of the EGFR mutation, which is a major surrogate biomarker for the response to gefitinib therapy, to evaluate the performance of the proposed model. In this study, we assume that a patient with an EGFR mutation will response to gefitinib therapy. We repeated the procedure 2,000 times to derive the average performance at different time points of the study.

To account for the fact that the patients' gene expression data (testing data) are collected sequentially in prospective clinical studies, we carefully normalized the gene expression data as follows: First, the training data (cell line gene expression data) were normalized using quantile normalization (28); then, the new (patient) gene expression data were normalized one by one, to have the same distribution as the expression of the training data. In this fashion, we normalized the training and testing data in the same way as much as possible, without assuming that the testing data come altogether. The genes with low variability (SEM less than 1) in the cell line data were removed and the remaining 1,473 genes were used as predictors to build the random forest model.

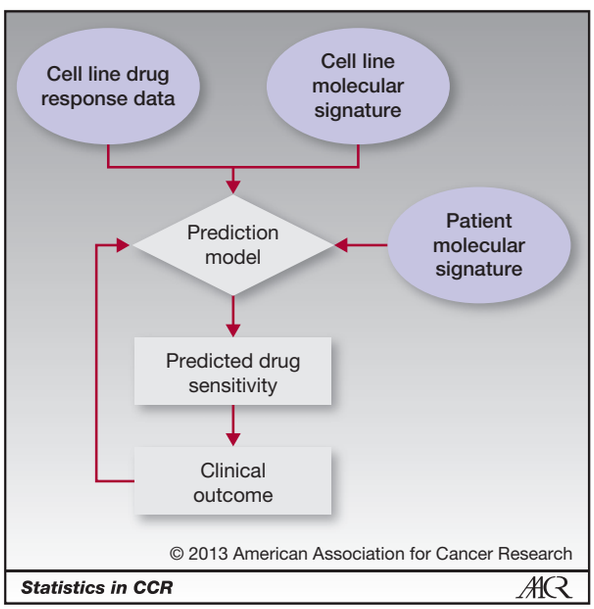

Figure 4. Flowchart of the adaptive prediction models. The prediction model was built on the cell line data and used to predict the tumor response of patients with cancer. As the clinical outcome of a patient is available, the information will be used to update the prediction model.

Figure 5A shows the improvement of the overall accuracy at different stages of the study. The accuracy at the beginning of the study was 0.74 and increased to 0.84 when the number of accumulated patients reached 40. Figure 5B presents the ROC curves for standard random forest, as well as the starting performance and ending performance of the RWRF model. The ROC curves of the RWRF represent the average performance of 2,000 simulations. It shows that the performance at the beginning of the study is the same as that of the standard random forest, as expected. As the study goes on, the weights of the classification trees were adjusted, using the information accumulated from earlier patients. The performance of the final model improved significantly from the initial model, as the prediction model adapted to the patient data.

We also checked the genes with increased adaptive scores in the prediction model to see whether they are cancer-related genes. Of note, 13 genes have adaptive scores bigger than a 30-fold increase in their contributions to prediction. As summarized in Supplementary Table S2, 11 out of 13 genes have been shown to be associated with cancer diagnosis and prognosis in other clinical studies. For example, gene *ZFHX1B* is an important transcriptional repressor in the EGFR pathway (31), which is the targeting pathway of gefitinib treatment, *THBS1* (*Homo sapiens* thrombospondin 1) is a tumorgenesis gene associated with the prognosis and drug response in many types of cancer (32), and CD24 (*Homo sapiens* CD24 antigen: small cell lung carcinoma cluster 4 antigen) is a prognostic marker of survival in NSCLC (33) and other cancer types. On the other hand, most genes with decreased weight (indicating they are unstable predictors) have been found to be differentially expressed between cell line data and patient data. For example, gene *IFITM2* had the greatest reduction in weight among all the genes and was differentially expressed between cell line and patient ($P < 0.0001$). These results indicate that the proposed model will increase the weights of good predictors and decrease the weights of unstable predictors, as expected. The ability to identify the important genes for translational study automatically diminishes the effect of the unstable predictors, which is a key advantage of having the new prediction model smoothly "evolved" from the original model built from the cell line data.

Discussion

In real practice, it is challenging to use the model developed from one data source to predict the outcome from another data source. It is especially difficult in molecular signature–based clinical studies, because microarray datasets are variable and the expression measurements tend to be different from dataset to dataset. In this article, we propose a RWRF model to account for the differences between training data and testing data. A key feature of the proposed method is to use the outcomes from earlier patients in the clinical study to adapt the prediction model to the current study cohort.

