Roadblocks and Roadmaps to Predictive Oncology

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

BRB Website
http://linus.nci.nih.gov/brb

- Powerpoint presentations and audio files
- Reprints & Technical Reports
- BRB-ArrayTools software
- BRB-ArrayTools Data Archive
  - 100+ published cancer gene expression datasets with clinical annotations
- Sample Size Planning for Targeted Clinical Trials

Oncology Needs

- Better treatments
- **Better targeting of treatments to the right patients**

- Many cancer treatments benefit only a small proportion of the patients to which they are administered
- Targeting treatment to the right patients can greatly improve the therapeutic ratio of benefit to adverse effects
  - Smaller clinical trials needed
  - Treated patients benefit
  - Economic benefit for society

“Biomarkers”

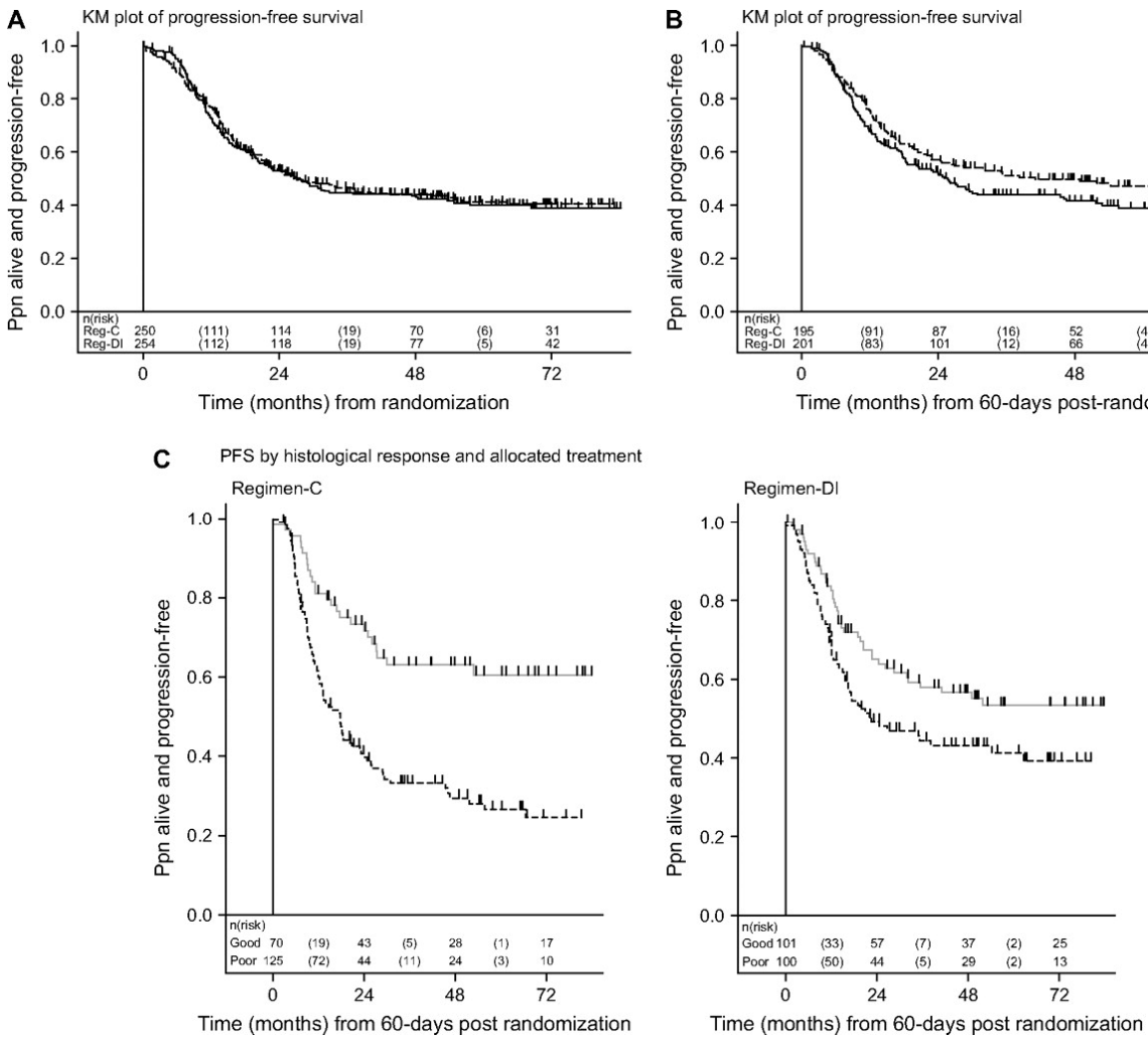
- Predictive classifier
  - A measurement made before treatment to predict whether a particular treatment is likely to be beneficial
- Surrogate endpoints
  - A measurement made before, during and after treatment to determine whether the treatment is working

