A Roadmap for the Development of Targeted Therapeutics in the Genomic Era

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 11:7872-8, 2005.

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, pp 3-15, 2005.

Simon R and Wang SJ. Use of genomic signatures in therapeutics development in oncology and other diseases. The Pharmacogenomics Journal, 2006.

Objectives

- How to use biological measurements to enhance the development and utilization of effective therapeutics
- What should constitute “validation” for such biological measurements
- Clarify misleading terminology
- Challenge erroneous conventional wisdom
- Transit from generalities to identifying specific and practical designs and analysis plans that sponsors can use

Biological Measurements

- Surrogate endpoint
  - A measurement made on a patient before, during and after treatment to determine whether the treatment is working
- Prognostic factor
  - A measurement made before treatment that correlates with outcome, usually for a heterogeneous set of patients
- Predictive factors
  - A measurement made before treatment to predict whether a particular treatment is likely to be beneficial

Biomarker

- Any biological measurement made on a patient

- “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

Surrogate Endpoints

- Intermediate endpoints are useful for phase I and phase II studies.
  - They don’t need to be “validated” surrogates for this purpose
- It is often more difficult to properly “validate” an endpoint as a surrogate than to use the clinical endpoint in phase III trials

Validity of a Surrogate Endpoint

- Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. Statistics in Medicine 8:431-40, 1989.

- Test of the null hypothesis of no effect on surrogate outcome S between treatment and control is a valid test of the null hypothesis of no effect on true outcome T between treatment and control.
  - (i) S fully captures the effect of treatment on T
  - (ii) S is informative about T

- “One rarely can establish that surrogate endpoints are valid. Even in that rare setting in which data on treatment Z would allow one to view S as a valid surrogate for T, one cannot extrapolate this surrogacy to any new treatment Z* that could have mechanisms of action that differ from those of Z.”
  - Fleming TR, Statistical Science 7:428-56, 1992

Partial Surrogate Endpoint

- Improvement of a partial surrogate endpoint is a necessary but not sufficient condition for improvement of the clinical endpoint
- When to hold and when to fold
- Partial surrogates are used for phase II trials
- Partial surrogates can be used for early termination of phase III trials. The trial should continue accrual and follow-up to evaluate true endpoint if treatment effect on partial surrogate is sufficient.

Prognostic Factors

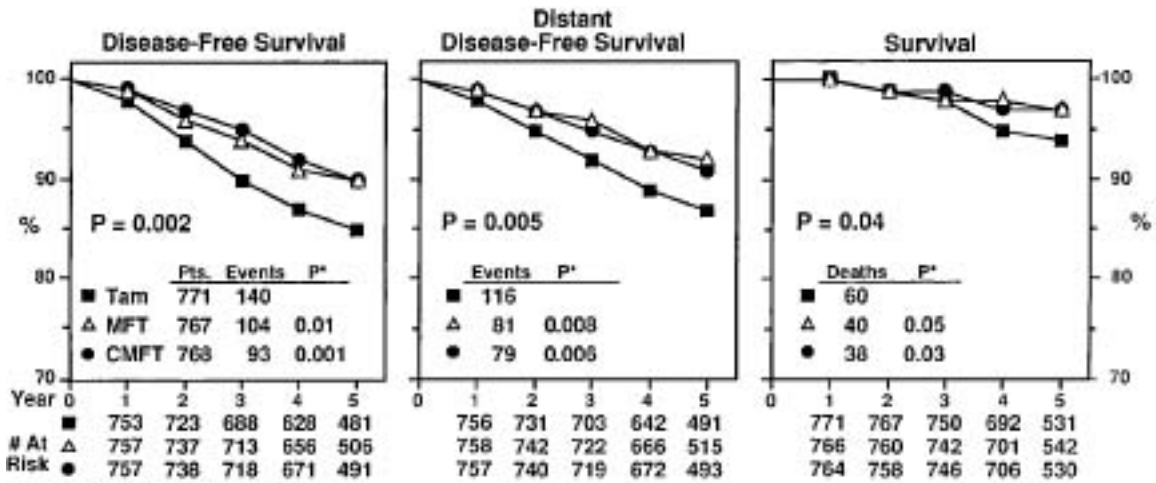
- Many prognostic factor studies utilize “convenience samples” of patients that are heterogeneous with regard to disease extent and treatment. The results are often not useful for therapeutic decision making.
- Non-therapeutically relevant prognostic factors are often forgotten or serve to drive up medical costs by fueling defensive medical practice

