Using Biomarkers & Gene Expression Classifiers in Clinical New Drug Development

Richard Simon, D.Sc.

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.

Biomarker

- Any biological measurement made on a patient

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

- "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

Surrogate Endpoints

- Properly validate a biomarker as a surrogate for clinical benefit requires a series of randomized trials with both the candidate biomarker and clinical outcome measured and then demonstration that treatment vs control conclusions for the surrogate are consistent with treatment vs control conclusions for clinical outcome

Cardiac Arrhythmia Supression Trial

- Ventricular premature beats was proposed as a surrogate for survival

- Antiarrythmic drugs supressed ventricular premature beats but killed patients at approximately 2.5 times that of placebo

- It is rare that we understand disease pathophysiology well enough to argue that a biomarker is self evidently a proper surrogate endpoint for clinical utility
- It is often more difficult and time consuming to properly "validate" an endpoint as a surrogate than to use the clinical endpoint in phase III trials
- The time frame for validating a surrogate is inconsistent with the time frame for initiating a pivotal study

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

Objectives of Phase I Trials

- Develop dose/schedule
- Determine whether the drug inhibits the targeted pathway

Dose/Schedule

- Ideal is to have a drug and target so specific for cancer cells that the drug can be delivered repeatedly at doses that completely shut down the de-regulated pathway without toxicity to normal cells
- Because most current targets are not specific to cancer cells, most targeted drugs are toxic

Dose/Schedule

- Few examples of drugs whose effectiveness at inhibiting target decreases with dose after maximum

- Titrating dose for maximum inhibition of target is difficult due to assay variability and need for tumor biopsies

- Titrating dose to plasma concentration at which target is inhibited in pre-clinical systems is more feasible

Dose/Schedule

- Determining dose just below MTD which can be delivered repeatedly is often the most appropriate and practical approach
- Accrue an additional cohort of patients at that selected dose to determine whether the target is inhibited

Objectives of Phase II Trials of Targeted Agents

- Determine whether there is a population of patients for whom the drug demonstrates sufficient anti-tumor activity to warrant a phase III trial

- Optimize the regimen in which the drug will be used in the phase III trial

- Optimize the target population for the phase III trial

Endpoints for Phase II

- Tumor shrinkage
- Inhibition of disease biomarker
  - Need not be validated
- Time to progression or proportion of patients without progression at a specified time

Using Time to Progression as Endpoint in Phase II Trials

- Requires comparison to distribution of progression times for patients not receiving drug

- Proportion of patients without progression at a specified time also requires comparison for evaluation

- Historical control vs randomized comparison

- Phase 2.5 trial design

Phase 2.5 Trial Design

- Simon R et al. Clinical trial designs for the early clinical development of therapeutic cancer vaccines. Journal of Clinical Oncology 19:1848-54, 2001

- Korn EL et al. Clinical trial designs for cytostatic agents: Are new approaches needed? Journal of Clinical Oncology 19:265-272, 2001

Phase 2.5 Trial Design

- Randomization to chemotherapy alone or with new drug

- Endpoint is progression free survival regardless of whether it is a validated surrogate of survival

- One-sided significance level can exceed .05 for analysis and sample size planning
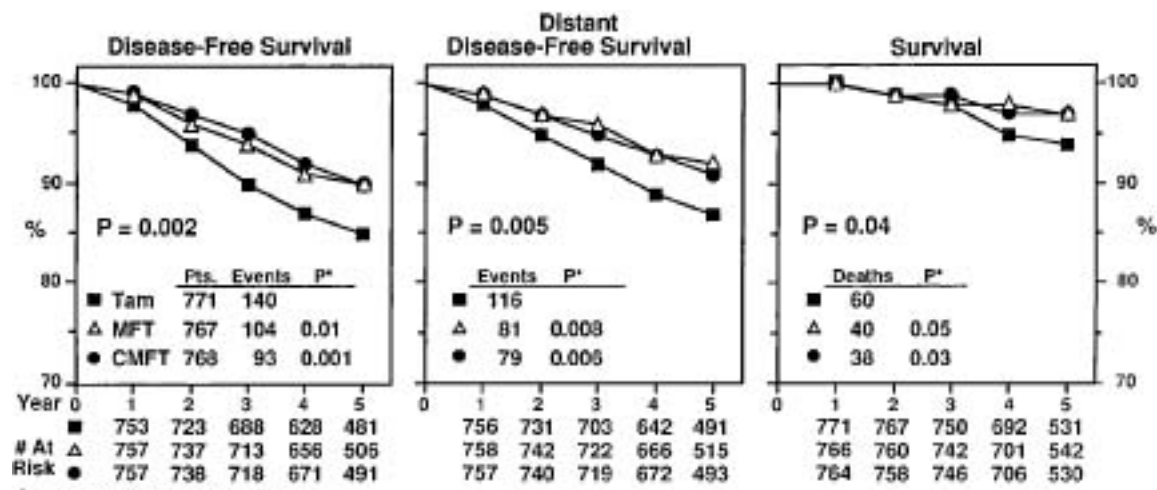
Total Sample Size
Randomized Phase 2.5
2 years accrual, 1.5 years followup