Surrogate Endpoints

- It is very difficult to properly validate a biomarker as a surrogate for clinical outcome. It requires a series of randomized trials with both the candidate biomarker and clinical outcome measured
  - Must demonstrate that treatment vs control differences for the candidate surrogate are concordant with the treatment vs control differences for clinical outcome
  - It is not sufficient to demonstrate that the biomarker responders survive longer than the biomarker non-responders

- Biomarkers for use as endpoints in phase I or II studies need not be validated as surrogates for clinical outcome
- Unvalidated biomarkers can also be used for early “futility analyses” in phase III trials

Progression-free survival according to allocated treatment



Lewis, I. J. et al. J. Natl. Cancer Inst. 2007 99:112-128; doi:10.1093/jnci/djk015

JNCI Journal of the National Cancer Institute

- It is usually more difficult and time consuming to properly “validate” an endpoint as a surrogate than to use the clinical endpoint in phase III trials
- Critical path objectives may be more effectively met by developing classifiers for treatment selection than by trying to validate surrogate endpoints

- FDA terminology of “valid biomarker” and “probable valid biomarker” are inappropriate
- “Validation” has meaning only as fitness for purpose and the purpose of treatment selection classifiers are completely different than for surrogate endpoints
- Criteria for validation of surrogate endpoints should not be applied to biomarkers used for treatment selection

- The components of multi-gene exrpession based classifiers should not have to be “valid biomarkers”
- It is often much easier to develop an accurate predictive classifier than to elucidate the role of the component genes in disease biology

Oncology Needs Predictive Markers not Prognostic Factors

- Most prognostic factors are not used because they are not therapeutically relevant
- Most prognostic factor studies are poorly designed and not focused on a clear objective; they use a convenience sample of patients for whom tissue is available. Generally the patients are too heterogeneous to support therapeutically relevant conclusions
- Prognostic and predictive studies should be designed with as much care and statistical rigor as clinical trials