Conventional Wisdom

- Clinical trials should have broad eligibility criteria
- Never believe subset analyses
- This approach needs re-examination in era when there is increasing evidence that many diseases are heterogeneous in pathogenesis and sensitivity to treatment

Broad Eligibility May Cause

- Large clinical trials that fail to establish the effectiveness of useful drugs
- Inconsistency in results among different trials
- Post-approval treatment of many patients who don’t benefit



* Comparison to Tamoxifen

	RELATIVE RISK (95% CONFIDENCE INTERVAL)		
MFT/Tam	0.72 (0.56-0.93)	0.68 (0.51-0.90)	0.67 (0.45-0.99)
CMFT/Tam	0.65 (0.50-0.84)	0.67 (0.50-0.89)	0.64 (0.42-0.95)

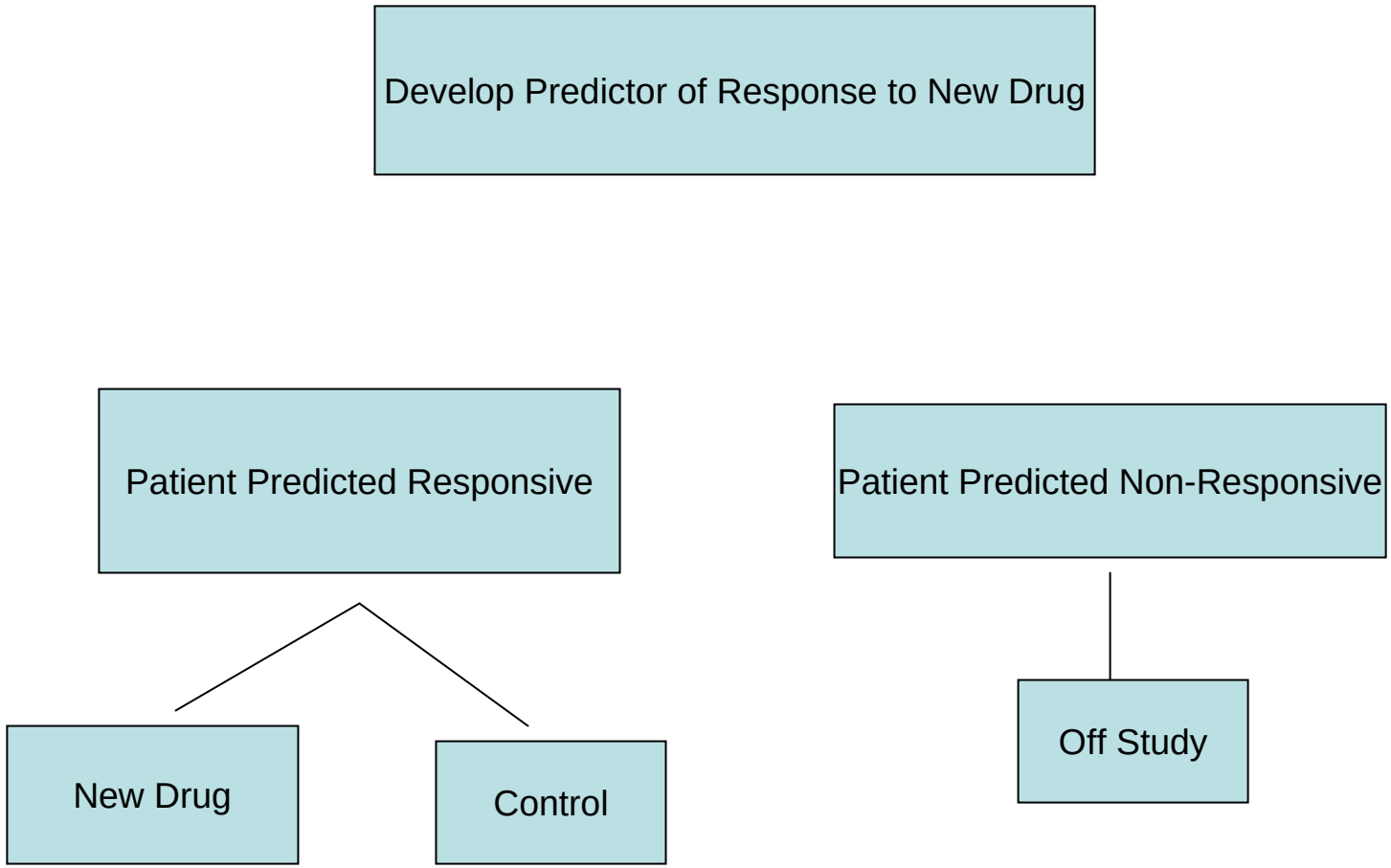
- Cancer clinical trials of molecularly targeted agents may benefit a relatively small population of patients with a given primary site/stage of disease
  - Iressa
  - Herceptin
- The benefit for the sensitive subset may be very substantial