Improvement in median PFS	Hazard Ratio	$\alpha=.05$	$\alpha=.10$	$\alpha=.20$
4 → 6 months	1.5	216	168	116
6 → 9 months	1.5	228	176	120
4 → 8 months	2	76	60	40
6→12 months	2	84	64	44

- It is important to characterize in phase II studies which tumors are most likely to be sensitive to the drug

Pharmacogenomic 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

Disease-Free Survival

Distant Disease-Free Survival

Survival

P = 0.002

P = 0.005

P = 0.04

	Pts.	Events	P*
■ Tam	771	140	
△ MFT	767	104	0.01
● CMFT	768	93	0.001

	Events	P*
■	116	
△	81	0.008
●	79	0.006

	Deaths	P*
■	60	
△	40	0.05
●	38	0.03

Year 0   1   2   3   4   5

# At Risk	Year 0	1	2	3	4	5
■		753	723	688	628	481
△		757	737	713	656	506
●		757	738	718	671	491

■	756	731	703	642	491
△	758	742	722	666	515
●	757	740	719	672	493

■	771	767	750	692	531
△	766	760	742	701	542
●	764	758	746	706	530

* 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)	0.54 (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 focused on patients selected based on tumor assays can be much more efficient than untargeted clinical trials

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 and interactive sample size calculations 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

- 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

You Can Evaluate How This Design Might Work For You

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

research programs of the division in developmental therapeutics, developmental diagnostics, diagnostic imaging and clinical trials. The members of the branch also conduct research in biostatistics, 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 Array Tools

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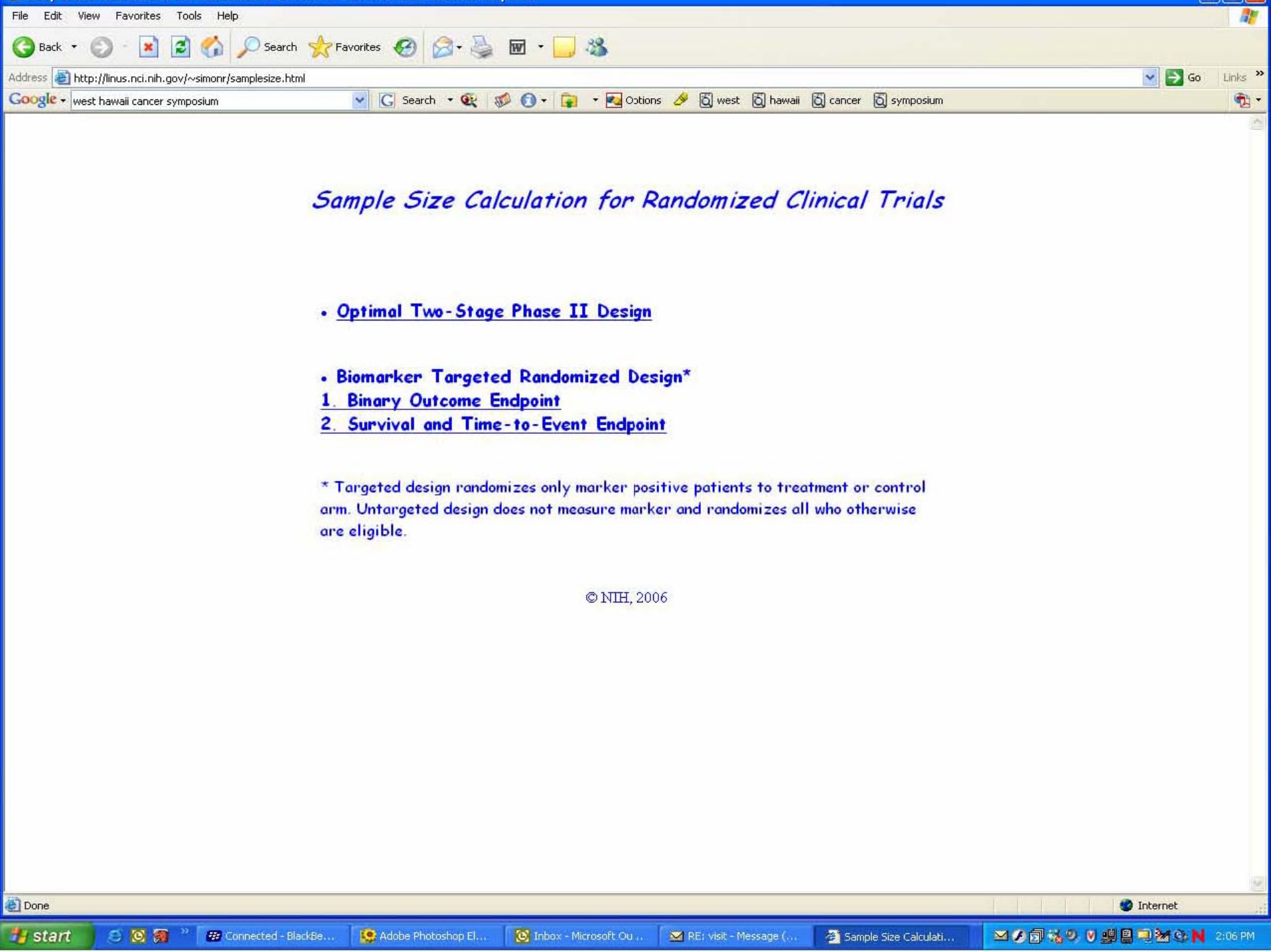- The Norton-Simon Hypothesis and Breast Cancer Mortality in National Randomized Trial

