A Roadmap for Developing & Validating Genomic Classifiers for Treatment Selection

Richard Simon, D.Sc.
Chief, Biometric Research Branch
National Cancer Institute
http://linus.nci.nih.gov/brb

Simon R, Korn E, McShane L, Radmacher M, Wright G, Zhao Y. *Design and analysis of DNA microarray investigations*, Springer-Verlag, 2003.

Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. Journal of Computational Biology 9:505-511, 2002.

Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the analysis of DNA microarray data. Journal of the National Cancer Institute 95:14-18, 2003.

Dobbin K, Simon R. Comparison of microarray designs for class comparison and class discovery, Bioinformatics 18:1462-69, 2002; 19:803-810, 2003; 21:2430-37, 2005; 21:2803-4, 2005.

Dobbin K and Simon R. Sample size determination in microarray experiments for class comparison and prognostic classification. Biostatistics 6:27-38, 2005.

Dobbin K, Shih J, Simon R. Questions and answers on design of dual-label microarrays for identifying differentially expressed genes. Journal of the National Cancer Institute 95:1362-69, 2003.

Wright G, Simon R. A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics 19:2448-55, 2003.

Korn EL, Troendle JF, McShane LM, Simon R.Controlling the number of false discoveries. Journal of Statistical Planning and Inference 124:379-08, 2004.

Molinaro A, Simon R, Pfeiffer R. Prediction error estimation: A comparison of resampling methods. Bioinformatics 21:3301-7,2005.

Simon R. Using DNA microarrays for diagnostic and prognostic prediction. Expert Review of Molecular Diagnostics, 3(5) 587-595, 2003.

Simon R. Diagnostic and prognostic prediction using gene expression profiles in high dimensional microarray data. British Journal of Cancer 89:1599-1604, 2003.

Simon R and Maitnourim A. Evaluating the efficiency of targeted designs for randomized clinical trials. Clinical Cancer Research 10:6759-63, 2004.

Maitnourim A and Simon R. On the efficiency of targeted clinical trials. Statistics in Medicine 24:329-339, 2005.

Simon R. When is a genomic classifier ready for prime time? Nature Clinical Practice – Oncology 1:4-5, 2004.

Simon R. An agenda for Clinical Trials: clinical trials in the genomic era. Clinical Trials 1:468-470, 2004.

Simon R. Development and Validation of Therapeutically Relevant Multi-gene Biomarker Classifiers. Journal of the National Cancer Institute 97:866-867, 2005.

Simon R. A roadmap for developing and validating therapeutically relevant genomic classifiers. Journal of Clinical Oncology 23(29), 2005.

Freidlin B and Simon R. Adaptive signature design. Clinical Cancer Research (In Press).

Simon R. Validation of pharmacogenomic biomarker classifiers for treatment selection. Disease Markers (In Press).

Simon R. Guidelines for the design of clinical studies for development and validation of therapeutically relevant biomarkers and biomarker classification systems. In Biomarkers in Breast Cancer, Hayes DF and Gasparini G, Humana Press (In Press).

Genomic Targeting

- Enables patients to be treated with drugs that actually work for them
- Avoids false negative trials for heterogeneous populations
- Avoids erroneous generalizations of conclusions from positive trials
- Enables clinical benefit to be reliably identified more easily with smaller clinical trials

