Adaptive Determination of the Intended Use Population

Richard Simon
R Simon Consulting
rmaceysimon@gmail.com
http://rsimon.us

Noah Simon
Department of Biostatistics
University of Washington
nrsimon@u.washington.edu

Old View

- Broad eligibility
  - Fear that an approved drug might not work in widespread community practice
- Assumption that disease is homogeneous and that treatment benefits all patients similarly

New Paradigm in Oncology Clinical Trials

- New century - recurrent somatic mutations in tumors were discovered
  - 50% of melanoma tumors contained the same point mutation in the BRAF gene
- Tumors of the same primary site can represent different diseases, with sensitivity to different treatments.
- This changed the approach to discovery and clinical evaluation of new treatments.

Number of Oncology Drugs/Indications Approved by FDA

- 2000 – 2005 35 drugs
- 2011 – 2016 85 drugs

The Most Important Decisions in Developing a Phase III Clinical Trial

- Whether to do a phase III trial
- What patient population

The Most Important Decisions in Developing a Phase III Clinical Trial

The Most Important Decisions in Developing a Phase III Clinical Trial

- Whether to do a phase III trial
- What patient population

$$\text{RandRat} = \frac{n_{untargeted}}{n_{targeted}}$$

If $TE_{-}=0$

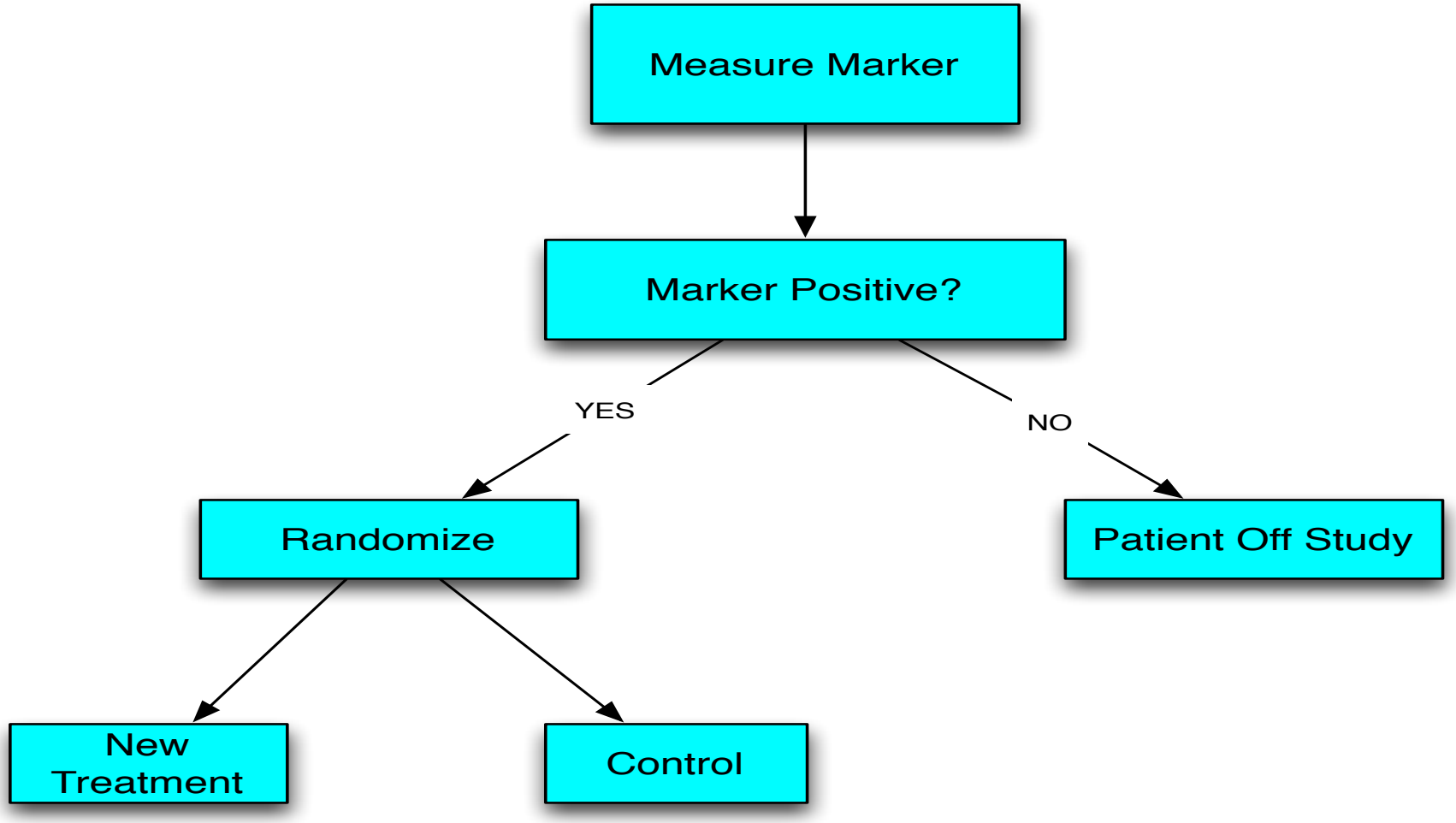
- $\text{RandRat} = \frac{1}{{p_+}^2}$
- if $p_+=0.5$, RandRat=4

Comparing E vs C on Survival or DFS

5% 2-sided Significance and 90% Power

% Reduction in Hazard	Number of Events Required
25%	509
30%	332
35%	227
40%	162
45%	118
50%	88

Enrichment Design



- Large randomized phase II trials can take a long time

Adaptive Determination of Intended Use Population

- Randomized population is only used as intended use population to avoid issues of subset analysis
- The proportion of patients in the randomized population who benefit from the test treatment in “positive” clinical trials is very small

Adaptive determination of intended use population

does not require adaptive changes during trial

Adaptive threshold design

Adaptive signature design

Cross-validated adaptive signature design

Biomarker-Adaptive Threshold Design: A Procedure for Evaluating Treatment With Possible Biomarker-Defined Subset Effect

Wenyu Jiang, Boris Freidlin, Richard Simon

Background Many molecularly targeted anticancer agents entering the definitive stage of clinical development benefit only a subset of treated patients. This may lead to missing effective agents by the traditional broad-eligibility randomized trials due to the dilution of the overall treatment effect. We propose a statistically rigorous biomarker-adaptive threshold phase III design for settings in which a putative biomarker to identify patients who are sensitive to the new agent is measured on a continuous or graded scale.