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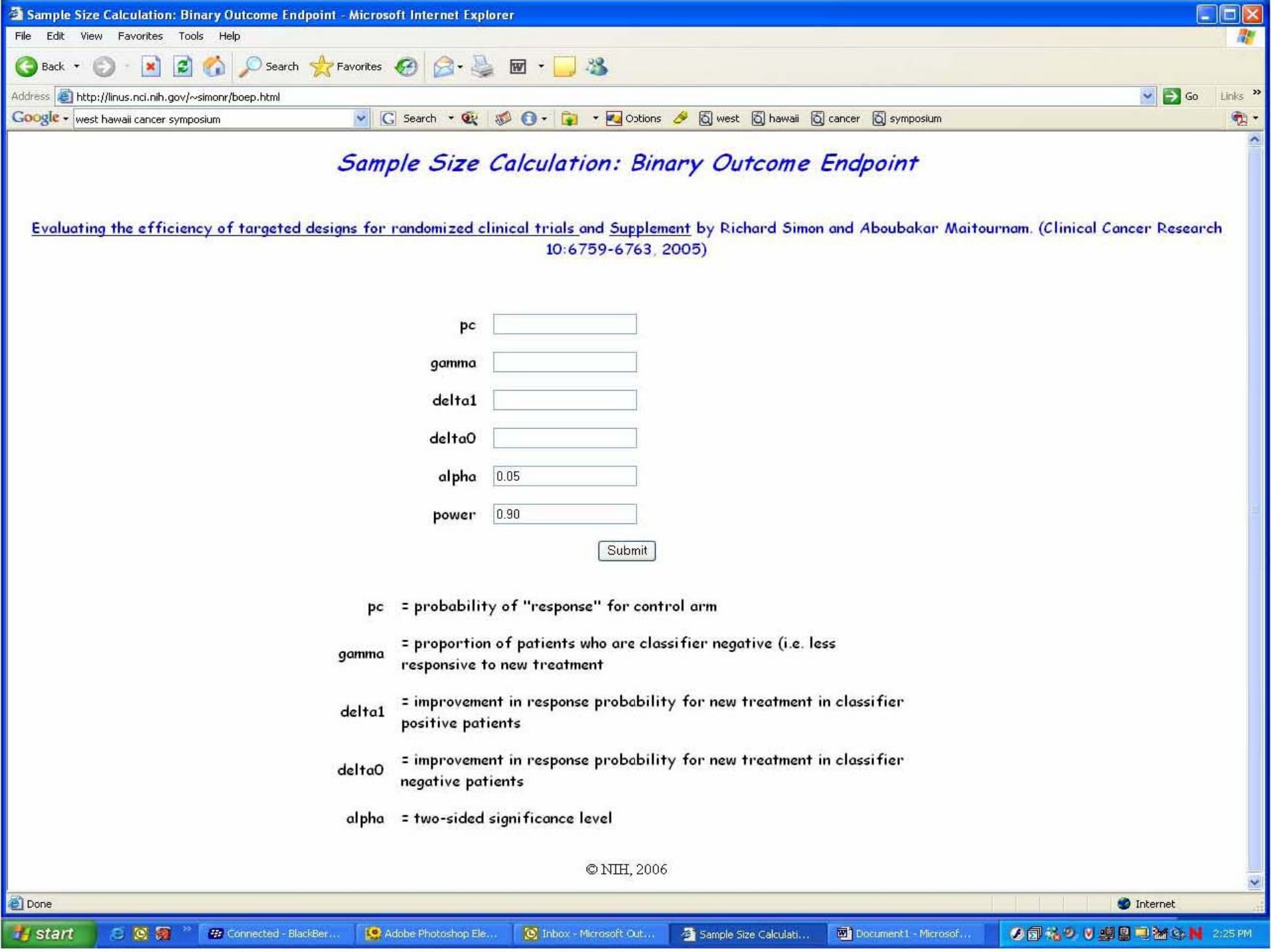
- **Optimal Two-Stage Phase II Design**