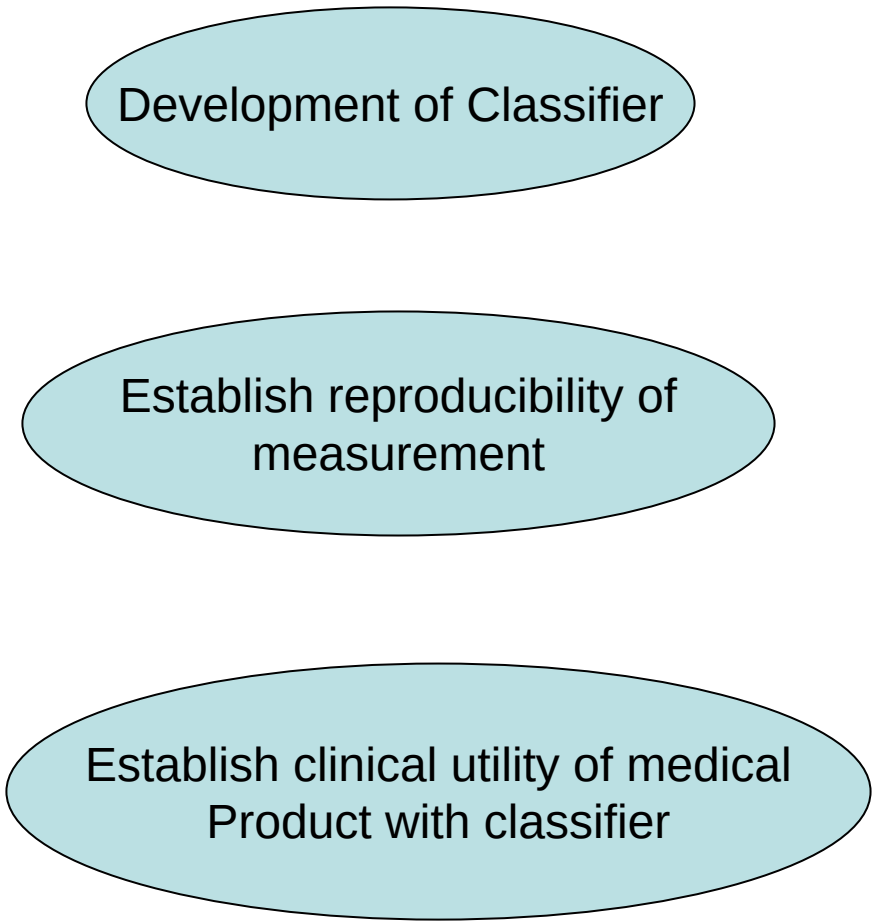
- Surrogate endpoint
  - A measurement made on a patient before, during and after treatment to determine whether the treatment is working
- Pharmacogenomic or treatment selection marker
  - A measurement made on a patient before treatment used to select treatment
- Biomarker
  - Any biological measurement made on a patient

Can of Worms

- Surrogate endpoints
- Validity of biomarkers
- Hypothesis formulation and testing on the same set of data
- Conducting pivotal clinical trials without clearly pre-planned analysis
- “Stratification”

The Roadmap

1. Develop a completely specified genomic classifier of the patients likely to benefit from a new medical product
2. Establish reproducibility of measurement of the classifier
3. Use the completely specified classifier to design and analyze a new clinical trial to evaluate effectiveness of the new treatment with a pre-defined analysis plan.


Development of Classifier
Establish reproducibility of measurement
Establish clinical utility of medical Product with classifier

Guiding Principle

- The data used to develop the classifier must be distinct from the data used to test hypotheses about treatment effect in subsets determined by the classifier
  - Developmental studies are exploratory
  - Studies on which treatment effectiveness claims are to be based should be definitive hypothesis testing studies based on completely pre-specified classifiers

A set of genes is not a classifier

- Gene selection
- Mathematical function for mapping from multivariate gene expression domain to prognostic or diagnostic classes
- Weights and other parameters including cut-off thresholds for risk scores

Linear Classifiers for Two Classes

$$l(\underline{x}) = \sum_{i \varepsilon G} w_i x_i$$

$\underline{x}$ = vector of expression measurements

G = genes included in model

w_i = weight for i'th gene

decision boundary $l(\underline{x})$ > or < d

Nearest Centroid Classifier

- In a training set, identify the features that distinguish the outcome classes.
- Select a pair-wise similarity measure that incorporates the selected features.
- Compute the centroid of the training set samples in each class.
- Classify a sample in the validation set as being in outcome class 1 or outcome class 2 based on which centroid it is most similar to.

Strategies for Development of Genomic Classifiers

- (a) Single gene or protein based on knowledge of therapeutic target. or
- (b) Empirically determined based on correlating gene expression or genotype to patient outcome after treatment.
- (a) During phase I/II development. or
- (b) After failed phase III trial using archived specimens
- There is no need for FDA to regulate methods of classifier “development”

Genomic Classifiers Used for Selecting and Stratifying Patients in Drug Development

- The components of the classifier should not have to be “valid disease biomarkers” in the FDA sense
- The FDA definitions are reasonable for biomarkers to be used as surrogate endpoints, but not for selecting patient populations

- “I don’t know what ‘clinical validation’ [of a biomarker] means. The first thing you have to do is define a purpose for the biomarker. Validation is all about demonstrating fitness for purpose.”
  - Dr. Stephen Williams, Pfizer

