November 15 2017 Distinguished Lecture

Jerry Lawless, PhD, University of Waterloo

Dr. Lawless is Distinguished Professor Emeritus in the Department of Statistics and Actuarial Science at the University of Waterloo. He was a faculty member there from 1972-2007, serving as Chair from 1979-84, and in 1987-88 and 2004-05. His research interests include biostatistics methodology and modeling generally, event history analysis and regression methodology, and applications areas that include cancer, chronic disease and genetics. He is the author of numerous papers and the books Statistical Models and Methods for Lifetime Data (John Wiley and Sons, 1982; second edition, 2003) and The Statistical Analysis of Recurrent Events (with R.J. Cook, Springer, 2007). He is a past editor of Technometrics and has served as a consultant to government, industry and the legal profession. He is a recipient of the Gold Medal of the Statistical Society of Canada and of the Shewhart Medal of the American Society for Quality, is a Fellow of the American Statistical Association, the Institute of Mathematical Statistics and the Royal Society of Canada, and is an Honorary Member of the Statistical Society of Canada.

Title: Two-Phase outcome-dependent genetic association studies

Abstract:
Phase 1 clinical trials aim to identify the optimal dose for the therapeutic agent that balances patient safety and potential efficacy. Cancer therapies that include two or more agents may increase efficacy as well as toxicity. Many adaptive dose-escalation designs have been proposed for trials of combination therapies. These designs can better assign dose combinations near the maximum tolerated dose combination (MTDC) to enrolled patients but require significant resources to design and monitor. Hence, relatively few adult oncology trials have used these designs, and to our knowledge, none have been used in pediatric trials. To motivate the use of adaptive designs in pediatric oncology, we performed a simulation study to compare the performance of dual-agent dose-escalation methods in a pediatric oncology framework.

Stephen George, PhD, Duke University of Medicine

Dr. George is Professor Emeritus of Biostatistics in the Department of Biostatistics and Bioinformatics in the Duke University School of Medicine. He served for over 20 years as Director of Biostatistics in the Duke Comprehensive Cancer Center and Director of the Statistical Center of the Cancer and Leukemia Group B (CALGB). He has been closely involved in the design, conduct, and analysis of cancer clinical trials and other research projects in cancer throughout his career, and has published extensively on this research and other topics. He served for four years as the biostatistician for the Oncologic Drugs Advisory Committee for the Food and Drug Administration (FDA) and for several years as a Special Government Employee (Consultant) to the Oncologic Drug Products Division of the FDA. He has served on and chaired data monitoring committees for cancer treatment and prevention trials, both for government and industry. Dr. George has served on study sections and review committees for the National Cancer Institute and other institutes. He has served as an editorial board member or associate editor for several professional journals and regularly serves as reviewer for journals in the biomedical field and in statistics. Dr. George is a Fellow of the American Statistical Association and a Fellow of the Society for Clinical Trials and is a past President for the Society for

Clinical Trials. He has served in several capacities for other professional societies including eight years as Treasurer and Executive Committee member of the International Biometric Society, the American Statistical Association, the American Association for Cancer Research, and the American Society of Clinical Oncology.

Title: Enrichment Strategies for Biomarker Stratified Clinical Trials

Abstract: In large clinical trials using a biomarker stratified design, the cost of treating and following all patients for clinical outcomes may be prohibitive. With a fixed number of randomized patients, the efficiency of testing certain parameters can be improved by increasing the proportion of biomarker positives on study, especially when the prevalence rate of biomarker positives is low in the underlying patient population. When the cost of assessing the true biomarker is high, one can further improve the study efficiency by oversampling biomarker positives with a cheaper auxiliary variable or biomarker that correlates with the true biomarker. An enriched biomarker stratified design enriches the cohort of randomized patients by directly oversampling the relevant patients with the true biomarker, while an auxiliary-variable-enriched biomarker stratified design enriches the randomized cohort based on an inexpensive auxiliary variable. The latter design avoids testing the true biomarker on all screened patients and reduces treatment waiting time. For both designs, the optimal enrichment proportion is derived for testing a single hypothesis or two hypotheses simultaneously. Numerical studies and examples are presented.

Two-Phase Outcome-Dependent Genetic Association Studies

Jerry Lawless
University of Waterloo

jlawless@uwaterloo.ca

OICR Biostatistics Training Initiative
November 15, 2017

INTRODUCTION

- Context: studies of disease processes (occurrence, progression, treatment ...)
  - Limited budgets for expensive measurements.
  - Studies of rare outcomes based on random samples are not cost-effective because the proportion of persons experiencing the outcome is small; outcome-dependent sample selection is therefore often used.
- Two-phase study designs: in phase 1 observe certain data on N individuals; in phase 2 get additional data on a subset of n individuals. Sample selection for phase 2 may depend on the phase 1 data.
- Common situation: outcomes Y and covariates Z are observed in phase 1. Additional covariates X are measured on phase 2 individuals.
  - e.g. Case-control studies: Y indicates presence or absence of some condition: in phase 2 we randomly select n_1 persons with the condition ("cases", $Y = 1$) and n_0 persons without the condition ("controls", $Y = 0$) and measure exposure or risk factors X.

- In genetic epidemiology, an early paper: Whittemore and Halperin (1997). Multi-stage sampling in genetic epidemiology. *Statist. Med.* **16**, 153-167.

- Survival Time Outcomes: T = time to some event (e.g. cancer occurrence or recurrence); censoring - by a given time, only some persons will have experienced the event.

- Basic setup: phase 1 cohort of N individuals followed over time. At some point in time let C_i denote the length of followup for individual i, $Y_i = min(T_i, C_i)$ and let $\Delta_i = I(T_i \leq C_i)$. Covariates Z_i are also known.

