Validation of Predictive Classifiers

Predictive Biomarker Classifiers

- In most positive clinical trials, only a small proportion of the eligible population benefits from the new rx
- Many chronic diseases are biologically heterogeneous and molecularly targeted treatments are likely to benefit only a subset
- Traditional subset analysis is error prone and a better paradigm is needed

An alternative Paradigm

- Test the strong null hypothesis that the new treatment E is uniformly ineffective relative to control C in a manner that preserves the type I error of the trial
- If the strong null is rejected, develop an internally validated labeling indication classifier for informing physicians in their decisions about which patients they treat with E

Adaptive Signature Design: An Adaptive Clinical Trial Design for Generating and Prospectively Testing A Gene Expression Signature for Sensitive Patients

Boris Freidlin and Richard Simon

Abstract

Purpose: A new generation of molecularly targeted agents is entering the definitive stage of clinical evaluation. Many of these drugs benefit only a subset of treated patients and may be overlooked by the traditional, broad eligibility approach to randomized clinical trials. Thus, there is a need for development of novel statistical methodology for rapid evaluation of these agents.

Experimental Design: We propose a new adaptive design for randomized clinical trials of targeted agents in settings where an assay or signature that identifies sensitive patients is not available at the outset of the study. The design combines prospective development of a gene expression–based classifier to select sensitive patients with a properly powered test for overall effect.

Results: Performance of the adaptive design, relative to the more traditional design, is evaluated in a simulation study. It is shown that when the proportion of patients sensitive to the new drug is low, the adaptive design substantially reduces the chance of false rejection of effective new treatments. When the new treatment is broadly effective, the adaptive design has power to detect the overall effect similar to the traditional design. Formulas are provided to determine the situations in which the new design is advantageous.

Conclusion: Development of a gene expression–based classifier to identify the subset of sensitive patients can be prospectively incorporated into a randomized phase III design without compromising the ability to detect an overall effect.

Developments in tumor biology have resulted in shift toward molecularly targeted drugs (1–3). Most human tumor types are heterogeneous with regard to molecular pathogenesis, genomic signatures, and phenotypic properties. As a result, only a subset of the patients with a given cancer is likely to benefit from a targeted agent (4). This complicates all stages of clinical development, especially randomized phase III trials (5, 6). In some cases, predictive assays that can accurately identify patients who are likely to benefit from the new therapy have been developed. Then, targeted randomized designs that restrict eligibility to patients with sensitive tumors should be used (7). However, reliable assays to select sensitive patients are often not available (8, 9). Consequently, traditional randomized clinical trails with broad eligibility criteria are routinely used to evaluate such agents. This is generally inefficient and may lead to missing effective agents.

Authors' Affiliation: Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, Maryland

Received 3/18/05; revised 7/15/05; accepted 8/4/05.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Boris Freidlin, Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, 6130 Executive Boulevard, EPN 8122, MSC 7434, Bethesda, MD 20892-7434. Phone: 301-402-0640; Fax: 301-402-0560; E-mail: freidlib@dcp.nci.nih.gov.

doi:10.1158/1078-0432.CCR-05-0605

Genomic technologies, such as microarrays and single nucleotide polymorphism genotyping, are powerful tools that hold a great potential for identifying patients who are likely to benefit from a targeted agent (10, 11). However, due to the large number of genes available for analysis, interpretation of these data is complicated. Separation of reliable evidence from the random patterns inherent in high-dimensional data requires specialized statistical methodology that is prospectively incorporated in the trial design. Practical implementation of such designs has been lagging. In particular, analysis of microarray data from phase III randomized studies is usually considered secondary to the primary overall comparison of all eligible patients. Many analyses are not explicitly written into protocols and done retrospectively, mainly as "hypothesis-generating" tools.

We propose a new adaptive design for randomized clinical trials of molecularly targeted agents in settings where an assay or signature that identifies sensitive patients is not available. Our approach includes three components: (*a*) a statistically valid identification, based on the first stage of the trial, of the subset of patients who are most likely to benefit from the new agent; (*b*) a properly powered test of overall treatment effect at the end of the trial using all randomized patients; and (*c*) a test of treatment effect for the subset identified in the first stage, but using only patients randomized in the remainder of the trial. The components are prospectively incorporated into a single phase III randomized clinical trial with the overall false-positive error rate controlled at a prespecified level.