There Should Be No Requirement For

- Demonstrating that the classifier or any of its components are “validated biomarkers of disease status”
- Ensuring that the individual components of the classifier are correlated with patient outcome or effective for selecting patients for treatment
- Demonstrating that repeating the classifier development process on independent data results in the same classifier

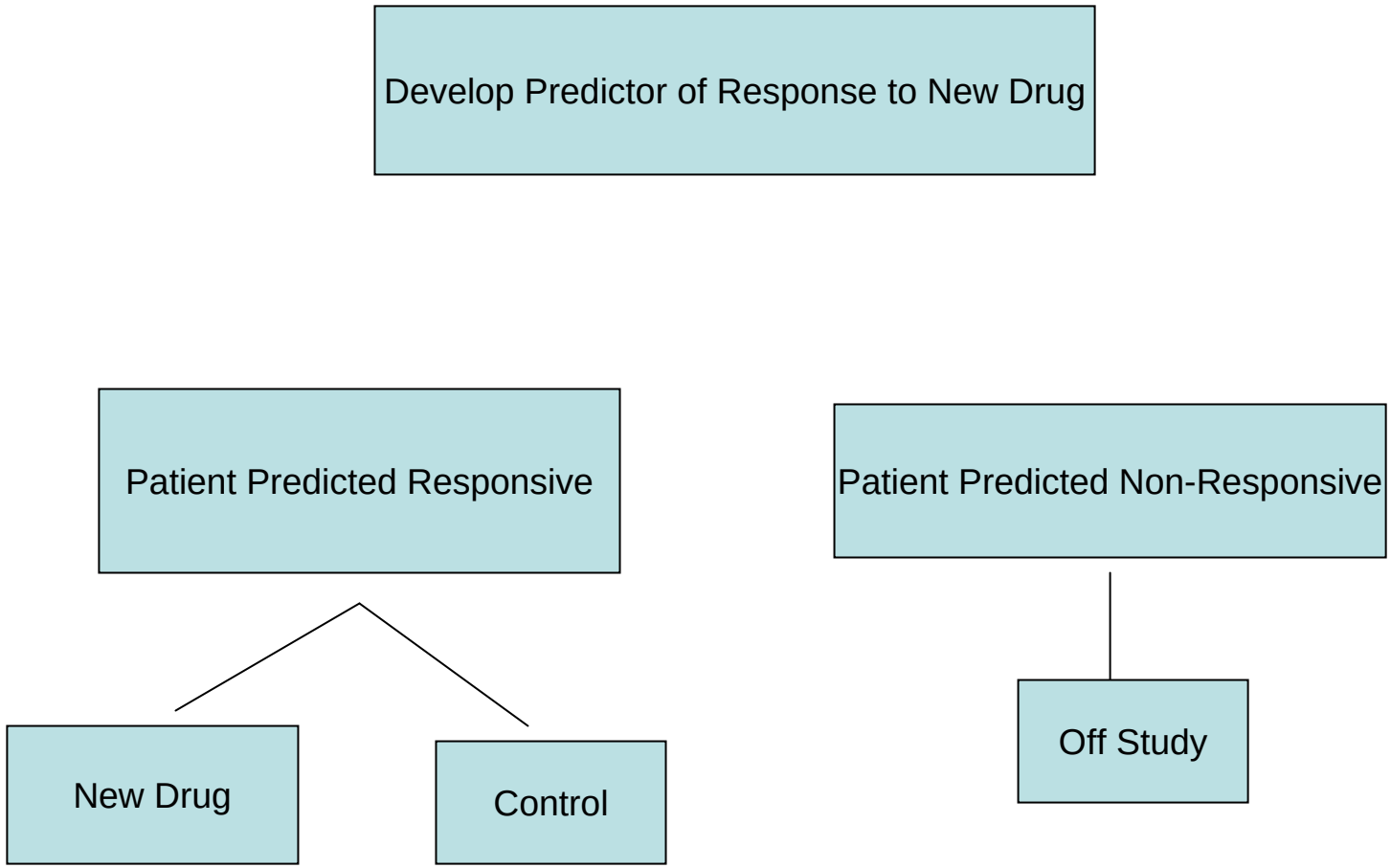
One Should Require That

- The classifier, as a whole, be reproducibly measurable

- The classifier as a whole, in conjunction with the medical product, has clinical utility

Using the Classifier in Evaluation of a New Therapeutic (I)