Methods The design combines a test for overall treatment effect in all randomly assigned patients with the establishment and validation of a cut point for a prespecified biomarker of the sensitive subpopulation. The performance of the biomarker-adaptive design, relative to a traditional design that ignores the biomarker, was evaluated in a simulation study. The biomarker-adaptive design was also used to analyze data from a prostate cancer trial.

Results In the simulation study, the biomarker-adaptive design preserved the power to detect the overall effect when the new treatment is broadly effective. When the proportion of sensitive patients as identified by the biomarker is low, the proposed design provided a substantial improvement in efficiency compared with the traditional trial design. Recommendations for sample size planning and implementation of the biomarker-adaptive design are provided.

Conclusions A statistically valid test for a biomarker-defined subset effect can be prospectively incorporated into a randomized phase III design without compromising the ability to detect an overall effect if the intervention is beneficial in a broad population.

J Natl Cancer Inst 2007;99:1036–43

Human cancers are heterogeneous with regard to their molecular and genomic properties. Recent advances in biotechnology have resulted in a shift toward molecularly targeted anticancer agents. These new therapeutics are likely to benefit only a subset of the patients with a given cancer. Definitive testing of such targeted agents requires the identification of the appropriate "sensitive" population. When biomarkers to identify the patients who are likely to benefit from the new therapy are available, targeted clinical trials that restrict eligibility to sensitive patients should be used (1). However, reliable assays to identify sensitive patients are often unavailable. In the absence of a reliable biomarker, broad-eligibility clinical trials are used routinely. Most of these trials use a conventional design, in which the primary analysis is based on comparison of all randomly assigned patients. This often leads to the failure to recognize effective agents due to dilution of the treatment effect by the presence of the patients who do not benefit from the agent. Retrospective analysis of trials with a conventional design can be used as an initial step in identifying biomarkers for the sensitive subpopulation. However, retrospectively identified biomarkers typically have to be validated in a confirmatory prospective randomized phase III clinical trial (2). This approach is inefficient and may considerably prolong clinical development.

Previously, we have proposed a design [adaptive signature design (3)] that combines a definitive test for treatment effect in a broad population with identification and validation of a genomic signature for the subset of sensitive patients if the broad population test is negative. The adaptive signature design was developed for high-dimensional data such as gene expression microarrays, where only a few unknown genes among thousands assayed may be relevant and where a classifier (signature) to identify sensitive patients is not available. The design incorporates both the identification and the validation of a pharmacogenomic signature for sensitive patients.

Often, preliminary information on a biomarker to identify the sensitive subset of patients is available but an appropriate cutoff

Affiliation of authors: Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD.

Correspondence to: Boris Freidlin, PhD, Biometric Research Branch, Division of Cancer Treatment and Diagnosis, EPN-8122, National Cancer Institute, Bethesda, MD 20892 (e-mail: freidlinb@ctep.nci.nih.gov).

See "Notes" following "References."

DOI: 10.1093/jnci/djm022

Adaptive Threshold Design

- Randomized clinical trial of E vs C
- Single candidate biomarker B with K candidate cut-points b_1, …, b_K in [0,1]
  e.g. b_1=0, b_2=0.25 & b_3=0.5
- Entry not restricted by biomarker value

Final Analysis

- Test H_k for k=0,1,…,K.
  - H_k : treatment effect is 0 for population with $B \geq b_k$
  - Compute p_k for treatment effect for each population with $B \geq b_k$
- Use $p^* = \min \{p_k\}$ as global test statistic.
- Test significance of p^* using a permutation test.
- If global null hypothesis was rejected, model treatment effect as a function of biomarker value
  - Compute bootstrap confidence intervals for the optimal cut-point

Threshold Model for Survival Data

$$\log\left(\frac{h(t,x,z)}{h_0(t)}\right) = \beta x + \gamma z I(x \geq b)$$

z=0,1 treatment indicator

x=biomarker value

I=indicator function

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.

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.