Developing & Validating Predictive Classifiers - Treatment vs Control

- Not testing treatment effect in multiple subsets defined by single covariates
- Not testing statistical significance of treatment by covariate interaction effects in a regression model

- Development and internal validation of predictive classifiers can also be conducted retrospectively on phase III trials to identify treatment hypotheses to be prospectively tested

Predictive Classifier Paradigm

- Suppose we have analyzed a training set T of data from an RCT and developed a predictive classifier C(x;T) ->{0,1} that indicates whether a patient with covariate vector x should be treated with C (0) or E (1)
- There are many ways of developing C

$$\log it(p_E) = \log \frac{p_E}{1 - p_E} = \alpha_E + \beta_E^{'} x$$

$$\log it(p_C) = \log \frac{p_C}{1 - p_C} = \alpha_C + \beta_C^{'} x$$

Prefer E if $p_E \geq p_C + \varepsilon$

i.e. $\alpha_E + \beta_E^{'} x \geq \alpha_C + \beta_C^{'} x + \varepsilon$

Prefer E if $\hat{p}_E \geq \hat{p}_C + \varepsilon$

i.e. $\hat{\alpha}_E + \hat{\beta}_E' x \geq \hat{\alpha}_C + \hat{\beta}_C' x + \varepsilon$

$$\lambda(t,z,x) = \lambda_0(t)\exp\left\{\alpha z + z\beta_E^{'}x + (1-z)\beta_C^{'}x\right\}$$

where z is (0,1) treatment indicator.

$$\lambda(t,z=1,x) = \lambda_0(t)\exp\left\{\alpha + z\beta_E^{'}x\right\}$$

$$\lambda(t,z=0,x) = \lambda_0(t)\exp\left\{\beta_C^{'}x\right\}$$

$$\log\frac{\lambda(t,z=1,x)}{\lambda(t,z=0,x)} = \alpha + \beta_E^{'}x - \beta_C^{'}x$$

Fit penalized PH model

Assign E if $\hat{\alpha} + \hat{\beta}_E^{'}x - \hat{\beta}_C^{'}x \le \varepsilon$.

- If we had an independent RCT, we could compare E vs C in
  - the subset S_+ of patients for which C(x;T)=1
  - the subset S_- of patients for which C(x;T)=0
  - We expect a greater treatment effect in the former subset of patients predicted to be "sensitive" to E.
- We could do significance tests of treatment effect in each of the two subsets, and we could estimate the size of the treatment effects.

- Using cross-validation or bootstrap re-sampling we can approximate

  The treatment effect in the intend-to-use population S_+

- We can also test the null hypothesis that no subset of the ITT population has a positive treatment effect

Cancer Therapy: Clinical

Clinical Cancer Research

The Cross-Validated Adaptive Signature Design

Boris Freidlin[1], Wenyu Jiang[2], and Richard Simon[1]

Abstract

Purpose: Many anticancer therapies benefit only a subset of treated patients and may be overlooked by the traditional broad eligibility approach to design phase III clinical trials. New biotechnologies such as microarrays can be used to identify the patients that are most likely to benefit from anticancer therapies. However, due to the high-dimensional nature of the genomic data, developing a reliable classifier by the time the definitive phase III trail is designed may not be feasible.

Experimental Design: Previously, Freidlin and Simon (*Clinical Cancer Research*, 2005) introduced the adaptive signature design that combines a prospective development of a sensitive patient classifier and a properly powered test for overall effect in a single pivotal trial. In this article, we propose a cross-validation extension of the adaptive signature design that optimizes the efficiency of both the classifier development and the validation components of the design.

Results: The new design is evaluated through simulations and is applied to data from a randomized breast cancer trial.