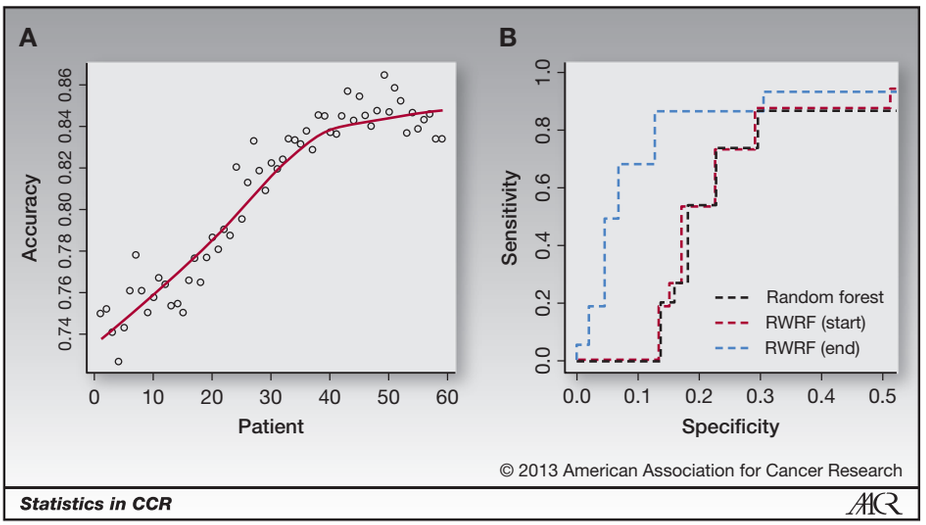

Figure 5. A, the prediction accuracy versus the number of patients enrolled in the study. The circle is the estimated accuracy obtained by averaging across 2,000 simulation runs, and the red line is the smoothed Lowess curve. The accuracy at the beginning of the trial was 0.74 and increased as the patients accumulated. The accuracy saturated at 0.84 as the number of patients was about 40. B, ROC curves for the prediction of patients' response to gefitinib treatment. The black, red, and blue lines represent the standard random forest, the starting performance of RWRF, and ending performance of RWRF. As expected, the starting performance of RWRF performs the same as the standard random forest and the prediction performance has significantly improved at the end of the study.

The procedure for updating the weights in our proposed method was inspired by AdaBoost (34), which uses the multiplicative rule to update the weights. In AdaBoost, the weights are adjusted for different observations to put more weight on the misclassified observations in previous iterations. In each iteration, the successive classifier becomes more focused on those observations misclassified by the previous one (35). In our proposed method, the weights are adjusted for different classification trees in the random forest to increase the weights of the trees with correct prediction in previous patients. The trees with correct prediction will gain more weights in the successive classifier. The trees suitable for translation from cell line to patient will increase in weight, which forces the classifier to be gradually adapted from the cell line expression profile to the patient expression profile. Intuitively, the multiplicative rule allows some classification trees that do not fit the new testing set to be gradually excluded from the model by having their weights exponentially decay to zero.

Our proposed method is similar to the stacking method (20) for model averaging but serves different purposes. In stacking, the final classifier is $\sum_{b=1}^{B} w_b^{st} f^b(x)$, and the stacking weight w^{st} is determined by $w^{st} = \arg\min \sum_{i=1}^{m} [y_i - \sum_{b=1}^{B} w_b^{st} f_{-i}^b(x)]^2$, where the $f_{-i}^b(x)$ is the prediction of observation i made by using a dataset without the ith training observation (20). In the stacking method, the weights are determined by cross validation. The stacking method can outperform each individual classifier by weighting them appropriately. In signature-based clinical studies, cross validation is not an option, as the patients are enrolled sequentially. In our proposed method, the weights are determined by the performance of each individual tree in previous patients in the clinical study, to make the classifier adapt to current patients' characteristics.

Our proposed method used a similar method as AdaBoost for adjusting the weights, and we use $\alpha = 0.1$ to illustrate the idea. The parameter α in equation 3 controls the speed of learning. If α is large ($\alpha > 1$), the model will adapt quickly but may lose stability. On the other hand, if α is small ($\alpha < 0.1$) the model adapts to new data slowly, but the prediction is relatively stable. As long as α has a moderate value (0.1–1), the model performs reasonably well. In practice, similar to adaptive designs for clinical trials, it may be necessary to conduct extensive simulation studies to pick an α value that gives the best operation characteristics. In original AdaBoost, α is a function of prediction error, and in the random forest prediction model, the prediction accuracy can be estimated internally using out-of-bag (OOB) estimator. Thus, we are studying how to control the learning speed using OOB estimation of prediction accuracy. If the prediction accuracy is much lower than that in the training data, indicating a large difference between the training and testing data, then α should be large to make the prediction model adapt quickly. On the other hand, if the prediction accuracies in the training and testing data are close, then α should be small to decrease the learning speed and gain stability.