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

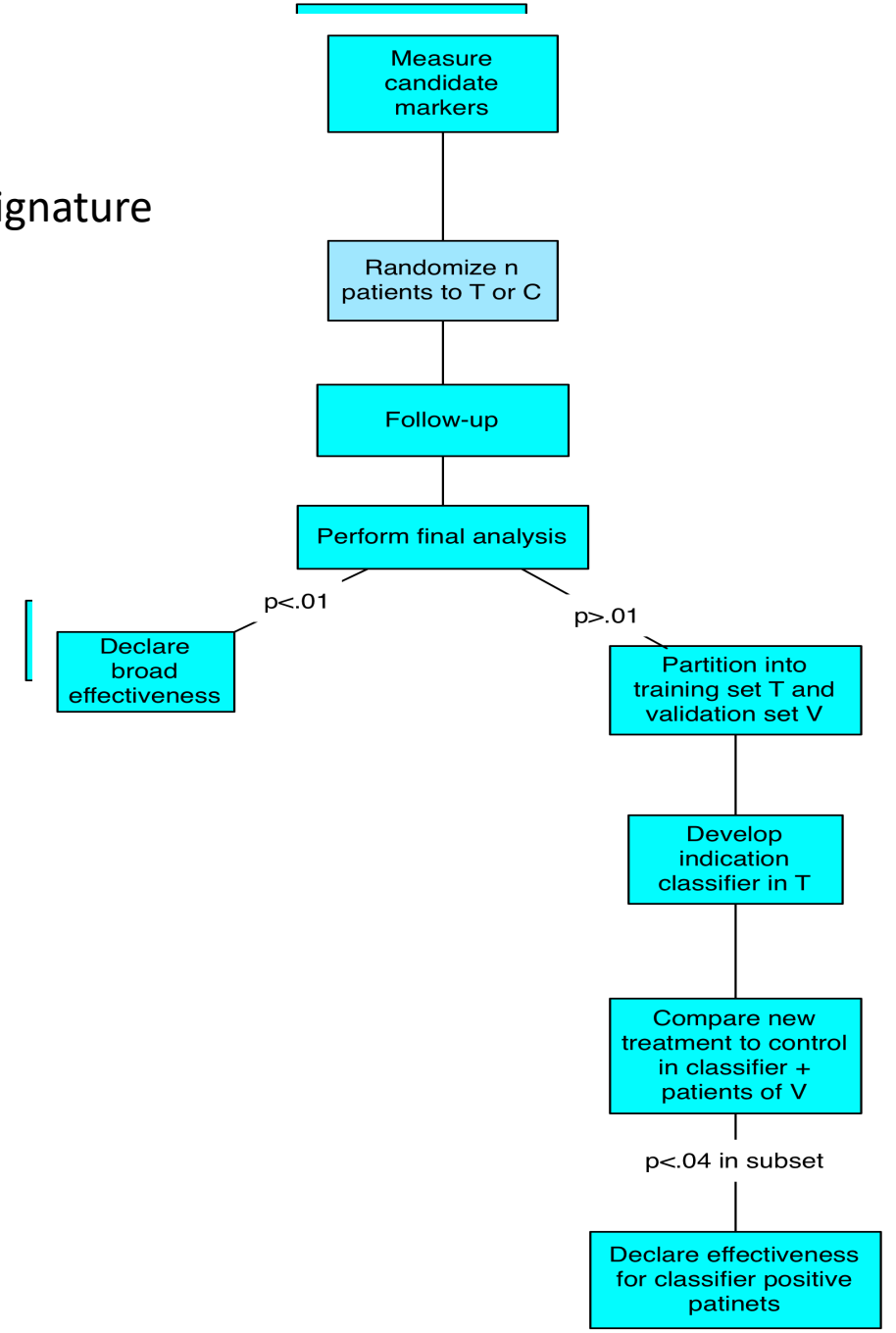
Received 3/18/05; revised 7/18/05; accepted 8/4/05.

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: freidlinb@ctep.nci.nih.gov.

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

Adaptive Siganture Design
Discover & Validate Predictive Signature



What is a Predictive Classifier?

Predictive Classifier

- A predictive classifier is not a prognostic classifier
- It is a binary classifier of whether the prognosis of a patient on E is better than the prognosis of the patient on C

- X vector ➔ {E,C}

Predictive Classifier

- The predictive classifier may be based on separate prognostic classifiers for patients on E and for patients on C

  - P(x|C) probability of response to rx C for patient with covariate vector x
  - P(x|E) probability of response to rx E for patient with covariate vector x

  - Predictive Classifier(x) = E if $P(x|E) > P(x|C) + \varepsilon$
  - = C otherwise

- Fit (penalized) PH model

$$\log \frac{h(t;z,x)}{h_0(t)} = \alpha z + z\beta' x + (1-z)\gamma' x$$

$z \sim (0,1)$ treatment indicator

x ~ covariate vector

$$\frac{h(t;z=1,x)}{h_0(t)} - \frac{h(t;z=0,x)}{h_0(t)} = \alpha + (\beta - \gamma)'x$$

Classify E if $\alpha + (\beta - \gamma)'x \leq \varepsilon,$ otherwise classify C

Example

- At interim analysis determine the mle's of the regression coefficients and their estimated covariance matrix
- Compute approximate mean and variance of

$$\Delta(x) = \hat{\alpha} + (\hat{\beta}_1 - \hat{\gamma}_1)x_1 + (\hat{\beta}_2 - \hat{\gamma}_2)x_2$$
$$\sim \mathrm{N}(\mu(\mathrm{x}), \sigma(\mathrm{x}))$$

$$\Pr\left[\Delta(\underline{x}) \leq \Delta^*\right] \approx \Phi\left\{\frac{\Delta^* - \mu(\underline{x})}{\sigma(\underline{x})}\right\}$$

Exclude future cases with covariate vectors for which this quantity is $\leq \varepsilon$

Evaluation of Predictive Classifier on Separate Test Set

Classify each patient in the test set using their covariate vectors using the classifier developed on the training set

Compute Kaplan-Meier curves of treatments for patients classified
as likelihood to benefit from T over C

Compute log-rank test comparing the two Kaplan-Meier curves

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

Cross-Validated Adaptive Signature Design

- Define predictive classifier development algorithm A
- Apply algorithm to full dataset D to develop predictive classifier M(x;D,A) for use with future cases

- How to evaluate the performance of this classifier?
  - How to avoid the bias of “re-substitution” since there is no separate test set?