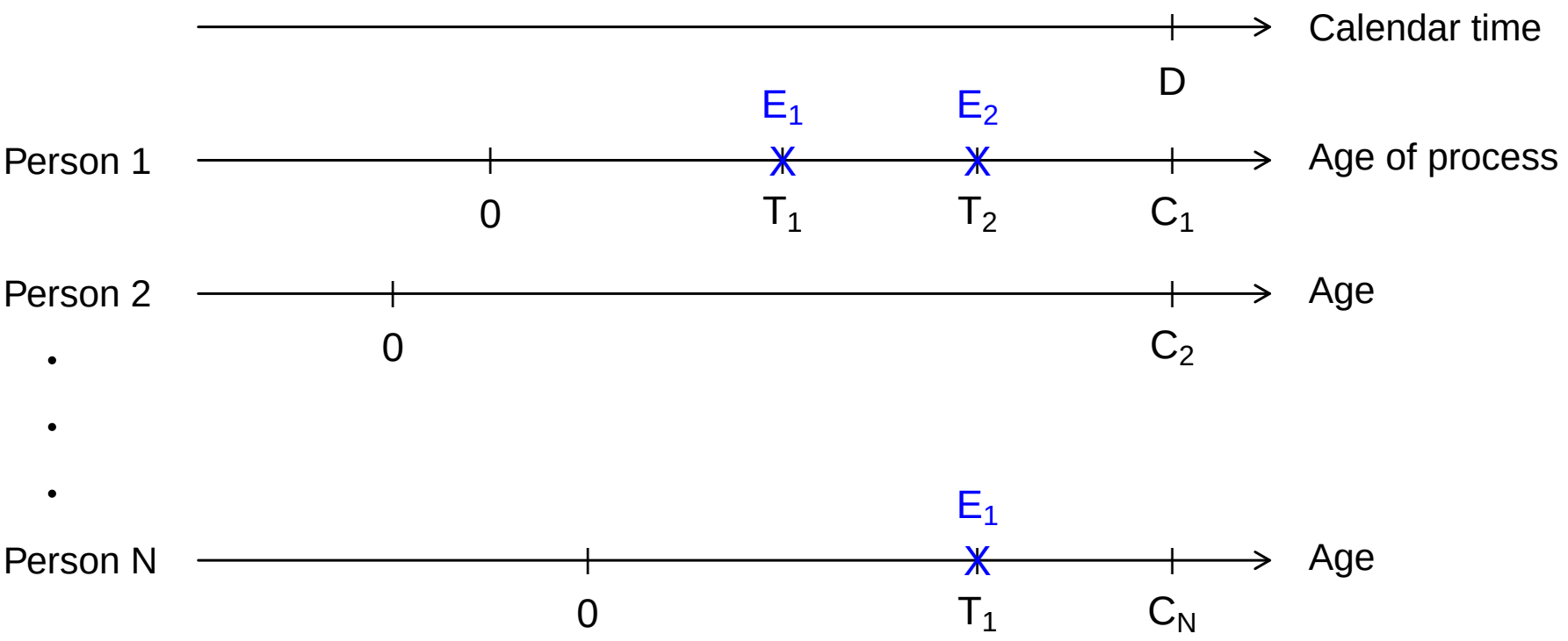
  Select a phase 2 sample of size n and measure X_i for those persons. Let $R_i = I(\text{person } i \text{ is selected})$; the study design specifies the sampling probabilities $\pi_i = P(R_i = 1 | Y_i, \Delta_i, Z_i)$.

- Includes case-cohort, nested case-control and other formats (e.g. Borgan and Samuelsen 2014). More generally, see the following graphic.

Cohorts with Disease History Information

E − some type of event

T − time or age when an event occurs



Examples

- Breslow et al. (2009): Atherosclerosis Risk in Communities (ARIC) study. Phase 1 - data on standard risk factors for coronary heart disease (CHD) and CHD outcomes for a cohort of about 16,000 individuals. A phase 2 sample consisted of 10-15 percent of the cohort, and candidate gene and biomarker data were obtained on each individual. Outcomes of interest are specific CHD events (e.g. stroke, heart attack).
- Sub-studies based on the Ontario Health Study (phase 1 data). For a study on cancer and heart disease (phase 2 data) select a sample consisting of 500 persons who have experienced a stroke or MI, 500 who have experienced cancer, and 1000 who have experienced neither. Sampling can be stratified by age or other known factors. Perform genotyping and RNA expression measurement on the phase 2 individuals. (D. Soave, P. Awadalla, OICR)

DESIGN AND ANALYSIS OF TWO-PHASE STUDIES

- At phase 1 suppose we have a response Y and covariate vector Z for each of N individuals. A sample of n individuals is selected for phase 2; for them we observe additional covariates X.
  - The objective is to fit regression models $f(y|x,z)$, conduct tests etc.
  - Other settings can also be considered (e.g. surrogate responses at phase 1).
- Stratified sampling: partition (Y,Z) into strata $S_1,\ldots,S_K$. Simple random samples of n_j individuals from S_j are selected for phase 2.
- Stratified case-cohort designs: sampling is based on (Δ_i, Z_i). When events are rare, we often take $\pi_i = P(R_i = 1|\Delta_i = 1, Z_i) = 1$ for cases.

Illustration (Breslow et al. 2009)

- Phase 1 cohort of 12,345 persons with data on age, sex, ethnicity and whether a person has had a CHD event; T = age at 1st CHD event.
- Base 8 covariate strata on: Age < 55, ≥ 55; Sex = Male, Female; Race = Black, White.
- Nine strata used for phase 2 sample selection (N_j = phase 1 stratum size):

  CHD = No $(\Delta = 0)$: Covariate strata 1-8 — CHD = Yes $(\Delta = 1)$: Stratum 9

  $N_1 \ldots N_8$ range from 393 to 2782 — N_9 = 730

- Phase 2 oversampled older ages, males and blacks plus persons with CHD. Phase 2 selection probabilities for strata $1, 2, ..., 9$ were

  .052-(F,B,A< 55), .075-(F,B,A≥ 55), .070-(M,B,A< 55), .181-(M,B,A≥ 55)
  .032-(F,W,A< 55), .070-(F,W,A≥ 55), .060-(M,W,A< 55), .081-(M,W,A≥ 55)

  for non-CHD $(\Delta = 0)$ strata 1, ..., 8 and .827 for CHD $(\Delta = 1)$ stratum 9.