- Develop a diagnostic classifier that identifies the patients likely to benefit from the new drug
- Use the diagnostic to restrict eligibility to a prospectively planned evaluation of the new drug
- Demonstrate that the new drug is effective in the prospectively defined set of patients determined by the diagnostic
- Demonstrate that the diagnostic can be reproducibly measured


Develop Predictor of Response to New Drug
Patient Predicted Responsive
Patient Predicted Non-Responsive
New Drug
Control
Off Study

Randomized Clinical Trials Targeted to Patients Predicted to be Responsive to the New Treatment Can Be Much More Efficient than Traditional Untargeted Designs

- Simon R and Maitnourim A. Evaluating the efficiency of targeted designs for randomized clinical trials. Clinical Cancer Research 10:6759-63, 2004.
- Maitnourim A and Simon R. On the efficiency of targeted clinical trials. Statistics in Medicine 24:329-339, 2005.
- reprints at http://linus.nci.nih.gov/brb

Two Clinical Trial Designs

- Un-targeted design
  - Randomized comparison of E to C in unselected patients
- Targeted design
  - Classify patients based on probability of benefit from E
  - Randomize only patients likely to benefit

- Compare the two designs with regard to the number of patients required to achieve a fixed statistical power for detecting treatment effectiveness and the number of patients needed for screening

Comparison of Targeted to Untargeted Design

Simon R, Development and Validation of Biomarker Classifiers for Treatment Selection, JSPI

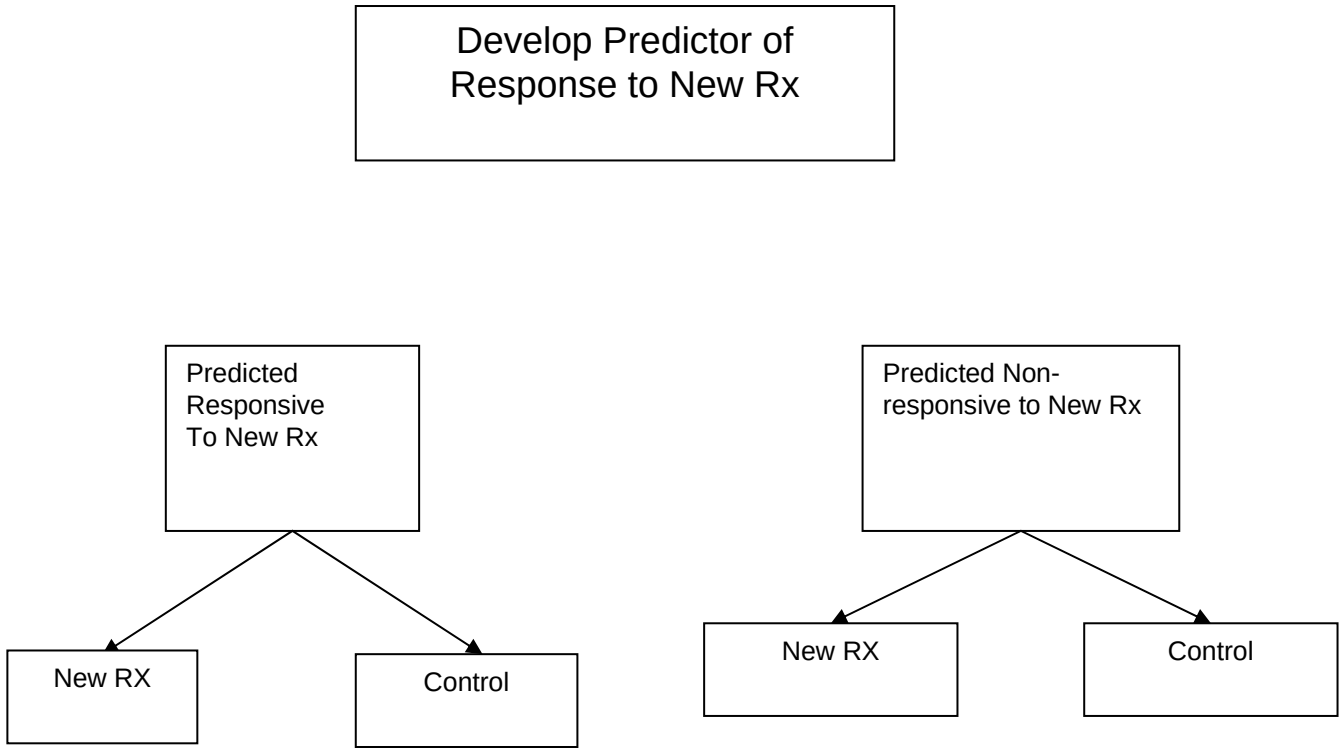
Treatment Hazard Ratio for Marker Positive Patients	Number of Events for Targeted Design	Number of Events for Traditional Design		
		Percent of Patients Marker Positive		
		20%	33%	50%
0.5	74	2040	720	316