Pre-validation solution

- Construct a “pre-validated” test set
- The pre-validated test set will contain all of the cases with their covariate vectors, treatment indicators and outcomes
- The synthesized predictive score for case i is $s_i' = M(x_i ; D^{-i} , A)$.

Convert the pre-validated scores s' to a binary classification,

Compute Kaplan-Meier curves of treatments for patients classified as likelihood to benefit from E over C

Compute log-rank statistic comparing the two Kaplan-Meier curves

Compute log-rank statistic comparing the two Kaplan-Meier curves

Use permutations of treatment indices to evaluate significance of the pre-validated Kaplan-Meier curves. The entire cross-validation must be repeated from scratch.

Key Ideas

- Replace multiple significance testing by development of one predictive classifier
- Internal validation by computing significance of treatment effect in adaptively determined intended use population
- Obtain almost unbiased estimate of the treatment effect of future classifier positive patients

Pre-trial planning for Adaptive Signature or Adaptive Threshold Design

- Analysis plan should be in protocol
- Analysis plan should specify candidate covariates and threshold cut-offs

Measures of predictive performance for survival data

- Spread of KM curves of the two treatment groups for the subset of patients classified as likely to benefit from T over C.
  - Log-rank statistic
  - Hazard ratio
- Area under time-dependent ROC curve for adaptively determined subset
- Simon's sensitivity, specificity, npv, ppv for binary predictive covariates and survival data

Figure 1: Overall analysis. The value of the log-rank statistic is 2.9 and the corresponding p-value is 0.09. The new treatment thus shows no benefit overall at the 0.05 level.

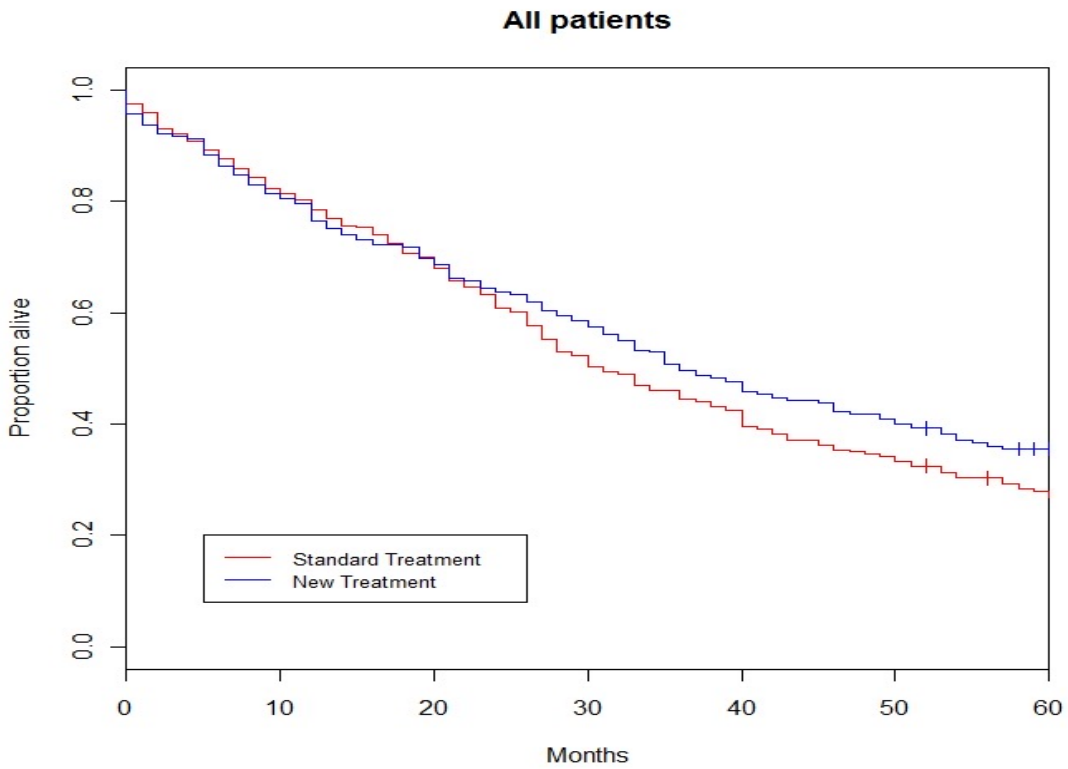
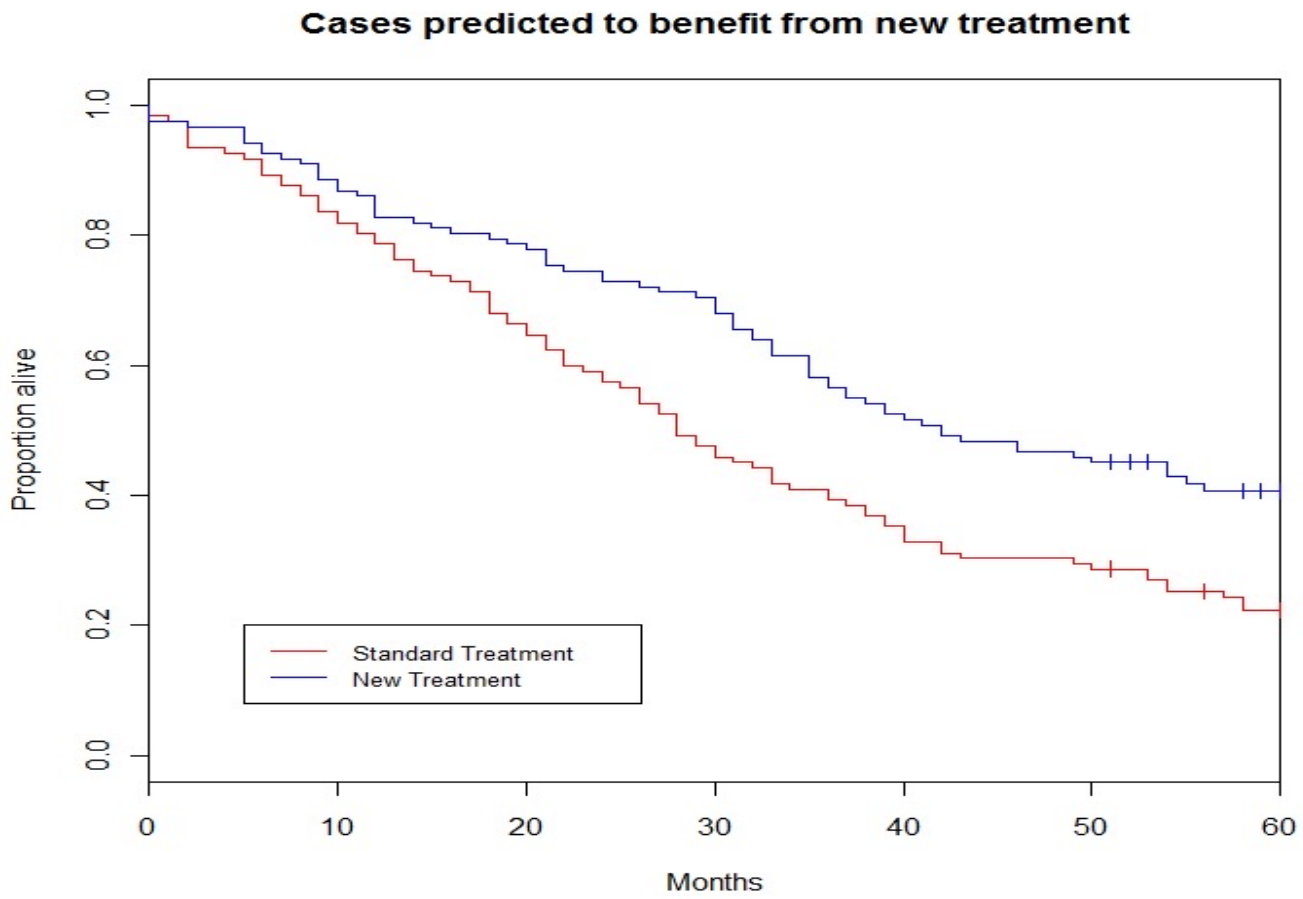