Analysis: Notation, Assumptions and Models

- Use only phase 2 data or use phase 1 data also?
- Data $(Y_i, \Delta_i, X_i, Z_i), i = 1, \ldots, N$ for the phase 1 cohort are assumed iid, based on distribution $f(t|x,z)r(c|x,z)g(x,z)$ for (T, C, X, Z).
- (Y_i, Δ_i, Z_i) are observed in phase 1; X_i is observed only for individuals selected for the phase 2 sample V.
- Let $R_i = I(i \in V)$, $R = (R_1, \ldots, R_N)$, $Y = (Y_1, \ldots, Y_N)$ etc. We assume that

$$P(R|Y, \Delta, X, Z) = P(R|Y, \Delta, Z)\ . \qquad \text{[X is missing at random (MAR)]}$$

- The study design specifies selection probabilities $\pi_i = P(R_i = 1|Y, \Delta, Z)$.
- Parameterize the target model as $f(y|x, z; \theta)$; our focus is on inference for θ. However, $g(x, z)$ and the π_i are important for design, and for some analysis methods.

- Approaches to analysis include likelihood and weighted estimating function methods. Most weighted estimation methods use only data on persons in the phase 2 sample V. To make this valid, the contributions from each person $i \in V$ are given weights π_i^{-1}.
- For the semiparametric Cox model, weights are applied to the partial likelihood; see e.g. Samuelsen et al. (2007) and Kulathinal et al. (2007). Cox model software will handle estimation.
- Maximum likelihood (ML) uses all data and is most efficient. It handles studies where some persons have $\pi_i = 0$; weighted estimation does not do this.
- The price for ML is that a model for the distribution of covariates X, Z is needed and thus there is more dependence on assumptions. In addition, software does not handle common models at present.

ASIDE: MAXIMUM LIKELIHOOD

- Let

$$f_1(t|z;\theta,\delta) = \int f(t|x,z;\theta)dG(x|z;\delta)$$

- The following likelihood function for (θ,δ) holds for a wide range of two-phase sampling schemes:

$$\begin{aligned} L(\theta,\delta) \;=\; & \prod_{R_i=1} f(y_i|x_i,z_i;\theta)^{\Delta_i} S(y_i|x_i,z_i;\theta)^{1-\Delta_i} g(x_i|z_i;\delta) \\ & \times \prod_{R_i=0} f_1(y_i|z_i;\theta,\delta)^{\Delta_i} S_1(y_i|z_i;\theta,\delta)^{1-\Delta_i}, \end{aligned} \tag{1}$$

  where S and S_1 are the survivor functions corresponding to f and f_1. See for example Zhao et al. (2009), Zeng and Lin (2014). Note that the likelihood contributions are the standard censored data likelihood expressions.

- Note that this does not involve selection probabilities π_i.

ASIDE: WEIGHTED ESTIMATING FUNCTIONS

- Weighted estimating functions

$$U_W(\theta) = \sum_{i=1}^{N} \frac{R_i}{\pi(Y_i, \Delta_i, Z_i)} \frac{\partial \log L(\theta; Y_i, \Delta_i, X_i, Z_i)}{\partial \theta} \tag{2}$$

  where $L(\theta; Y_i, \Delta_i, X_i, Z_i)$ is the censored data likelihood for $i \in V$.

  - Improve by estimating $\pi(Y, \Delta, Z)$ or augmentation (AIPW: include $i \in \bar{V}$).

- Cox PH Models: Weighted Pseudo Likelihood
  For models with hazard function $\lambda(t|x, z) = \lambda_0(t) \exp(\beta' x + \gamma' z)$, maximize

$$PL_W(\theta) = \prod_{i \in V} \left\{ \frac{\exp(\beta' x_i + \gamma' z_i)}{\sum_{\ell=1}^{N} Y_\ell(t_i) \frac{R_\ell}{\pi_\ell} \exp(\beta' x_\ell + \gamma' z_\ell)} \right\}^{\Delta_i R_i / \pi_i}, \tag{3}$$

  where $Y_\ell(t) = I(Y_\ell \geq t) = I(T_\ell \geq t, C_\ell \geq t)$.

ILLUSTRATION: TESTS OF ASSOCIATION FOR EXPENSIVE COVARIATES

- Objective: test the effect of X in $f(y|x,z)$, where X represents genotypes or biomarkers. See e.g. Barnett et al. (2013), and Derkach et al. (2015) for two-phase studies with uncensored responses.
- Suppose survival time T has conditional density function of the form

$$f(t|x,z;\theta) = f_0(t|\mu(x,z);\theta) \tag{4}$$

where $\mu(x,z) = \beta_0 + \beta' x + \gamma' z$, f_0 is a known function and $\theta = (\beta_0, \beta, \gamma, \sigma)$, with σ a vector of scale or shape parameters.

  - Includes proportional hazards, accelerated failure time, proportional odds and other models.
  - We want to test $H_0 : \beta = 0$; I'll consider semi-parametric ML.

- Let $\phi_0'(y, \Delta|\mu,\sigma) = \Delta\ \partial \log f_0(y|\mu,\sigma)/\partial\mu + (1-\Delta)\ \partial \log S_0(y|\mu,\sigma)/\partial\mu$. Letting $\hat{\beta}_0, \hat{\gamma}, \hat{\sigma}$ be the MLEs under H_0 from the phase 1 data, the likelihood score statistic for testing H_0 when X and Z are independent reduces to (Derkach et

al. 2015, Lawless 2017)

$$U = \sum_{i \in V} \phi_0'(y_i, \Delta_i | \hat{\mu}_i, \hat{\sigma})(x_i - \bar{x}) \tag{5}$$

where $\hat{\mu}_i = \hat{\beta}_0 + \hat{\gamma}' z_i$ and $\bar{x} = \sum_{i \in V} x_i / n$.