Pusztai et al. The Oncologist 8:252-8, 2003

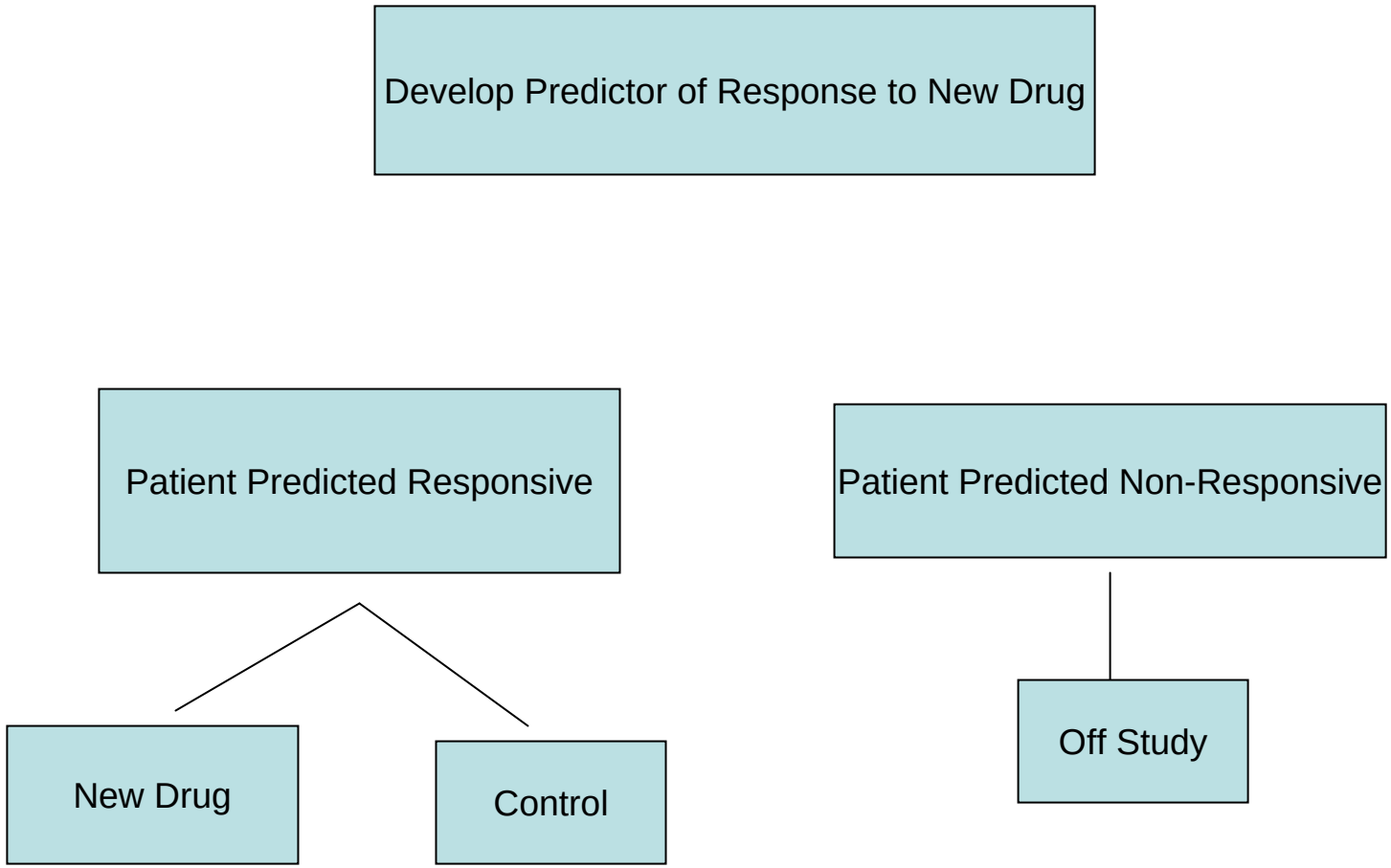
- 939 articles on “prognostic markers” or “prognostic factors” in breast cancer in past 20 years
- ASCO guidelines only recommend routine testing for ER, PR and HER-2 in breast cancer
- “With the exception of ER or progesterone receptor expression and HER-2 gene amplification, there are no clinically useful molecular predictors of response to any form of anticancer therapy.”

- Clinical trials of molecularly targeted drugs focused on patients whose tumors are expected to be susceptible to the drug can be much more efficient than traditional broad clinical trials

- In new drug development
  - The focus should be on evaluating the new drug in a population defined by a predictive classifier, not on “validating” the classifier
- In developing a predictive classifier for use in restricting a widely used treatment
  - The focus should be on evaluating the classifier; Is clinical outcome better if the classifier is used than if it is not used?

New Drug 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

Applicability of Design I

- Primarily for settings where the classifier is based on a single gene whose protein product is the target of the drug
- With substantial biological basis for the classifier, it will often be unacceptable ethically to expose classifier negative patients to the new drug

- Traditional parameters of sensitivity and specificity are not applicable to estimating relative efficacy of a new regimen versus a control with survival or progression-free survival endpoint
  - The relevant parameters are treatment effect in classifier positive and classifier negative subsets
- “When your only tool is a hammer, everything looks like a nail”
- Forcing predictive medicine based drug development into square boxes developed for traditional medical devices creates a serious roadblock to the introduction of effective pharmacogenomic based therapeutics

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 and interactive sample size calculations at http://linus.nci.nih.gov/brb

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

Web Based Software for Comparing Sample Size Requirements

- http://linus.nci.nih.gov/brb/

biomathematics, and computational biology, on topics ranging from methodology to facilitate understanding at the molecular level of the pathogenesis of cancer to methodology to enhance the conduct of clinical trials of new therapeutic and diagnostic approaches.