Conclusion: The cross-validation approach is shown to considerably improve the performance of the adaptive signature design. We also describe approaches to the estimation of the treatment effect for the identified sensitive subpopulation. *Clin Cancer Res; 16(2); 691–8. ©2010 AACR.*

Due to the molecular heterogeneity of most human cancers, only a subset of treated patients benefit from a given therapy. This is particularly relevant for the new generation of anticancer agents that target specific molecular pathways (1–3). Genomic (or proteinomic) technologies such as microarrays provide powerful tools for identifying a genetic signature (diagnostic test) for patients who are most likely to benefit from a targeted agent. Ideally, such diagnostic test should be developed and validated before commencing the definitive phase III trial (4). However, due to the complexity of signaling pathways and the large number of genes available for analysis, the development of a reliable diagnostic classifier using early nonrandomized phase II data is often not feasible. Conducting a phase III randomized clinical trial (RCT) requires considerable time and resources. Therefore, clinical trial designs that allow combining the definitive evaluation of a new agent with the development of the companion diagnostic test can considerably speed up the introduction of new cancer therapies.

Previously, the adaptive signature design (ASD) has been proposed for settings where a signature to identify sensitive patients is not available (5). The design combines the prospective development of a pharmacogenomic diagnostic test (signature) to select sensitive patients with a properly powered test for overall effect. It was shown that when the proportion of patients sensitive to the new drug is low, the ASD substantially reduces the chance of false rejection of effective new treatments. When the new treatment is broadly effective, the power of the adaptive design to detect the overall effect is similar to that of the traditional design.

The signature component of the ASD carries out signature development and validation on the mutually exclusive subgroups of patients (e.g., half of the study population is used to develop a signature and another half to validate it). Although the conceptual simplicity of this approach is appealing, it also limits its power as only half of the patients are used for signature development and half for validation. This is especially relevant in the present setting because (*a*) signature development in high dimensional data requires large sample sizes, and (*b*) when the fraction of sensitive patients is low, a large number of patients needs to be screened to identify the sufficient number of sensitive patients to achieve acceptable power.

In this article, we describe an extension of the ASD in which signature development and validation are embedded in a complete cross-validation procedure. This allows the use of virtually the entire study population in both signature development and validation steps. We develop a procedure that preserves the study-wise type I error while substantially increasing the statistical power for establishing a statistically significant treatment effect for an identified subset of patients who benefit from the experimental treatment. We also examine approaches to estimation of treatment effect for the identified sensitive subset.

Authors' Affiliations: [1]Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, Maryland and [2]Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada

Corresponding Author: Boris Freidlin, Biometric Research Branch, EPN-8122, National Cancer Institute, Bethesda, MD 20892. Phone: 301-402-0640; Fax: 301-402-0560; E-mail: freidlinb@ctep.nci.nih.gov.

doi: 10.1158/1078-0432.CCR-09-1357

- At the conclusion of the trial randomly partition the patients into K approximately equally sized sets P_1 , … , P_K
- Let D_{-i} denote the full dataset minus data for patients in P_i
- Omit patients in P_1
- Apply the defined algorithm to analyze the data in D_{-1} to obtain a classifier M_{-1}
- Classify each patient j in P_1 using model M_{-1}
- Record the treatment recommendation E or C

- Repeat the above steps for all K loops of the cross-validation (develop classifier from scratch in each loop and classify omitted patients)
- When cross-validation is completed, all patients have been classified once as what their optimal treatment is predicted to be

- Let S denote the set of patients for whom treatment E is predicted optimal
- Compare outcomes for patients in S who actually received E to those in S who actually received C
  - Compute Kaplan Meier curves of those receiving E and those receiving C
  - Let z = standardized log-rank statistic

- Compute statistical significance of z by randomly permuting treatment labels and repeating the entire cross-validation procedure to obtain a new set S' and a new logrank statistic z'
  - Do this 1000 or more times to generate the permutation null distribution of treatment effect for the patients in S

- The size of the E vs C treatment effect in the subset S estimates the treatment effect in the subset defined by applying the algorithm to the full dataset