- Note that $\phi_0'(y_i, \Delta_i | \hat{\mu}_i, \hat{\sigma})$ is a score residual and that the test statistic U has the same form as for a random sample V.
- When X, Z are not independent (5) does not apply; for discrete Z we get

$$U = \sum_{i \in V} r_i x_i + \sum_{i \in \bar{V}} r_i \bar{x}(z_i)$$

where $\bar{x}(z_i)$ is the mean X-value for individuals in V with $Z_i = z_i$.

- These tests are easily implemented for models with existing software.
- The case when X, Z are related and Z has continuous elements is more difficult. Zeng and Lin (2014) consider kernel density estimation of $g(x|z)$ but this is not feasible when dimension(Z) exceeds 2 or 3 or when X is high-dimensional.

Illustration: ANN Breast Cancer Study (S. Bull, Y. Yilmaz)

- Phase 1: cohort of women, followed post treatment for over 20 years (median FU about 8.5 years).

  T = time to distant recurrence

  Z = traditional prognostic factors, including lymphatic invasion (LVI)

  About 85-90% censored, i.e. recurrences in 10-15% of women.

- Phase 2: X - Tissue microrarray measurement of protein expression
  Phase 2 sampling - stratify full cohort according to (Y, Δ) and LVI positive or negative

- Cure rate survival models have been found important, since the majority of women will never experience recurrence.

- Best sampling plan? Sample all "cases" ($\Delta_i = 1$) and some non-cases ($\Delta_i = 0$).

Sample simulation

- $X = I$(protein expressed); X is Bin(1,0.2). Covariates Z_1, Z_2 are Bin(1,0.6) and max(0, Normal(2.5,1)) respectively.
- T is time to recurrence, in years; $T|X, Z_1, Z_2$ is Weibull (shape = 0.5, scale = $exp(-2(\gamma_0 + \gamma' Z + \beta' X))$. C is Uniform(5,30).
- Regression coefficients are chosen such that $S(10|X, Z)$ is in range 0.87 - 0.97 and $S(20|X, Z)$ is in range 0.76 - 0.94.
- Consider $N = 2000, n = 1000$. Sampling all N_1 individuals with $\Delta = 1$ and $n_0 = 1000 - N_1$ individuals from those with $\Delta = 0, C > 15$ years performs slightly better than randomly selecting n_0 non-cases.

  - With $\alpha = 0.05, exp(\beta) = 1.5$: two-phase power ($n = 1000$) is 0.703 versus 0.734 if X was available for the full cohort ($N = n = 2000$).

- Challenges in the real setting: X and Z are not independent, and Z may include 10-20 variables.

EXTENSIONS AND DISCUSSION

- Various application scenarios: e.g. cases where Z is a surrogate for X (or Y).
  Z - GWAS-based SNP genotypes (observed or imputed)
  X - genotypes from fine mapping
- Gaps in methodology: high-dimensional or continuous covariates; assessment of robustness, model adequacy; predictive modeling; software.
- Design: Large literature on types of study design (e.g. case-control, case-cohort, nested case-control) under weighted estimation, but less on ML.
  - See Ding et al. (2016) for a review emphasizing PH and additive hazards models.
- More complex studies: outcomes include recurrent events and multiple event-types. For example, studies on a range of diseases and related events.
- Models are more complex; maximum likelihood is generally difficult to apply, but inverse probability-weighted pseudo likelihood is usually straightforward.

- General life history processes: phase 1 data D_i observed over some time interval (A_i, C_i) for individual i. Based on these D_i, select n persons for phase 2 data collection. (Remark: vital status as a selection factor.)
- Recurrent or multiple events: at a given calendar time, stratify individuals in the full cohort according to number or types of events, and Z_i.
  - Oversampling persons with (more) events makes sense, but allowance should be made for variable lengths of followup $C_i - A_i$.
  - e.g. Cohort of persons with recurrent skin cancers (A. Whittemore).
- Multistate models: states $1, \ldots, K$, where $Y_i(t)$ is the state occupied at time t.
  - At a given time A_i, stratify individuals according to process history.
  - e.g. Progression of retinopathy in a cohort with Type 1 diabetes (Andrew Paterson; Cook and Lawless 2014): oversample persons with slow or fast progression through increasingly severe states of retinopathy.
- Final remark: great opportunities to base sub-studies on individuals in administrative data bases and other big data bases.

SELECTED REFERENCES

I Barnett, S Lee, X Lin (2013). Detecting rare variant effects using extreme phenotype sampling in sequencing association studies (2013). *Genet. Epidemiol.* **37**, 142-151.

O Borgan, S Samuelsen (2014). Nested case-control and case-cohort studies. *Handbook of Survival Analysis*. Chapman and Hall/CRC Press, pages 343-367.

N Breslow et al. (2009). Using the whole cohort in the analysis of case-cohort data. *Amer. J. Epidemiology* **169**, 1398-1405.

R Cook, J Lawless (2014). Statistical issues in modeling chronic disease in cohort studies. *Statistics in Biosciences* **6**, 127-161.

A Derkach, J Lawless, L Sun (2015). Score tests for association under response-dependent sampling designs for expensive covariates. *Biometrika* **102**, 988-994.

J Ding, T-S Lu, J Cai, H Zhou (2017). Recent progresses in outcome-dependent sampling with failure time data. *Lifetime Data Analysis*, **23**, 57-82.

S Kang, J Cai (2009). Marginal hazard model for case-cohort studies with multiple disease outcomes. *Biometrika* **96**, 887-901.

S Kulathinal et al. (2007) Case-cohort design in practice - experiences from the MORGAM project. *Epidemiol. Perspectives and Innov.*, **4:15**.

J Lawless (2017). Two-phase outcome-dependent studies for failure times and testing for effects of expensive covariates. To appear in LIDA; doi: 10.1007/s10985-016-9386-8.

J Lawless, J Kalbfleisch, C Wild (2009). Semiparametric methods for response-selective and missing data problems in regression. *J. Roy. Statist. Soc. B* **61**, 413-438.