Research Areas

Clinical trials, Drug Discovery, Molecular Cancer Diagnosis, Biomedical Imaging, Computational and Systems Biology, and Biostatistical Research

Technical Reports and Talks

Download the PDF version of our most recent research papers and PowerPoint version of the talk slides

BRB Staff

Investigators and contact information

BRB ArrayTools

Download the most advanced tools for microarray data analysis

BRB Alumni

Sample Size Calculation

BRB Annual Report 2005

Mathematics And Oncology

- The Norton-Simon Hypothesis
- The Norton-Simon Hypothesis and Breast Cancer Mortality in National Randomized Trial

Position Available

Post-doctoral fellow positions available

Software Download

- Accelerated Titration Design Software
- Optimal Two-Stage Phase II Design Software

Sample Size Calculation for Randomized Clinical Trials

- **Optimal Two-Stage Phase II Design**

- **Biomarker Targeted Randomized Design***

1. **Binary Outcome Endpoint**
2. **Survival and Time-to-Event Endpoint**

* Targeted design randomizes only marker positive patients to treatment or control arm. Untargeted design does not measure marker and randomizes all who otherwise are eligible.

Sample Size Calculation: Binary Outcome Endpoint

Evaluating the efficiency of targeted designs for randomized clinical trials and Supplement by Richard Simon and Aboubakar Maitournam. (Clinical Cancer Research 10:6759-6763, 2005)

pc	
gamma	
delta1	
delta0	
alpha	0.05
power	0.90

Submit

pc = probability of "response" for control arm

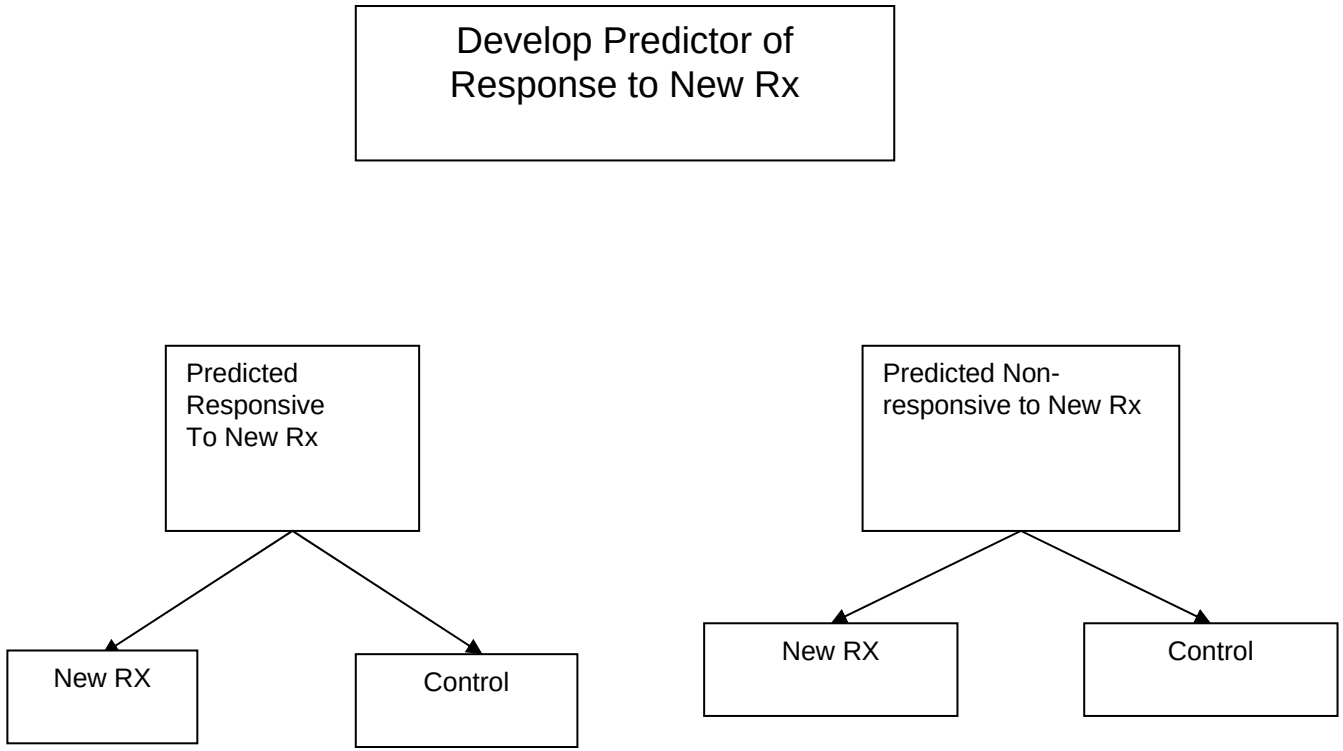
gamma = proportion of patients who are classifier negative (i.e. less responsive to new treatment