Figure 2: Cross-validated survival curves for patients predicted to benefit from the new treatment. log-rank statistic

= 10.0, permutation p-value is .002



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], John Crowley[3]

[1]Department of Data Science, The Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan; [2]Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Rockville, Maryland, USA; [3]Cancer Research And Biostatistics, Seattle, Washington, USA; [4]The Myeloma Institute for Research and Therapy, University of Arkansas for Medical Science, Little Rock, Arkansas, USA

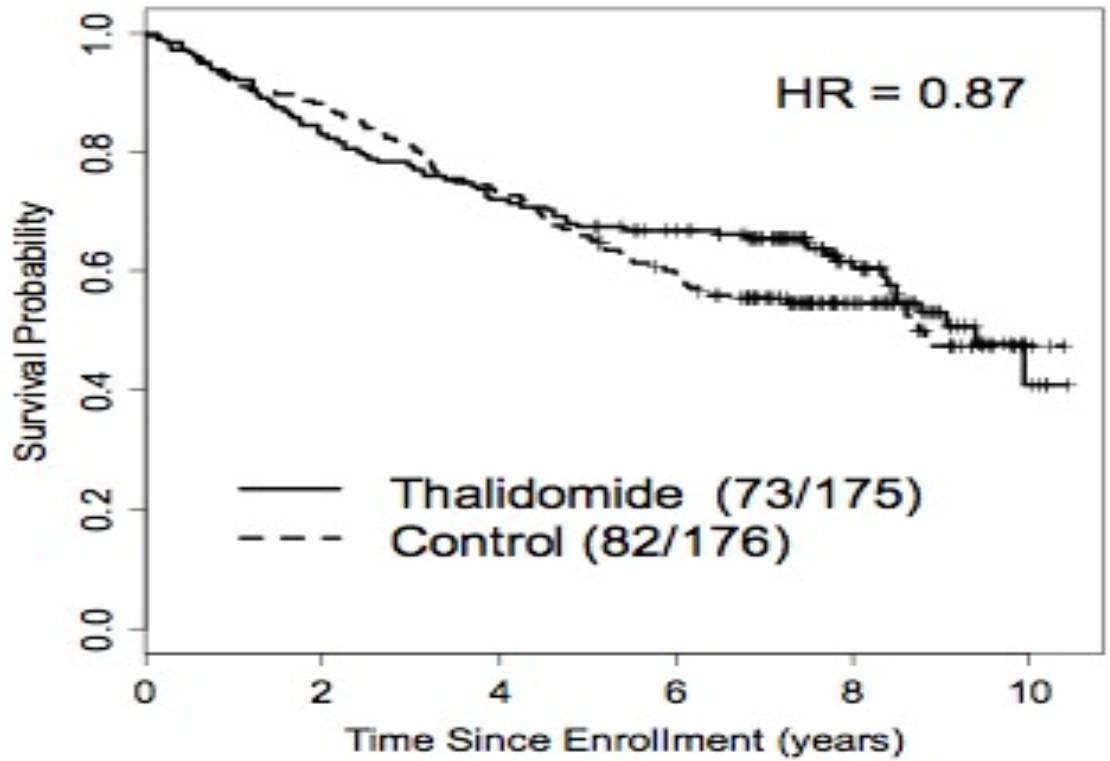


Fig. 1. Survival curves for all 351 patients with genomic data in the randomized trial for multiple myeloma.