To make use of the newly acquired data in the prediction model, instead of gradually adjusting, a tempting alternative is to rebuild the entire prediction model using newly acquired data as a part of the training set whenever new informative is available. The major problem with totally rebuilding a prediction model for a clinical study is its instability. A prediction model must go through a series of tests and validation steps before use in the genomic

signature–based clinical studies. However, these tests are time consuming and we cannot afford to test and validate each time the model is built. In real practice, we would rather make gradual updates on a well-tested and validated prediction model than totally rebuild it. Furthermore, in our proposed method, the new model and the initial model are based on the same set of genes, and the only difference is the weight of the classification tree that makes the model more stable. Another advantage of this method is that it can automatically select genes that are important in the translational study from cell line to patient. The genes identified in the real data example have been shown to be important cancer genes and are of great interest for future biologic studies. In addition, all of the adaptive procedures/parameters used in the RWRF model are predefined before recruiting the new cohort, which is an important practical advantage for prospective clinical studies using adaptive designs (36, 37).

It is worth noting that, as the adaptive prediction model gradually improves the prediction accuracy, the patients who enter early in the study are less likely to be correctly assigned to effective therapy than patients who enter late in the study. The patients entered in the study benefit from the adaptive prediction model with better accuracy, but the early-enrolled patients benefit less, whereas the late-enrolled patients benefit more. One possible alternative to the proposed adaptive prediction model is to use an "adaptation" dataset, which contains molecular profiles and clinical outcomes for patients. This adaptation dataset would be used to adapt the model to new situations, but no patient would be assigned on the basis of the model prediction until a predetermined threshold of acceptable predictive accuracy is reached.

In summary, the proposed RWRF model can effectively adapt the predictive models to current patients' characteristics and, therefore, improve the prediction accuracy significantly. The RWRF model provides a rigorous statistical framework with predefined procedures, to account for the potential heterogeneity between the training and testing cohorts. The method can facilitate using molecular signatures to predict the clinical outcomes of patients in prospective clinical studies.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: G. Xiao, J. Minna, Y. Xie
Development of methodology: G. Xiao, S. Ma, Y. Xie
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Minna, Y. Xie
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): G. Xiao, Y. Xie
Writing, review, and/or revision of the manuscript: G. Xiao, S. Ma, Y. Xie
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Y. Xie
Study supervision: Y. Xie

Grant Support

This work was supported by NIH grants 5R01CA152301 to Y. Xie, 1R01CA172211 to G. Xiao and Y. Xie, 4R33DA027592 to G. Xiao, University of Texas SPORE in Lung Cancer (P50CA70907) to J. Minna and Y. Xie, and Cancer Prevention Research Institute of Texas award RP101251 to G. Xiao and Y. Xie.

Received August 1, 2013; revised November 20, 2013; accepted November 25, 2013; published OnlineFirst December 9, 2013.