$f(x,z) =$ expected outcome for a patient with covariate vector x receiving rx z

$C(;D) =$ predictive classifier trained on full dataset D

$S_+(D) = \{x : C(x,D) = 1\}$

$$\Delta\left(S_+(D)\right) = \int_{x \in S_+(D)} \left[f(x,1) - f(x,0)\right] dF\left(x \mid x \in S_+(D)\right)$$

Cross-validation estimates $E_T\left[\Delta\left(S_+(T)\right)\right]$

Published OnlineFirst August 27, 2012; DOI: 10.1158/1078-0432.CCR-12-1206

Predictive Biomarkers and Personalized Medicine

Clinical Cancer Research

Developing and Validating Continuous Genomic Signatures in Randomized Clinical Trials for Predictive Medicine

Shigeyuki Matsui[1], Richard Simon[2], Pingping Qu[3], John D. Shaughnessy Jr[4], Bart Barlogie[4], and John Crowley[3]

Abstract

Purpose: It is highly challenging to develop reliable diagnostic tests to predict patients' responsiveness to anticancer treatments on clinical endpoints before commencing the definitive phase III randomized trial. Development and validation of genomic signatures in the randomized trial can be a promising solution. Such signatures are required to predict quantitatively the underlying heterogeneity in the magnitude of treatment effects.

Experimental Design: We propose a framework for developing and validating genomic signatures in randomized trials. Codevelopment of predictive and prognostic signatures can allow prediction of patient-level survival curves as basic diagnostic tools for treating individual patients.

Results: We applied our framework to gene-expression microarray data from a large-scale randomized trial to determine whether the addition of thalidomide improves survival for patients with multiple myeloma. The results indicated that approximately half of the patients were responsive to thalidomide, and the average improvement in survival for the responsive patients was statistically significant. Cross-validated patient-level survival curves were developed to predict survival distributions of individual future patients as a function of whether or not they are treated with thalidomide and with regard to their baseline prognostic and predictive signature indices.

Conclusion: The proposed framework represents an important step toward reliable predictive medicine. It provides an internally validated mechanism for using randomized clinical trials to assess treatment efficacy for a patient population in a manner that takes into consideration the heterogeneity in patients' responsiveness to treatment. It also provides cross-validated patient-level survival curves that can be used for selecting treatments for future patients. *Clin Cancer Res; 18(21); 6065–73. ©2012 AACR.*

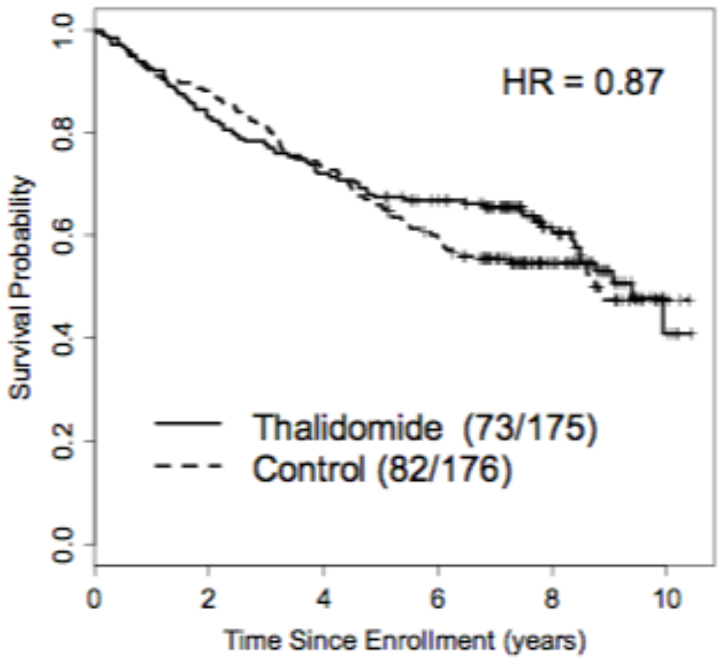


Fig. 1. Survival curves for all 351 patients with genomic data in the randomized trial for multiple myeloma.