- Targeted clinical trials can be much more efficient than untargeted clinical trials, if we know who to target
- Targeted, not enriched

- Adoption of a classifier to restrict the use of a treatment in wide use should be based on demonstrating that use of the classifier leads to better clinical outcome
- In new drug development, the role of a classifier is to select a target population for treatment
  - The focus should be on evaluating the new drug, not on validating the classifier

Developmental Strategy (I)

- **Develop** a diagnostic classifier that identifies the patients likely to benefit from the new drug
- Develop a reproducible assay for the classifier
- **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


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

Evaluating the Efficiency of Strategy (I)

- 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

- Compare the two targeted design to the standard untargeted design 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
0.67	200	5200	1878	820

Randomized Ratio
(normal approximation)

- RandRat = $n_{untargeted}/n_{targeted}$

$$RandRat \approx \left(\frac{\delta_1}{\lambda\delta_1 + (1-\lambda)\delta_0} \right)^2$$

- δ_1= rx effect in marker + patients
- δ_0= rx effect in marker - patients
- λ=proportion of marker + patients
- If δ_0=0, RandRat = $1/\lambda^2$
- If $\delta_0 = \delta_1/2$, RandRat = $4/(\lambda+1)^2$

Randomized Ratio $n_{untargeted}/n_{targeted}$