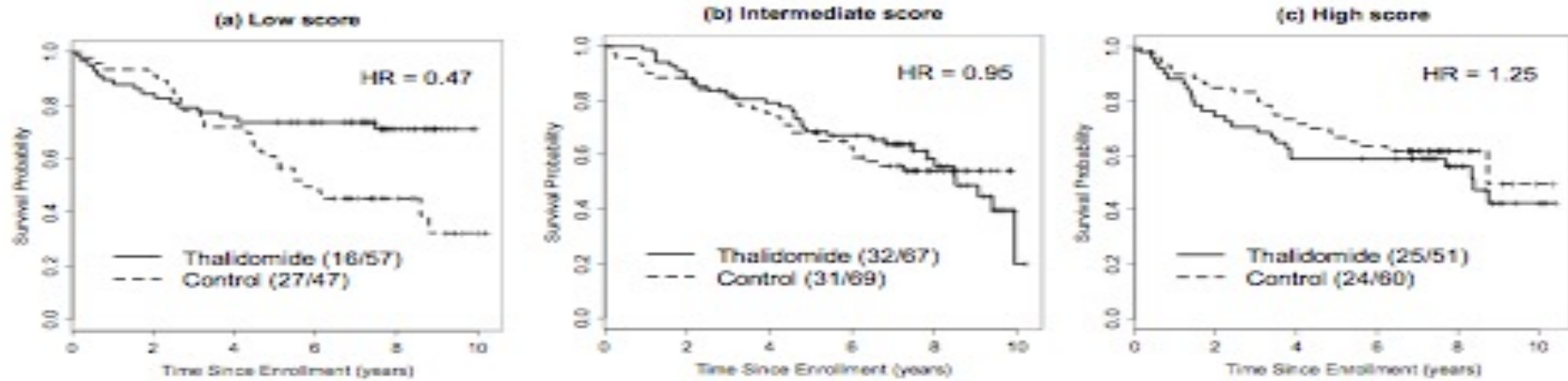


Fig. 5. Survival curves for each of the three subclasses, "Low", "Intermediate" and "High" derived from using thresholds of 33rd and 66th percentiles in the predicted signature score S (panels a-c).

Biostatistics (2013), **14**, 4, *pp.* 613–625
doi:10.1093/biostatistics/kxt010
Advance Access publication on March 21, 2013

Adaptive enrichment designs for clinical trials

NOAH SIMON*

Department of Statistics, Stanford University, Stanford, CA 94305, USA

nsimon@stanford.edu

RICHARD SIMON

Biometric Research Branch, National Cancer Institute, Bethesda, MD 20892, USA

SUMMARY

Modern medicine has graduated from broad spectrum treatments to targeted therapeutics. New drugs recognize the recently discovered heterogeneity of many diseases previously considered to be fairly homogeneous. These treatments attack specific genetic pathways which are only dysregulated in some smaller subset of patients with the disease. Often this subset is only rudimentarily understood until well into large-scale clinical trials. As such, standard practice has been to enroll a broad range of patients and run post hoc subset analysis to determine those who may particularly benefit. This unnecessarily exposes many patients to hazardous side effects, and may vastly decrease the efficiency of the trial (especially if only a small subset of patients benefit). In this manuscript, we propose a class of adaptive enrichment designs that allow the eligibility criteria of a trial to be adaptively updated during the trial, restricting entry to patients likely to benefit from the new treatment. We show that our designs both preserve the type 1 error, and in a variety of cases provide a substantial increase in power.

Keywords: Adaptive clinical trials; Biomarker; Cutpoint; Enrichment.

Adaptive Enrichment Designs

- Includes one or more interim analyses that may modify eligibility criteria based on candidate covariates
- Single significance test at final analysis
- All patients included in final analysis

Adaptive Enrichment Designs

- Single binary covariate
- Quantitative covariate
- Multiple candidate covariates
- Patient strata

- Δ_k =statistic for comparing outcome of treatment and control group of **all** patients who entered study during period k (k=1,2,…K).
  - Under null we assume that Δ_k has known distribution with mean 0 and independent of t_j for all j≠k.

- At end of trial, one significance test performed using test statistic $w_1 \Delta_1 + \ldots + w_K \Delta_K$

- There is no final subset analysis
- Power is gained by increasing alternative means of later Δ_k by restricting eligibility

- At each interim analysis time j, a decision is made concerning whether/how to change eligibility criteria for subsequent periods. That decision may use data accrued for patients who entered the trial up until the current interim analysis.

- The decision made at interim time j can be based on use of a “surrogate” endpoint instead of the endpoint to be used at the final analysis.
- The validity of the significance test performed at the end of the trial does not depend on the decision made at interim times concerning eligibility.
  - A Bayesian model for managing eligibility decisions can be used although the final significance test is frequentist (Frasian)

Biostatistics (2018) **19**, 1, *pp.* 27–41
doi:10.1093/biostatistics/kxw054
Advance Access publication on May 17, 2017

Using Bayesian modeling in frequentist adaptive enrichment designs

NOAH SIMON*

Department of Biostatistics, University of Washington, Box 357232, Seattle, WA 98195, USA
nrsimon@uw.edu

RICHARD SIMON

Biometric Research Branch of the National Cancer Institute (at the National Institutes of Health), 9609 Medical Center Dr, Rockville, MD 20850, USA