delta1 = improvement in response probability for new treatment in classifier positive patients

delta0 = improvement in response probability for new treatment in classifier negative patients

alpha = two-sided significance level

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.

- This analysis strategy is designed to not penalize sponsors for having developed a classifier
- It provides sponsors with an incentive to develop genomic classifiers

- The alternative design of separate testing of treatment effect in classifier positive and negative subsets is generally not viable
  - With classifier tightly linked to drug target, it may be ethically unacceptable to expose classifier negative patients
  - With empirically based classifier, it will be better to not measure classifier than to be forced to demonstrate effectiveness in both subsets separately

FDA Subset Catch 22

- Do not accept claims based on subset analysis
- Require sponsors to do subset analysis to establish that a claim based on overall treatment effect applies to all subsets

Predictive Medicine not Correlative Science

- The purpose of the RCT is to evaluate treatment T vs C overall and for the pre-defined subset
- The purpose is not to re-evaluate the components of the classifier, or to modify or refine the classifier
- The purpose is not to demonstrate that repeating the classifier development process on independent data results in the same classifier

The Roadmap

1. Develop a completely specified genomic classifier of the patients likely to benefit from a new drug
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.

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

- From a non-targeted “negative” clinical trial to develop a binary classifier of a subset thought to benefit from treatment
- Test that subset hypothesis in a separate clinical trial
  - Prospective targeted type (I) trial
  - Prospective type (II) trial
  - Using archived specimens from a second previously conducted clinical trial

Development of Genomic Classifiers

- Single gene or protein based on knowledge of therapeutic target
- Empirically determined based on correlating gene expression to patient outcome after treatment

Development of Genomic Classifiers

- During phase II development or
- After failed phase III trial using archived specimens.
- Adaptively during early portion of phase III trial.

Development of Empirical Gene Expression Based Classifier

- 20-30 phase II responders are needed to compare to non-responders in order to develop signature for predicting response
  - Dobbin KK, Simon RM. Sample size planning for developing classifiers using high dimensional DNA microarray data, Biostatistics 8:101-117, 2007.

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

Validation of Predictive Classifiers for Use with *Available* Treatments

- Should establish that the classifier is reproducibly measurable and has clinical utility
- Studies of predictive classifiers should be viewed as either *developmental* or *validation* studies

- Developmental studies should develop classifiers for homogeneously treated patients and provide split-sample or cross-validated estimates of prediction accuracy
- Validation studies should establish whether patient outcome is improved using the pre-specified new classifier for treatment selection compared to using current practice standards

ARTICLE

Critical Review of Published Microarray Studies for Cancer Outcome and Guidelines on Statistical Analysis and Reporting

Alain Dupuy, Richard M. Simon

Background Both the validity and the reproducibility of microarray-based clinical research have been challenged. There is a need for critical review of the statistical analysis and reporting in published microarray studies that focus on cancer-related clinical outcomes.