λ Assay+	δ_0=0	$\delta_0 = \delta_1/2$
0.75	1.78	1.31
0.5	4	1.78
0.25	16	2.56

Screened Ratio

λ Assay+	δ_0=0	$\delta_0 = \delta_1/2$
0.75	1.33	0.98
0.5	2	0.89
0.25	4	0.64

- 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

One Should Require That

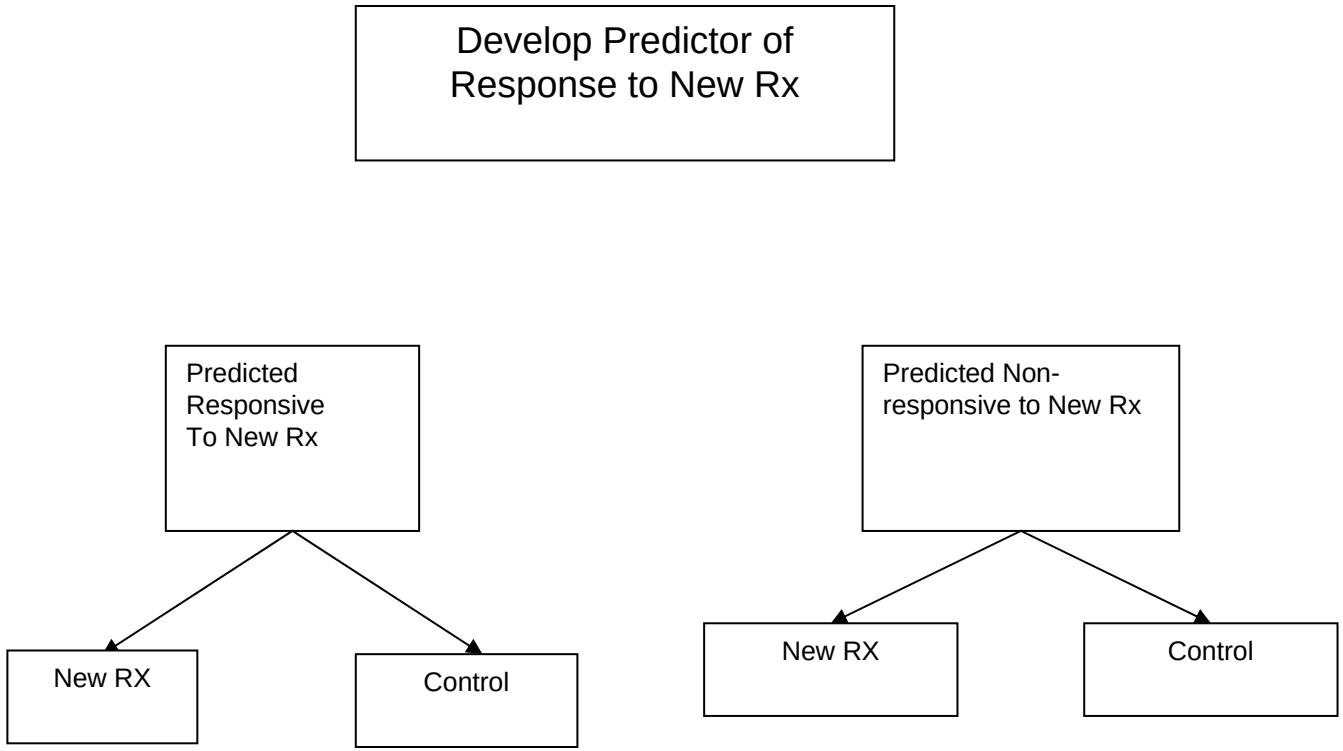
- The classifier, as a whole, be reproducibly measurable

- As a whole, the classifier in conjunction with the new drug has clinical utility

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 selection of the same components (genes)

Developmental Strategy (II)



Developmental Strategy (II)

- Do not use the diagnostic to restrict eligibility, but to structure a prospective analysis plan.
- Compare the new drug to the control overall for all patients ignoring the classifier.
  - If $p_{overall} \leq 0.04$ claim effectiveness for the eligible population as a whole
- Otherwise perform a single subset analysis evaluating the new drug in the classifier + patients
  - If $p_{subset} \leq 0.01$ claim effectiveness for the classifier + patients.

Sample Size Planning for Developmental Strategy (II)

- For overall test at 0.04 level
- For subset test at 0.01 level
- For overall test at 0.04 level and continue accrual to subset if overall test not significant

Key Features of Design (II)

- The purpose of the RCT is to evaluate treatment T vs C overall and for the pre-defined subset; not to re-evaluate the components of the classifier, or to modify or refine the classifier