O Saarela, S Kulathinal, E Arjas, E. Laara (2008). Nested case-control data utilized for multiple outcomes: a likelihood approach and alternatives. *Statist. Med.* **27**, 5991-6008.

S Samuelsen, H Anestad, A Skrondal (2007). Stratified case-cohort analysis of general cohort sampling designs. *Scand. J. Statist.* **34**, 103-119.

Y Zhao, J Lawless, D McLeish (2009). Likelihood methods for regression models with expensive variables missing by design. *Biomet. J.* **51**, 123-136.

D Zeng, D Lin (2014). Efficient estimation of semiparametric transformation models for two-phase cohort studies. *J. Amer. Statist. Assoc.* **109**, 371-383.

Enrichment Strategies for Biomarker Stratified Cancer Trials

Stephen L. George

Department of Biostatistics & Bioinformatics
Duke University School of Medicine

Ontario Institute for Cancer Research

November 15, 2017

joint work with Xiaofei Wang, Ting Wang, Haibo Zhou, Jianwen Cai and Kathy Zhou

Background

This talk is based primarily on the following two papers:

- Wang, X., Zhou, J., Wang, T., and George, S. L. (2017). On enrichment strategies for biomarker stratified clinical trials. *Journal of Biopharmaceutical Statistics* **21**: 1-17.

- Wang, T., Wang, X., Zhou, H., Cai, J., and George, S. L. (2017). Auxiliary-variable-enriched biomarker stratified design. (submitted).

Outline

Biomarkers

- In this talk, the term “biomarker” or “marker” refers generically to a distinct pattern of demographic, clinical, imaging, pathological, molecular or genetic data.
- Prognostic biomarker - provides information about a patient’s outcome, regardless of therapy.
- Predictive biomarker - provides information about the effect of a therapeutic intervention and the potential to tailor treatments to individuals based on the value of the biomarker.

Example: Companion Diagnostics for Biomarkers in NSCLC

Companion Diagnostic	Approval year	Biomarker	Drug/Agent
PD-L1 IHC	2015	PD-L1	nivolumab
Therascreen EGFR	2015	EGFR	gefitinib
Ventana ALK	2015	ALK	crizotinib
Cobras EGFR	2016	EGFR	erlotinib
Ventana PD-L1	2016	PD-L1	ateziolizumab
Dako 22C3	2016	PD-L1	pembrolizumab
Ventana ALK	2017	ALK	ceritinib
Oncomine Dx	2017	BRAF ROS1 EGFR	dabrafenib/trametinib crizotinib gefitinib

Biomarker-Driven Clinical Trials

- Targeted design (biomarker-positive only)
- Biomarker stratified design (all comers)
- Biomarker strategy design

See Mandrekar and Sargent (2009) and Freidlin et al. (2010) for reviews.

What is "Enrichment"?

A deliberate selection of a biased sample with a higher proportion of patients with certain characteristics (e.g., those who are marker-positive).

Why enrich?

- Efficiency (fewer patients required)
- Lower cost

Notation

- Treatment D
  $D = 1$ for experimental therapy E
  $D = 0$ for control therapy C

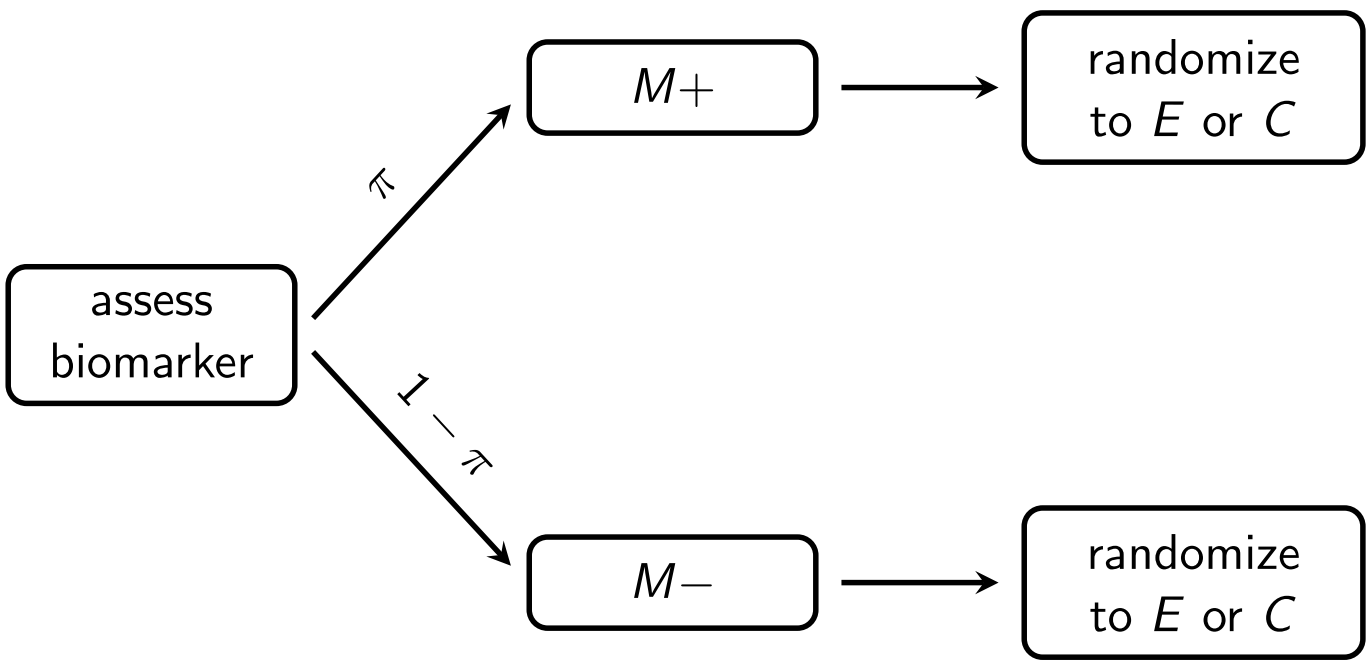
- Binary marker M
  $M = 1$ or $M+$ for marker positive
  $M = 0$ or $M-$ for marker negative