Methods Studies published through 2004 in which microarray-based gene expression profiles were analyzed for their relation to a clinical cancer outcome were identified through a Medline search followed by hand screening of abstracts and full text articles. Studies that were eligible for our analysis addressed one or more outcomes that were either an event occurring during follow-up, such as death or relapse, or a therapeutic response. We recorded descriptive characteristics for all the selected studies. A critical review of outcome-related statistical analyses was undertaken for the articles published in 2004.

Results Ninety studies were identified, and their descriptive characteristics are presented. Sixty-eight (76%) were published in journals of impact factor greater than 6. A detailed account of the 42 studies (47%) published in 2004 is reported. Twenty-one (50%) of them contained at least one of the following three basic flaws: 1) in outcome-related gene finding, an unstated, unclear, or inadequate control for multiple testing; 2) in class discovery, a spurious claim of correlation between clusters and clinical outcome, made after clustering samples using a selection of outcome-related differentially expressed genes; or 3) in supervised prediction, a biased estimation of the prediction accuracy through an incorrect cross-validation procedure.

Conclusions The most common and serious mistakes and misunderstandings recorded in published studies are described and illustrated. Based on this analysis, a proposal of guidelines for statistical analysis and reporting for clinical microarray studies, presented as a checklist of "Do's and Don'ts," is provided.

J Natl Cancer Inst 2007;99:147–57

DNA microarray technology has found many applications in biomedical research. In oncology, it is being used to better understand the biological mechanisms underlying oncogenesis, to discover new targets and new drugs, and to develop classifiers (predictors of good outcome versus poor outcome) for tailoring individualized treatments (1–4). Microarray-based clinical research is a recent and active area, with an exponentially growing number of publications. Both the reproducibility and validity of findings have been challenged, however (5,6). In our experience, microarray-based clinical investigations have generated both unrealistic hype and excessive skepticism. We reviewed published microarray studies in which gene expression data are analyzed for relationships with cancer outcomes, and we propose guidelines for statistical analysis and reporting, based on the most common and serious problems identified.

Medicine, followed by hand screening of abstracts and articles. The detailed process of selection is presented in Supplementary Note 1 (available online). The inclusion criteria were as follows: the work was an original clinical study on human cancer patients, published in English before December 31, 2004; it analyzed gene expression data of more than 1000 spots; and it presented statistical analyses relating the gene expression profiling to a clinical outcome. Two types of outcome were considered: 1) A relapse or death occurring during the course of the disease. 2) A therapeutic response.

Affiliations of authors: Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, MD (AD, RMS); Université Paris VII Denis Diderot, Paris, France (AD); Assistance Publique-Hôpitaux de Paris, Service de Dermatologie, Hôpital Saint-Louis, Paris, France (AD).

Correspondence to: Richard M. Simon, DSc, National Cancer Institute, 9000 Rockville Pike, MSC 7434, Bethesda, MD 20892 (e-mail: rsimon@nih.gov).

Major Flaws Found in 40 Studies Published in 2004

- Cluster Analysis of samples
  - 13/28 studies invalidly claimed that expression clusters based on differentially expressed genes could help distinguish clinical outcomes
- Outcome related gene finding
  - 9/23 studies had unclear or inadequate methods to deal with false positives
    - 10,000 genes x .05 significance level = 500 false positives
- Supervised prediction
  - 12/28 reported a misleading estimate of prediction accuracy
- 50% of studies contained one or more major flaws

and criteria used for selection of cases.