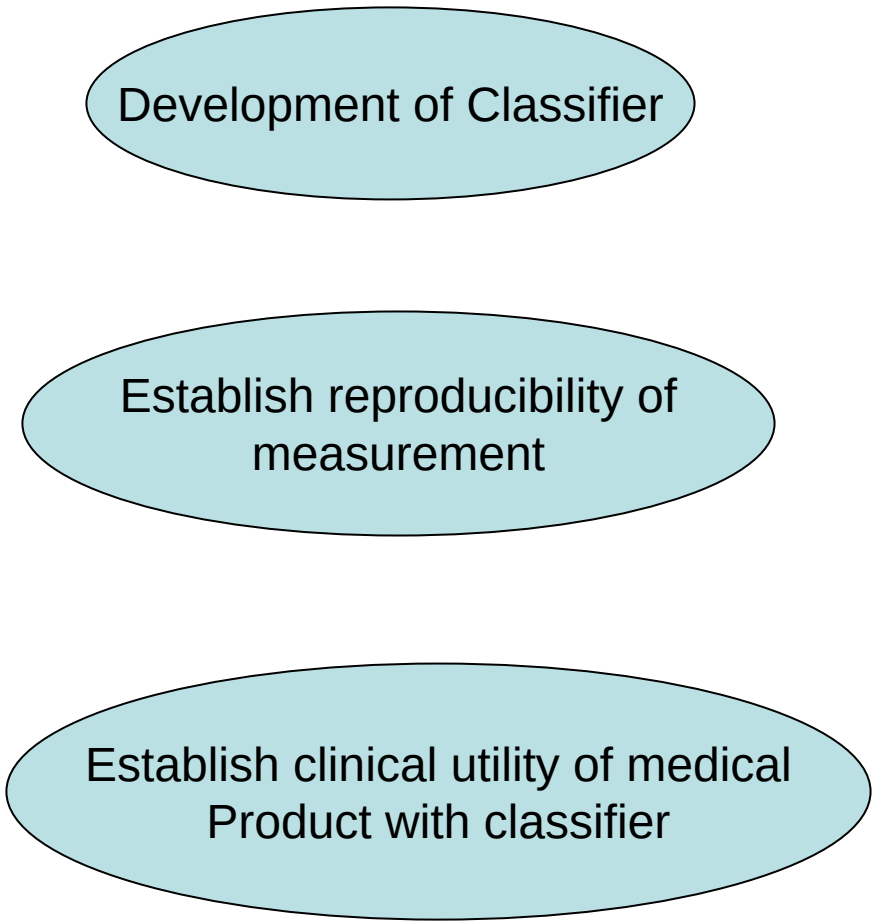
- There will be opportunity to examine whether the treatment is effective in classifier negative patients.
- In some cases there will be strong biological justification for testing T vs C only in classifier positive patients.

Key Features of Design (II)

- Pre-specified analysis plan
- Single pre-defined subset
- Overall study type I error of 0.05 is split between overall test and subset test
- Saying that the study should be “stratified” is not sufficient
  - It doesn’t matter whether randomization is stratified except that it helps ensure that all patients have specimens available to assay for classification

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 studies that test a treatment hypothesis in a patient population completely pre-specified by the classifier

Use of Archived Samples

- Archived samples from a conventional non-targeted “negative” clinical trial can be used to define a binary classifier of a subset thought to benefit from treatment T.
- That subset hypothesis should be tested in a separate clinical trial
  - Prospective targeted type (I) trial
  - Prospective type (II) trial
  - Analysis of archived specimens from a second previously conducted clinical trial to identify classifier positive patients

Development of Genomic Classifiers

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

Use of DNA Microarray Expression Profiling

- For settings where you don’t know how to identify the patients likely to be responsive to the new treatment based on its mechanism of action
- Only pre-treatment specimens are needed
- Expression profiling should be used to identify informative genes and form a binary classifier that can be used to select patients for study of for a pre-defined subset analysis

A set of genes is not a classifier

- Gene selection
- Mathematical function for combining expression levels of different genes to predict prognostic or diagnostic classes
- Weights and other parameters including cut-off thresholds for risk scores

There Should Be No Requirement For

- Demonstrating that the classifier or any of its components are “validated biomarkers of disease status”
- Demonstrating that repeating the classifier development process on independent data results in the same classifier
- FDA regulation of how DNA microarrays are used for classifier development

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 11:7872-8, 2005

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 it increasingly feasible to identify which patients are most likely to benefit or suffer severe adverse events from a new treatment
- FDA can either expedite or slow effective utilization of this technology
  - Over-regulating classifier development
  - Not providing sponsors with a clear and practical roadmap of what is required
- 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

- 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 biomarkers are inappropriate for treatment selection biomarkers

Conclusions

- Technology is sufficiently mature today to effectively identify which patients benefit from new treatments and to dramatically improve the efficiency of clinical trials
- What is lacking today is leadership in establishing specific guidelines for the design and analysis of adequate clinical trials that test new treatments in patient populations pre-defined based on completely specified diagnostic classifiers
- Trial designs are available that will support broad labeling indications in cases where drug activity is sufficient, and provide strong evidence of effectiveness for a prospectively defined subset where appropriate

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

Collaborators

- Boris Freidlin
- Aboubakar Maitournam
- Sue-Jane Wang