- Prevalence (π) of $M+$ patients
  $\pi = P(M = 1)$

Measures of Treatment Effects

- Treatment effect in $M+$ patients
- Treatment effect in $M-$ patients
- Overall treatment effect
- Interaction between treatment D and M
- Difference in clinical benefit between a biomarker-guided and biomarker-unguided approach

Biomarker Stratified Design (BSD)



BSD Drawbacks

The BSD design allows testing of treatment effects in biomarker-defined subgroups and in all patients but:

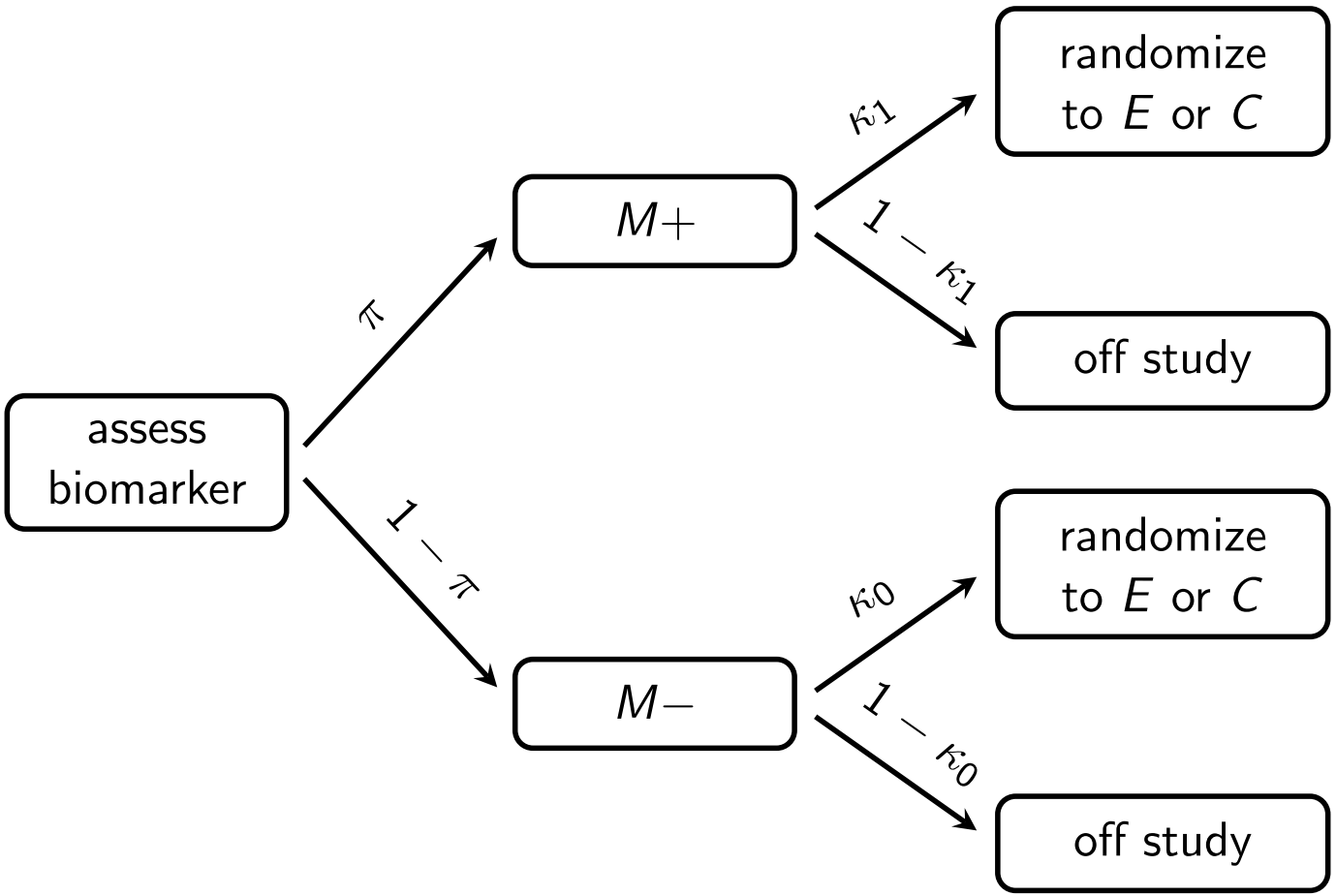
- It is inefficient if there is a low prevalence of biomarker-positive patients
- It is costly since the true biomarker status must be ascertained on all patients prior to randomization
- An excessive number of $M-$ patients, those who may have less benefit, are required to be treated

Two Possible Enrichment Strategies

- Enrichment through true biomarker M (EBSD)
- Enrichment through auxiliary biomarker (variable) $\tilde{M}$ (AEBSD)