- For Herceptin, even a relatively poor assay enabled conduct of a targeted phase III trial which was crucial for establishing effectiveness
- Recent results with Herceptin in early stage breast cancer show dramatic benefits for patients selected to express Her-2

Targeted Design

- Achieves critical path objectives, enabling clinical benefit to be easily seen in small clinical trials
- To fully achieve this potential
  - FDA must not insist on traditional large trials in unselected patients in order to demonstrate that the classifier is “necessary”
  - Sponsors must expand the size and intensity of development of genomic classifiers in early clinical development

Using the Classifier in Evaluation of a New Therapeutic (II)



Using Genomics in Development of a New Therapeutic (II)

- Develop a diagnostic classifier that identifies the patients likely to benefit from the new drug
- Do not use the diagnostic to restrict eligibility, but to structure a prospectively planned analysis strategy of a randomized trial of the new drug.
- Compare the new drug to the control overall for all patients ignoring the classifier.
  - If the treatment effect on the primary pre-specified endpoint is significant at the 0.04 level, then claim effectiveness for the eligible population as a whole.
- If the overall test is not significant at the 0.04 level, then perform a single subset analysis evaluating the new drug in the classifier + patients.
  - If the treatment effect is significant at the 0.01 level, then claim effectiveness for the classifier + patients.
- Demonstrate that the diagnostic can be reproducibly measured

Adaptive Signature Design

An adaptive design for generating and prospectively testing a gene expression signature for sensitive patients

Boris Freidlin and Richard Simon

Clinical Cancer Research (In Press)

Adaptive Signature Design
End of Trial Analysis

- Compare E to C for **all patients** at significance level 0.04
  - If overall H_0 is rejected, then claim effectiveness of E for eligible patients
  - Otherwise

- Otherwise:
  - Using only the first half of patients accrued during the trial, develop a binary classifier that predicts the subset of patients most likely to benefit from the new treatment E compared to control C
  - Compare E to C for patients accrued in second stage who are predicted responsive to E based on classifier
    - Perform test at significance level 0.01
    - If H_0 is rejected, claim effectiveness of E for subset defined by classifier

Treatment effect restricted to subset.
10% of patients sensitive, 10 sensitivity genes, 10,000 genes, 400 patients.

Test	Power
Overall .05 level test	46.7
Overall .04 level test	43.1
Sensitive subset .01 level test (performed only when overall .04 level test is negative)	42.2
Overall adaptive signature design	85.3

Overall treatment effect, no subset effect.
10,000 genes, 400 patients.

Test	Power
Overall .05 level test	74.2
Overall .04 level test	70.9
Sensitive subset .01 level test	1.0
Overall adaptive signature design	70.9

Conclusions

- New technology and biological knowledge make is increasingly feasible to identify which patients are most likely to benefit or suffer severe adverse events from a new treatment
- Targeting treatment can greatly improve the therapeutic ratio of benefit to adverse effects
  - Smaller clinical trials needed
  - Treated patients benefit
  - Economic benefit for society

Conclusions

- Effectively defining and utilizing genomic classifiers in drug development offers multiple challenges
- Much of the conventional wisdom about how to develop and utilize biomarkers is flawed and does not lead to definitive evidence of treatment benefit for a well defined population
- Some aspects of the guidelines of the FDA on co-development are inappropriate for treatment selection biomarkers and are not consistent with the critical path objectives

Conclusions

- With careful prospective planning, genomic classifiers can be used in a manner that provides definitive evidence of treatment effect
  - Trial designs are available that will support broad labeling indications in cases where drug activity is sufficient, and the opportunity to obtain strong evidence of effectiveness in a well defined subset where overall effectiveness is not established

Conclusions

- Prospectively specified analysis plans for phase III data are essential to achieve reliable results
  - Biomarker analysis does not mean exploratory analysis except in developmental studies
  - Biomarker classifiers used in phase III evaluations should be completely specified based on external data
- In some cases, definitive evidence can be achieved from prospective analysis of patients in previously conducted clinical trials with extensive archival of pre-treatment specimens