SUMMARY

Our increased understanding of the mechanistic heterogeneity of diseases has pushed the development of targeted therapeutics. We do not expect all patients with a given disease to benefit from a targeted drug; only those in the *target population*. That is, those with sufficient dysregulation in the biomolecular pathway targeted by treatment. However, due to complexity of the pathway, and/or technical issues with our characterizing assay, it is often hard to characterize the *target population* until well into large-scale clinical trials. This has stimulated the development of *adaptive enrichment trials*; clinical trials in which the target population is adaptively learned; and enrollment criteria are adaptively updated to reflect this growing understanding. This paper proposes a framework for group-sequential adaptive enrichment trials. Building on the work of Simon & Simon (2013). Adaptive enrichment designs for clinical trials. Biostatistics 14(4), 613–625), it includes a frequentist hypothesis test at the end of the trial. However, it uses Bayesian methods to optimize the decisions required during the trial (regarding how to restrict enrollment) and Bayesian methods to estimate effect size, and characterize the *target population* at the end of the trial. This joint frequentist/Bayesian design combines the power of Bayesian methods for decision making with the use of a formal hypothesis test at the end of the trial to preserve the studywise probability of a type I error.

Keywords: Adaptive enrichment; Bayesian statistics; Clinical trials.

Received: 24 August 2016 | Revised: 24 April 2017 | Accepted: 28 June 2017
DOI: 10.1002/sim.7422

FEATURED ARTICLE

WILEY Statistics in Medicine

Inference for multimarker adaptive enrichment trials

Richard Simon[1] | Noah Simon[2]

[1]Biometric Research Program, National Cancer Institute, Rockville, MD 20850, USA
[2]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA

Correspondence
Noah Simon, Department of Biostatistics, University of Washington, Seattle, WA 98195, USA.
Email: nrsimon@uw.edu

Funding information
NIH, Grant/Award Number: DP5OD019820

Identification of treatment selection biomarkers has become very important in cancer drug development. Adaptive enrichment designs have been developed for situations where a unique treatment selection biomarker is not apparent based on the mechanism of action of the drug. With such designs, the eligibility rules may be adaptively modified at interim analysis times to exclude patients who are unlikely to benefit from the test treatment.We consider a recently proposed, particularly flexible approach that permits development of model-based multifeature predictive classifiers as well as optimized cut-points for continuous biomarkers. A single significance test, including all randomized patients, is performed at the end of the trial of the strong null hypothesis that the expected outcome on the test treatment is no better than control for any of the subset populations of patients accrued in the K stages of the clinical trial.

In this paper, we address 2 issues involving inference following an adaptive enrichment design as described above. The first is specification of the intended use population and estimation of treatment effect for that population following rejection of the strong null hypothesis. The second issue is defining conditions in which rejection of the strong null hypothesis implies rejection of the null hypothesis for the intended use population.

KEYWORDS
adaptive clinical trials, biomarker, enrichment, resampling

1 | INTRODUCTION

Adaptive Threshold Enrichment

- Randomize patients without regard to value of biomarker but measure biomarker pre-randomization on all patients
- Pre-specify K candidate thresholds for the biomarker $B_1,\ldots,B_K$

Adaptive Threshold Enrichment

- Perform interim analysis using intermediate endpoint
- Find largest candidate cut-point B_k such that

$Pr[\Delta(B) \geq \Delta_a] < \varepsilon$ for $B < B_k$

where Δ_a is the treatment effect to be detected under the alternative hypothesis.

- Continue accrual only for patients with $B \geq B_k$

Simulation with Binary Response

- p_0= response probability for control group
- p_0= response probability for treatment group if $B<b^*$
- b^*= true cut-point
- p_1= response probability for treatment group if $B \geq b^*$
- K=number of candidate cut-points
- B is uniform on (0,1)
- One interim analysis

p_0=.2, p_1=.5, K=5, N_{tot}=200, all pts 100/yr

True cut-point	Power adaptive	Power non-adaptive	Accrual adaptive	Accrual non-adaptive
.25	.968	.955	2.55	2.25
.5	.897	.726	3.19	3.25
.67	.768	.424	3.97	4.75

Two period cut-point enrichment
Response to T =p_1 if B≥cut-point
B uniform on (0,1)

- Simulations show that adaptive enrichment can substantially increase the statistical power with adaptive threshold determination and multi-biomarker modeling

6. REDESIGN OF CETUXIMAB TRIAL

In addition to the previous simulations, we illustrate the approach to adaptive enrichment design described here using the clinical trial described by Bokemeyer *and others* (2009). The trial compared the standard of care chemotherapy regimen FOLFOX-4 to the same regimen with the addition of the anti-EGFR antibody cetuximab as first-line treatment for newly diagnosed patients with metastatic colorectal cancer. A total of 337 patients were randomized equally to the two treatments in a clinical trial involving 79 centers. The sample size was established to have 90% power at a 0.05 significance level for detecting an odds ratio of 2.33 for response rate comparing the two treatment groups overall. At the start of the trial there was uncertainty about the influence of EGFR expression or KRAS mutation on the probability of response to cetuximab.

The primary analysis of Bokemeyer *and others* (2009) gave a response rate for the cetuximab containing arm of 46% compared to 36% for the chemotherapy only control. The p-value reported was 0.064, interpreted as not significant at the 0.05 level.

We developed an adaptive enrichment design for this clinical trial that included EGFR expression level and KRAS mutation status as candidate predictive biomarkers. Response is modeled separately for the cetuximab and control groups using logistic regression. For details on how the parameters of the models were estimated from the published data see the supplemental material available at *Biostatistics* online.