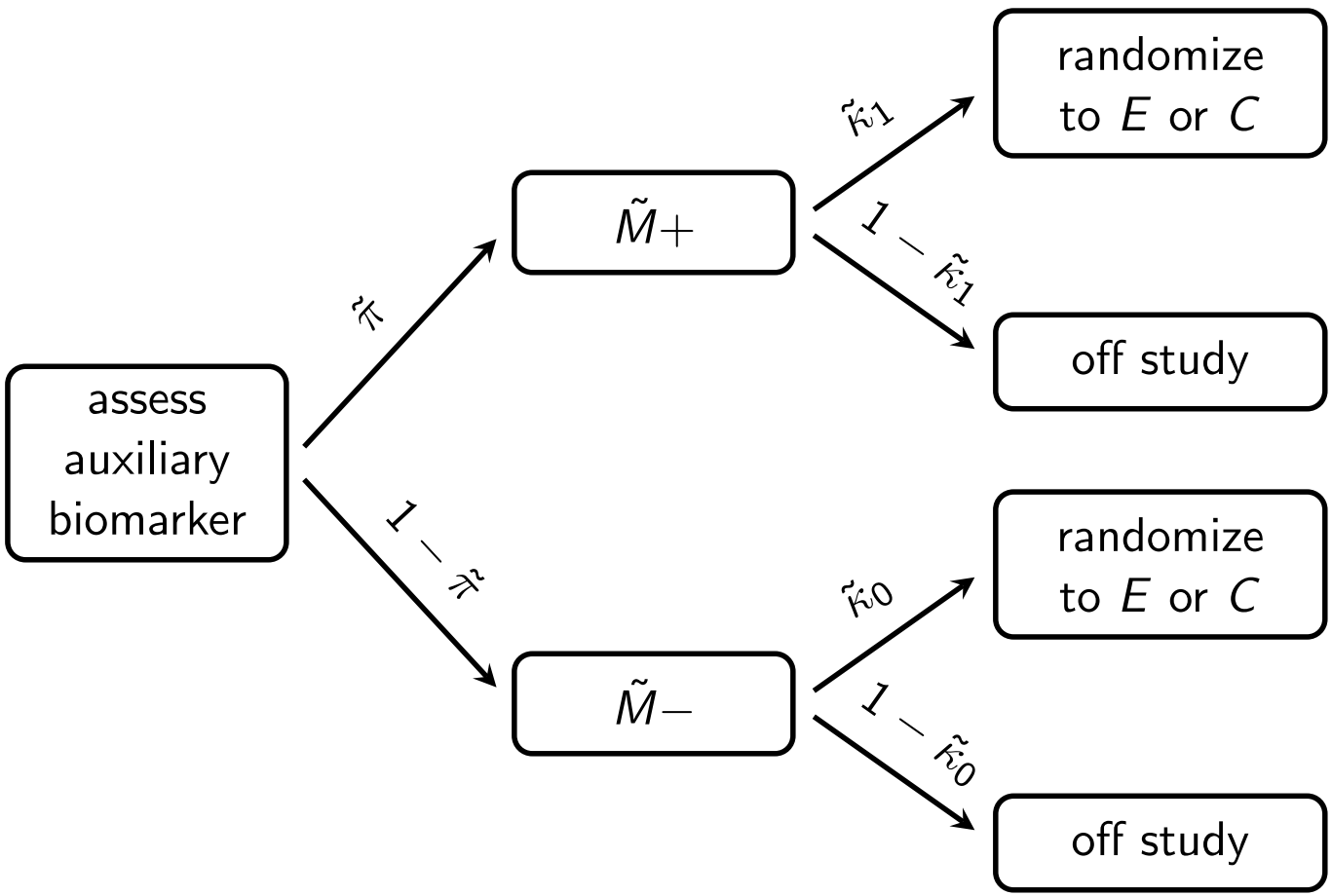
  An auxiliary variable is one that is:
  - inexpensive to obtain relative to the true biomarker
    e.g. smoking status, histology, gender, race, routine lab tests, etc.
  - positively correlated to the true biomarker

Enriched Biomarker Stratified Design (EBSD)



κ_1 and κ_0 are the selection probability for biomarker positives and biomarker negatives into the randomized cohort.

Auxiliary-Enriched Biomarker Stratified Design (AEBSD)

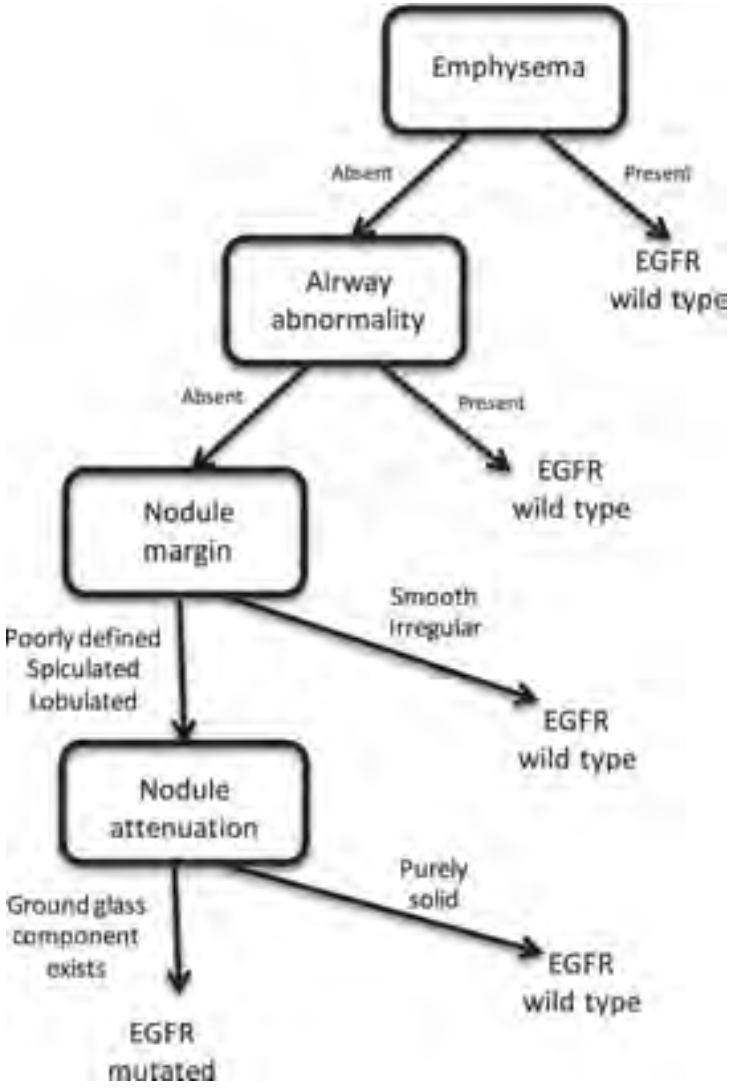


$\tilde{\pi}$ is the prevalence of auxiliary positives in the population; $\tilde{\kappa}_1$ and $\tilde{\kappa}_0$ are the selection probability for auxiliary positives and auxiliary negatives into the randomized cohort. The true marker M is assessed only for the randomized patients.