References

1. Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, et al. Gene expression predictors of breast cancer outcomes. Lancet 2003; 361:1590–6.
2. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: A multiple random validation strategy. Lancet 2005; 365:488–92.
3. Nevins JR, Huang ES, Dressman H, Pittman J, Huang AT, West M. Towards integrated clinico-genomic models for personalized medicine: Combining gene expression signatures and clinical factors in breast cancer outcomes prediction. Hum Mol Genet 2003;12:R153–7.
4. Xie Y, Xiao G, Coombes KR, Behrens C, Solis LM, Raso G, et al. Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. Clin Cancer Res 2011;17:5705–14.
5. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet 2003;362:362–9.
6. Tang H, Xiao G, Behrens C, Schiller J, Allen J, Chow CW, et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. Clin Cancer Res 2013;19:1577–86.
7. Minna JD, Girard L, Xie Y. Tumor mrna expression profiles predict responses to chemotherapy. J Clin Oncol 2007;25:4329–36.
8. Xie Y, Minna JD. Predicting the future for people with lung cancer. Nat Med 2008;14:812–3.
9. Xie Y, Minna JD. Non-small-cell lung cancer mrna expression signature predicting response to adjuvant chemotherapy. J Clin Oncol 2010;28:4404–7.
10. Xie Y, Minna JD. A lung cancer molecular prognostic test ready for prime time. Lancet 2012;379:785–7.
11. 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.
12. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. J Clin Oncol 2005;23:2020–7.
13. Simon R, Wang SJ. Use of genomic signatures in therapeutics development in oncology and other diseases. Pharmacogenomics J 2006; 6:166–73.
14. Wang SJ. Biomarker as a classifier in pharmacogenomics clinical trials: A tribute to 30th anniversary of psi. Pharm Stat 2007;6:283–96.
15. Breiman L. Random forests. Machine Learning 2001;45:5–32.
16. Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. BMC Bioinformatics 2006;7:3.
17. Wu B, Abbott T, Fishman D, McMurray W, Mor G, Stone K, et al. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. Bioinformatics 2003;19:1636–43.
18. Huang X, Pan W, Han X, Chen Y, Miller LW, Hall J. Borrowing information from relevant microarray studies for sample classification using weighted partial least squares. Comput Biol Chem 2005;29:204–11.
19. Zhang Z, Chen D, Fenstermacher DA. Integrated analysis of independent gene expression microarray datasets improves the predictability of breast cancer outcome. BMC Genomics 2007;8:331.
20. Wolpert DH. Stacked generalization. Neural Networks 1992;5:241–59.

21. Pan W, Xiao G, Huang X. Using input dependent weights for model combination and model selection with multiple sources of data. Statistica Sinica 2006;16:523–40.
22. Tsuboi M, Ohira T, Saji H, Miyajima K, Kajiwara N, Uchida O, et al. The present status of postoperative adjuvant chemotherapy for completely resected non-small cell lung cancer. Ann Thorac Cardiovasc Surg 2007;13:73–7.
23. Herbst RS, Maddox AM, Rothenberg ML, Small EJ, Rubin EH, Baselga J, et al. Selective oral epidermal growth factor receptor tyrosine kinase inhibitor zd1839 is generally well-tolerated and has activity in non-small-cell lung cancer and other solid tumors: Results of a phase i trial. J Clin Oncol 2002;20:3815–25.
24. Baselga J, Rischin D, Ranson M, Calvert H, Raymond E, Kieback DG, et al. Phase i safety, pharmacokinetic, and pharmacodynamic trial of zd1839, a selective oral epidermal growth factor receptor tyrosine kinase inhibitor, in patients with five selected solid tumor types. J Clin Oncol 2002;20:4292–302.
25. Nakagawa K, Tamura T, Negoro S, Kudoh S, Yamamoto N, Takeda K, et al. Phase i pharmacokinetic trial of the selective oral epidermal growth factor receptor tyrosine kinase inhibitor gefitinib ('iressa', zd1839) in japanese patients with solid malignant tumors. Ann Oncol 2003;14:922–30.
26. Ranson M, Hammond LA, Ferry D, Kris M, Tullo A, Murray PI, et al. Zd1839, a selective oral epidermal growth factor receptor-tyrosine kinase inhibitor, is well tolerated and active in patients with solid, malignant tumors: Results of a phase i trial. J Clin Oncol 2002;20:2240–50.
27. van't Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. Nature 2008;452:564–70.
28. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185–93.
29. Minna JD, Gazdar AF, Sprang SR, Herz J. Cancer. A bull's eye for targeted lung cancer therapy. Science 2004;304:1458–61.
30. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. Science 2004;304:1497–500.
31. Garcia JA. Hifing the brakes: Therapeutic opportunities for treatment of human malignancies. Sci STKE 2006;2006:p25.
32. Dudek AZ, Mahaseth H. Circulating angiogenic cytokines in patients with advanced non-small cell lung cancer: Correlation with treatment response and survival. Cancer Invest 2005;23:193–200.
33. Kristiansen G, Schluns K, Yongwei Y, Denkert C, Dietel M, Petersen I. Cd24 is an independent prognostic marker of survival in nonsmall cell lung cancer patients. Br J Cancer 2003;88:231–6.
34. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. Comput Syst Sci 1997;55:119–39.
35. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York: Springer-Verlag; 2001.
36. McShane L, Cavenagh M, Lively T, Eberhard D, Bigbee W, Williams P, et al. Criteria for the use of omics-based predictors in clinical trials: Explanation and elaboration. BMC Medicine 2013;11:220.
37. McShane LM, Cavenagh MM, Lively TG, Eberhard DA, Bigbee WL, Williams PM, et al. Criteria for the use of omics-based predictors in clinical trials. Nature 2013;502:317–20.