Table 3. *Operating characteristics for the three redesign scenarios of Bokemeyer and others (2009), averaged over 1000 simulated trials, for two blocks with $n = 168$ patients per block. Each column is a different trial design: AC, FC (0.3), FC (0.5), FC (0.7) are all enrichment trials. These designs do not restrict entry in the first block. 'A-C' uses an adaptive cutpoint in the second block with candidate values $(0.3, 0.5, 0.7)$, 'F-C(η)' uses a fixed cutpoint with value η. 'Non-A' is a standard unenriched design. 'Sensitivity' and 'Specificity' are given for detecting the population which benefits. 'Effect size' is the true effect size for the designated population. 'Bias' is the amount by which each procedure over-estimates that effect size. 'root-MSE' is the root-mean-square error of the effect size estimate for the designated population. All operating characteristics, other than power and accrual time, are averaged only over successful trials*

		AC	Non-A	FC (0.3)	FC (0.5)	FC (0.7)
1	Power	0.68	0.23	0.62	0.69	0.73
	Sensitivity	0.99	1	0.99	0.99	0.99
	Specificity	0.92	0	0.92	0.93	0.89
	Accrual time	454	336	423	452	478
	Effect size	0.21	0.06	0.21	0.21	0.2
	Bias	0.04	0.07	0.05	0.04	0.04
	Root-MSE	0.06	0.08	0.07	0.06	0.06
2	Power	0.81	0.68	0.79	0.84	0.86
	Sensitivity	0.98	1	0.98	0.98	0.98
	Specificity	0.56	0	0.51	0.58	0.58
	Accrual time	412	336	376	406	442
	Effect size	0.18	0.13	0.18	0.18	0.18
	Bias	0.04	0.03	0.05	0.04	0.03
	Root-MSE	0.06	0.05	0.06	0.06	0.06
3	Power	0.38	0.03	0.22	0.37	0.45
	Sensitivity	1	1	1	1	1
	Specificity	0.76	0	0.83	0.8	0.74
	Accrual time	488	336	452	488	488
	Effect size	0.15	0.01	0.17	0.16	0.14
	Bias	0.08	0.12	0.09	0.07	0.08
	Root-MSE	0.09	0.12	0.1	0.09	0.09

Table 4. *Operating characteristics for the three redesign scenarios of Bokemeyer and others (2009), averaged over 1000 simulated trials, for three blocks with $n = 112$ patients per block. Each column is a different trial design: AC, FC (0.3,0.3),..., FC (0.5,0.7) are all enrichment trials. These designs do not restrict entry in the first block. 'A-C' uses* adaptive cutpoints *in the second and third blocks with candidate values* $(0.3, 0.5, 0.7)$. *'F-C(η_1,η_2)' uses* fixed cutpoints *with value η_1 in block 2 and η_2 in block 3. 'Non-A' is a standard unenriched design. 'Sensitivity' and 'Specificity' are given for detecting the population which benefits. 'Effect size' is the true effect size for the designated population. 'Bias' is the amount by which each procedure* over-estimates *that effect size. 'root-MSE' is the root-mean-square error of the effect size estimate for the designated population. All operating characteristics, other than power and accrual time, are averaged only over successful trials*

		AC	Non-A	FC(0.3,0.3)	FC(0.5,0.5)	FC(0.7,0.7)	FC(0.5,0.7)
1	Power	0.78	0.21	0.74	0.82	0.82	0.83
	Sensitivity	0.98	1	0.99	0.98	0.97	0.98
	Specificity	0.92	0	0.93	0.95	0.89	0.92
	Accrual time	511	336	461	503	552	525
	Effect size	0.21	0.06	0.21	0.22	0.2	0.21
	Bias	0.03	0.07	0.03	0.02	0.02	0.02
	Root-MSE	0.06	0.08	0.06	0.06	0.06	0.06
2	Power	0.86	0.65	0.8	0.86	0.87	0.88
	Sensitivity	0.97	1	0.98	0.98	0.96	0.97
	Specificity	0.62	0	0.58	0.67	0.64	0.67
	Accrual time	459	336	400	451	512	480
	Effect size	0.19	0.13	0.18	0.19	0.19	0.19
	Bias	0.03	0.03	0.04	0.03	0.02	0.03
	Root-MSE	0.06	0.05	0.06	0.06	0.06	0.05
3	Power	0.49	0.03	0.35	0.5	0.58	0.57
	Sensitivity	1	1	1	1	0.99	1
	Specificity	0.8	0	0.84	0.83	0.74	0.79
	Accrual time	544	336	502	556	564	569
	Effect size	0.16	0.01	0.17	0.17	0.14	0.15
	Bias	0.06	0.13	0.07	0.06	0.07	0.06
	Root-MSE	0.08	0.13	0.08	0.07	0.08	0.08

Determining which patients benefit from a new treatment once the study-wise null is rejected is a problem for all clinical trials, including standard broad eligibility trials. Serious adverse drug reactions (ADRs) from FDA approved drugs administered according to the labeling indication is a major medical problem. One study claimed that such ADRs represented the fourth to sixth leading cause of death in the United States (Lazarou *and others*, 1998). Using the eligible population in pivotal studies as the intended use population simplifies statistical significance testing but can result in many patients being exposed to the risk of ADR without likelihood of benefit. This is evident in the small effect sizes seen in many "positive" pivotal trials with broad eligibility. An approach to labeling based on a quantitative risk-benefit assessment that takes account of covariates effecting risk and benefit, rather than statistical significance for a population, might provide physicians more effective guidance. Although specifying an intended use population may be necessary for determining how the drug may be advertised, a more individualized based approach such as described in this paper for benefit may provide a framework for a more satisfactory way of communicating the uncertainties in benefits and risks.

Acknowledgements

- Noah Simon, U.W.
- Boris Freidlin, NCI
- Wenu Jiang, Queens U.
- Abubakar Maitournam, Niger
- Shigayaki Matsui, Nagoya U.