9	Don't	Transform time-to-outcome data into a binary outcome variable if the goal is to predict groups with different survival probabilities.	Use statistical methods suited for time-to-event data, unless you can ensure the absence of bias due to transformation. See text and Supplementary Fig. 2 (available online).
Outcome-related gene finding†			
10	Don't	Use only fold changes between groups to select the differentially expressed genes.	This does not take into account the variance of the genes' data values.
11	Don't	Use a .05 *P* value threshold to select the differentially expressed genes.	A set of 10000 genes will yield on average 500 false-positive genes if this threshold is used.
12	Do	Use a method for controlling the number of falsely differentially expressed genes.	Lowering the *P* value threshold for selection (e.g., to .001) is the simplest method. Others are available.
13	Do	Use a permutation test to assess the probability of finding the same number of differentially expressed genes as the one you found from your dataset.	The result should be significant at .05 *P* value level.
Class discovery			
14	Don't	Use class discovery methods if you are interested in classifying new samples in the future.	Supervised prediction should be used for this purpose. It utilizes the outcome information to optimize predictive accuracy. See text.
15	Don't	Use a selection of outcome-related differentially expressed genes if you intend to correlate cluster-defined classes with the outcome.	Supervised clustering leads to a spurious correlation between cluster and outcome. See text and Fig. 1.
16	Don't	Select the clustering method that gives the best result.	Class discovery should not be result driven.
17	Do	Use methods for testing the reproducibility of cluster finding.	Assessing the reproducibility of cluster finding without using external information makes class discovery more convincing. See text.
18	Don't	Use conventional statistical tests for computing the statistical significance of genes that are differentially expressed between two clusters.	These tests assume independence between class definition and expression profile data, which is not the case for cluster-defined classes.
Supervised prediction			
19	Do	Frame a therapeutically relevant question and select a homogeneous set of patients accordingly.	Classifiers developed outside a specific therapeutically relevant context are unlikely to be useful and utilized. See text.
20	Don't	Violate the fundamental principle of classifier validation, i.e., no preliminary use of the tested samples.	Most of the "Don't" items on validation procedures are illustrations of how this principle can be violated. See text and Fig. 2 and Supplementary Fig.1 (available online).
21	Don't	Attempt to predict cluster-defined classes.	Classes should be defined independently from the expression profile data.
Evaluating the prediction on a separate test set			
22	Don't	Use any information from the test set for developing the classifier.	The test set is to be used exclusively for evaluating the classifier performance. See text and Fig. 2.
23	Do	Access the test set only once and only for testing the samples with the fully specified classifier developed from the training set.	The test set must not be used to choose the best classifier. See text and Fig. 2.
24	Do	Use the same outcome definition as the one used in the training set.	

(Table continues)

Checklist			Comment
	Evaluating the prediction with a cross-validation procedure		
25	Don't	Use all the samples from the dataset to develop the classifier and test them.	The resubstitution estimate is not a cross-validation procedure. See text and Fig. 2.
26	Don't	Use the same feature selection for all iterations.	This inflates the estimate of the prediction accuracy. See text and Fig. 2.
27	Don't	Perform a cross-validation procedure on a selection of outcome-related differentially expressed genes.	Idem. Invalid although commonly done.
28	Do	Report the estimates for all the classification algorithms if several have been tested, not just the most accurate.	
29	Don't	Consider that testing a few additional independent samples adds value to a correctly cross-validated estimate of the classifier prediction accuracy.	However, this may be valuable if the additional samples are in sufficient number and are representative of the samples in which the classifier might be used in the future. See text.
30	Do	Report the fully specified classifier with its parameters.	So it can be used by others. Parameters are obtained from the whole training set in a separate test set procedure and from the whole dataset in a cross-validation procedure.
31	Do	Report the correctly validated sensitivity and specificity or positive and negative apparent predictive values (for a binary outcome).	Receiver-operating characteristic curves may also be used. See text.
32	Don't	Use an odds ratio to assess the performance of the prediction (for a binary outcome).	The odds ratio is a measure of association, not of prediction accuracy. See text and Supplementary Fig. 3 (available online).
33	Do	Report the statistical significance of the prediction accuracy and, even better, of the sensitivity and specificity (for a binary outcome).	It states the probability of obtaining a prediction accuracy as high as actually observed if there was no relationship between the expression data and the outcome. See text.
34	Don't	Use a Fisher's exact test or chi-square test to assess the statistical significance of the prediction accuracy for a binary outcome.	They do not test the statistical significance of the prediction. See text and Supplementary Fig. 3 (available online).
35	Do	Pay attention to the imbalance between outcome categories when interpreting the prediction accuracy of a binary outcome.	90% prediction accuracy may be inadequate if outcome categories are highly imbalanced. See text and Supplementary Fig. 3 (available online).
36	Don't	Use the log-rank test for testing the difference in survival between cross-validated groups.	The test is invalid because of a dependency among cases after cross-validation.
37	Don't	Use standard regression models, e.g., logistic regression or proportional hazards model, with cross-validated predicted groups.	Idem.
38	Don't	Assess the utility of the prediction based on the value of the regression coefficient or on its *P* value from multivariable regression models.	Regression coefficients are poor measures of prediction accuracy, and the test of statistical significance simply assesses if the coefficient is different from 0. See text.
39	Do	Assess the added value of the classifier by examining its performance within the levels of the standard prognostic factors.	Other approaches can be used. See text.
40	Do	Assess the utility of the classifier in a clinical context, for the therapeutically relevant question, and plan, if appropriate, further studies for external validation.	