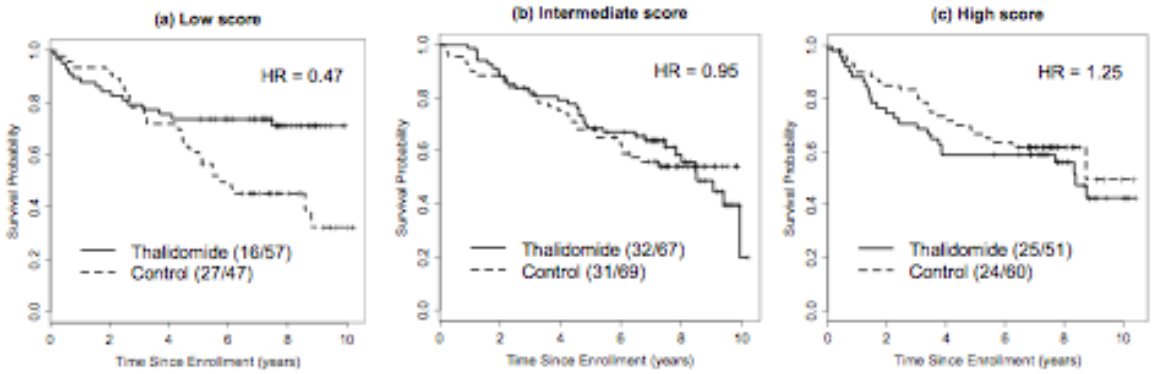
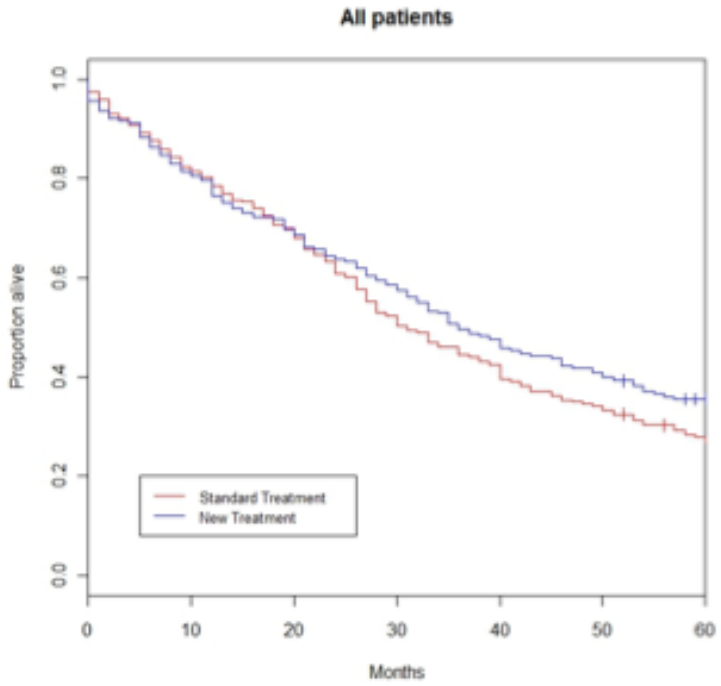
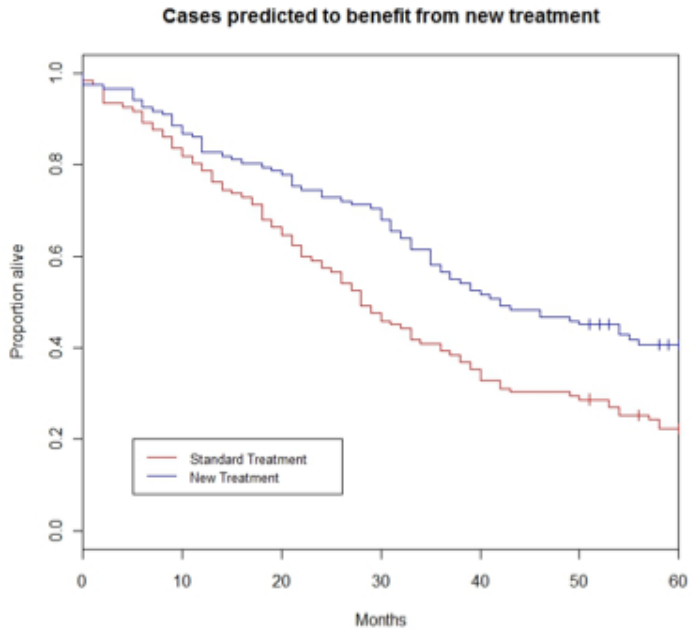


Fig. 5. Survival curves for each of the three subclasses, "Low", "Intermediate" and "High" derived from using thresholds of 33rd and 66th percentiles in the predicted signature score S (panels a-c).