- 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

© NIH, 2006

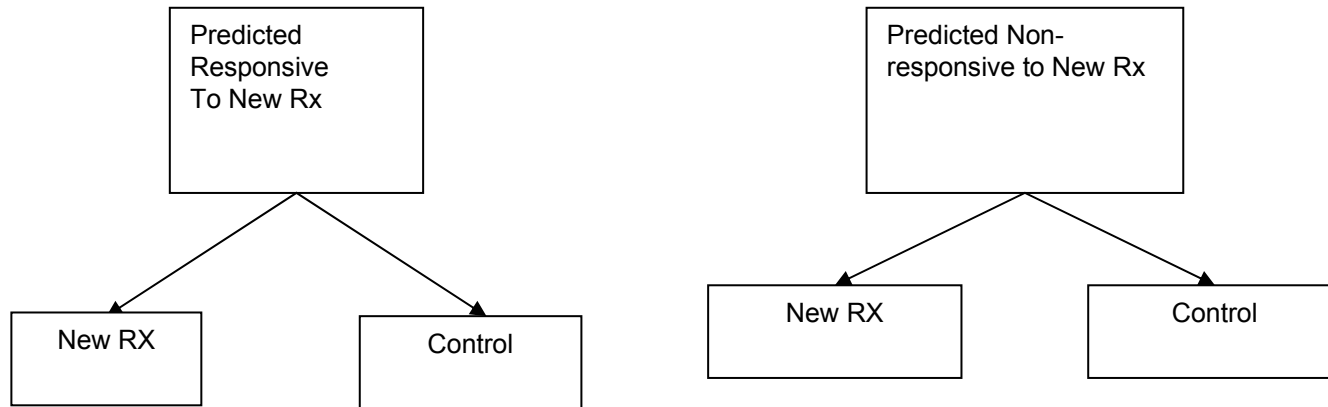# Sample Size Calculation: Survival or Time-to-Event Endpoint*

	Median survival of the control group (years)

or

	Proportion surviving beyond [ ] years
	Total accrual rate (both marker positive and negative patients/year)
	Percent of patients marker negative
	% Reduction in hazard for treatment of marker positive patients
	% Reduction in hazard for treatment of marker negative patients
	Years of follow-up following end of accrual
0.05	Two-sided significance
0.90	Desired power for targeted design

Submit

*Assumes exponential distribution of survival for treatment and control group within marker positive and marker negative subsets. Uses formulas in Rubinstein, Gail & Santner (J Chronic Disease 34:469-79, 1981) for targeted design and simulation for untargeted design. Simulation uses Poisson process assumptions of Rubinstein et al.

© NIH, 2006

Done   Internet

start   Connected - BlackBer...   Adobe Photoshop Ele...   Inbox - Microsoft Out...   Sample Size Calculati...   Document1 - Microsof...   2:26 PM

Developmental Strategy (II)

Develop Predictor of
Response to New Rx

Predicted
Responsive
To New Rx

Predicted Non-
responsive to New Rx

New RX

Control

New RX

Control

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

- Empirically determined based on correlating gene expression to patient outcome after treatment

Development of Genomic Classifiers

- During phase I/II development. or

- After failed phase III trial using archived specimens

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

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

- 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

- Much "correlative science" activity is unproductive and focused on development of surrogate endpoints

- "Correlative science" activity should be refocused on developing predictive markers of which patients are likely to benefit from a treatment

Conclusions

- Technology is sufficiently mature today to effectively identify which patients benefit from new treatments and to dramatically improve the efficiency of clinical trials

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 previous developmental studies

Acknowledgement

- Kevin Dobbin
- Boris Freidlin
- Aboubakar Maitournam
- Sue-Jane Wang
- Yingdong Zhao