BRB-ArrayTools

- Contains analysis tools that I believe are valid and most informative
- Extensive analysis tools for developing and validating predictive classifiers
  - Binary or survival endpoints
- Guided selection of analysis tools for biomedical scientists
- Imports data from all platforms and major databases

BRB-ArrayTools

- Extensive linkage to gene annotation websites
- Extensive analyses for integrating gene expression with other biological information
- Publicly available for non-commercial use
  - http://linus.nci.nih.gov/brb

BRB-ArrayTools

December 2006

- 6635 Registered users
- 1938 Distinct institutions
- 68 Countries
- 311 Citations

Conclusions

- Developments in biotechnology and tumor biology make it increasingly feasible to identify which patients are most likely to benefit from a specified treatment

Achieving the potential of new technology requires

- Paradigm changes in study design, moving from “correlative science” to predictive medicine
- New organizational structures and resource allocations to foster excellence in interdisciplinary research among biostatistical, laboratory and clinical scientists
  - Traditional core support structures are ineffective for high level collaboration
  - Major studies continue to be poorly designed and analyzed
  - Over-emphasis on software engineering at the expense of biostatistical collaboration
- FDA policies that encourage development of classifier targeted therapeutics

Collaborators

- Alain Dupuy
- Boris Freidlin
- Aboubakar Maitournam
- Yingdong Zhao

Using Genomic Classifiers In Clinical Trials

•Dupue A and Simon R. Critical review of published microarray studies for clinical outcome and guidelines for statistical analysis and reporting, Journal of the National Cancer Institute 99:147-57, 2007
.

•Dobbin K and Simon R. Sample size planning for developing classifiers using high dimensional DNA microarray data. Biostatistics 8:101-117, 2007.

•Simon R. Development and validation of therapeutically relevant predictive classifiers using gene expression profiling, Journal of the National Cancer Institute 98:1169-71, 2006.

•Simon R. Validation of pharmacogenomic biomarker classifiers for treatment selection. Cancer Biomarkers 2:89-96, 2006.

•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, pp 3-15, Humana Press, 2006.

•Simon R. A checklist for evaluating reports of expression profiling for treatment selection. Clinical Advances in Hematology and Oncology 4:219-224, 2006.

•Simon R. Identification of pharmacogenomic biomarker classifiers in drug development. In Pharmacogenomics, Anti-cancer Drug Discovery and Response, F Innocenti (ed), Humana Press (In Press).

•Simon R, Lam A, Li MC, et al. Analysis of gene expression data using BRB-ArrayTools, Cancer Informatics 2:1-7, 2006

•Simon R. New challenges for 21st century clinical trials, Controlled Clinical Trials (In Press).

•Simon R. Development and validation of biomarker classifiers for treatment selection, Journal of Statistical Planning and Inference (In Press).

Using Genomic Classifiers In Clinical Trials

.

•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 23:7332-41,2005.

•Freidlin B and Simon R. Adaptive signature design. Clinical Cancer Research 11:7872-78, 2005.

•Simon R. and Wang SJ. Use of genomic signatures in therapeutics development in oncology and other diseases, The Pharmacogenomics Journal 6:166-73, 2006.

•Trepicchio WL, Essayan D, Hall ST, Schechter G, Tezak Z, Wang SJ, Weinreich D, Simon R. Designing prospective clinical pharmacogenomic trials- Effective use of genomic biomarkers for use in clinical decision-making. The Pharmacogenomics Journal 6:89-94,2006.