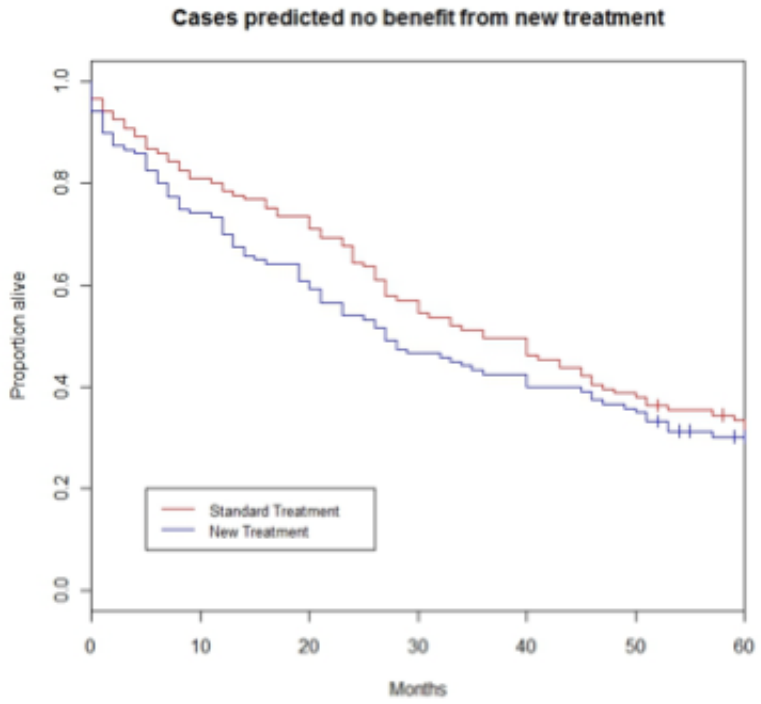
Overall analysis. Log-rank statistic is 2.9 with p=0.09



Cross-validated survival curves for patients predicted to benefit from new treatment. Log-rank = 10.0, p=0.002



Cross-validated survival curves for cases predicted not to benefit from the new treatment. Log-rank = 0.54



Performance Indices of Predictive Classifiers with Survival Data

- RCT of E vs C
- Survival endpoint
- Predictive binary classifier M
  - M=1 indicates patient should benefit from E
  - M=0 indicates patient should not

- Given two patients i and j with same x vector, one receives E (z_i=1) and the other receives C (z_j=0).
- sensitivity = Pr[M=1 | $S_i > S_j$]
- specificity = Pr[M=0 | $S_i < S_j$]
- ppv = Pr[$S_i > S_j$ | M=1]
- npv = Pr[$S_i < S_j$ | M=0]

- If the survival distributions have proportional hazards with δ_+ denoting the hazard ratio of C versus T (>1), estimated from the cross-validated set S_+ then

$$ppv = \frac{\delta_+}{1+\delta_+}$$

$$npv = \frac{1}{1+\delta_+}$$

$$sensitivity = \left\{1 + \frac{1-npv}{ppv}\frac{\Pr(M=0)}{\Pr(M=1)}\right\}^{-1}$$

$$specificity = \left\{1 + \frac{1-ppv}{npv}\frac{\Pr(M=1)}{\Pr(M=0)}\right\}^{-1}$$

Example
Amgen pannitumumab trial

- Treatment effect for patients with wild-type KRAS was 2.22 favoring panitumumab
- Treatment effect for patients with mutated KRAS was 1.01
- Prevalence of wild type KRAS (M=1) was 0.62
- ppv=0.69, npv=0.50
- Sensitivity = 0.62, specificity = 0.50