Auxiliary Variable for EGFR Mutation in NSCLC Patients Based on Demographic and Clinical Variables

- Sex (Females)
- Smoking history (Never)
- Ethnicity (Asian)
- Histopathology (adenocarcinoma)
- Tumor grade (well differentiated)
- Primary tumor size (< 5 cm)
- Stage (IV)

Auxiliary Variable for EGFR Mutation Based on CT Scan Gevaert et al. (2017)



Auxiliary Variable for EGFR Mutation Based on Plasma Li et al (2014)

	Tumor mutation	Tumor wild-type	Total
Plasma mutation	27	3	30
Plasma wild-type	29	62	91
Total	56	65	121

$PPV = \frac{27}{30} = 0.90, \quad NPV = \frac{62}{91} \approx 0.68,$

$\pi = \frac{56}{121} \approx 0.46, \quad \tilde{\pi} = \frac{30}{121} \approx 0.25$

AEBSD Concepts

- Use the value of $\tilde{M}$ to select patients for randomization with probability $\tilde{\kappa}_1$ for $\tilde{M}+$ patients and $\tilde{\kappa}_0$ for $\tilde{M}-$ patients.
- The true marker M is assessed only only for patients selected for randomization.
- This will result in increased efficiency as well as reduced costs, including savings in the ascertainment costs of the marker M.

PPV, π and ρ

It can be shown that

$$PPV = \pi + \rho \frac{\sqrt{\pi(1-\pi)\tilde{\pi}(1-\tilde{\pi})}}{\tilde{\pi}} \tag{1}$$

where $\rho = Corr(M, \tilde{M})$.

Thus, if $\pi \approx \tilde{\pi}$, then $PPV \approx \pi + \rho(1-\pi)$

Expected Number of Screened Patients

Let $S = 1$ if selected for randomization, where

$$P(S = 1) = \tilde{\pi}\tilde{\kappa}_1 + (1 - \tilde{\pi})\tilde{\kappa}_0$$

Then, the expected number of screened patients (n_s) necessary to achieve a fixed number (n) of randomized patients is

$$\begin{aligned} n_s &= \frac{n}{P(S = 1)} \\ &= \frac{n}{\tilde{\pi}\tilde{\kappa}_1 + (1 - \tilde{\pi})\tilde{\kappa}_0} \end{aligned} \tag{2}$$

Enriched proportion of $M+$ patients

The enriched proportion of $M+$ patients is

$$\tilde{\pi}_e = Pr(M = 1|S = 1) = \frac{\tilde{\pi}\tilde{\kappa}_1 PPV + \tilde{\kappa}_0(\pi - \tilde{\pi} PPV)}{\tilde{\pi}\tilde{\kappa}_1 + (1 - \tilde{\pi})\tilde{\kappa}_0} \tag{3}$$

If $\tilde{\kappa}_1 = 1$, then (3) is

$$\tilde{\pi}_e = \pi + \frac{\tilde{\pi}(PPV - \pi)(1 - \tilde{\kappa}_0)}{\tilde{\pi}(1 - \tilde{\kappa}_0) + \tilde{\kappa}_0} \tag{4}$$

$$\Rightarrow \tilde{\pi}_e \in [\pi, PPV]$$

Why We Always Choose $\tilde{\kappa}_1 = 1$

For any $\tilde{\pi}_e \in [\pi, PPV]$, we can write

$$\tilde{\pi}_e = \frac{\tilde{\pi}\tilde{\kappa}_1 PPV + \tilde{\kappa}_0(\pi - \tilde{\pi} PPV)}{\tilde{\pi}\tilde{\kappa}_1 + (1 - \tilde{\pi})\tilde{\kappa}_0}$$

Any pair $(\tilde{\kappa}_0, \tilde{\kappa}_1)$ satisfying $\tilde{\kappa}_0 = k\tilde{\kappa}_1$ will work, where

$$k = \frac{\tilde{\pi}(PPV - \tilde{\pi}_e)}{\tilde{\pi}(PPV - \tilde{\pi}_e) + (\tilde{\pi}_e - \pi)}$$

$\Rightarrow$

The solution that maximizes the number of randomized patients is $(k, 1)$.

Model

Proportional hazard model:

$$\lambda(t) = \lambda_0(t) exp(\beta_1 M + \beta_2 D + \beta_3 MD) \tag{5}$$

- $\lambda_0(t)$ baseline hazard function
- M biomarker indicator (0 marker-negative; 1 marker-positive)
- D treatment indicator(0 control; 1 experimental treatment)
- MD interaction term

Testing Different Treatment Parameters

Case	H_0	v.s.	H_a	Interpretation
1	$log\Delta_1 = 0$		$log\Delta_1 \neq 0$ [1]	Treatment effect on $M+$
2	$log\Delta_{overall} = 0$		$log\Delta_{overall} \neq 0$ [2]	Overall effect
3	$log\Delta_I = 0$		$log\Delta_I \neq 0$ [3]	Interaction effect
4	$log\Delta_1 = 0$ $log\Delta_I = 0$		$log\Delta_1 \neq 0$ $log\Delta_I \neq 0$	Treatment effect on $M+$ Interaction
5	$log\Delta_1 = 0$ $log\Delta_{overall} = 0$		$log\Delta_1 \neq 0$ $log\Delta_{overall} \neq 0$	Treatment effect on $M+$ Overall treatment effect

[1] $log\Delta_1 = \beta_2 + \beta_3$ in equation (5)

[2] $log\Delta_{overall} = \pi log\Delta_1 + (1-\pi)log\Delta_0 = \beta_2 + \pi\beta_3$ in equation (5)

[3] $log\Delta_I = \beta_3$ in equation (5)

Scenario 1: Testing the Treatment Effect in $M+$ Patients

- In equation (1), the treatment effect on $M+$ patients is $log\Delta_1 = \beta_2 + \beta_3$.
- $H_0 : log\Delta_1 = 0$ vs. $H_a : log\Delta_1 = log\Delta_1^*$
- The power can be written as

$$Pw = 1 - Pr(\left|\frac{log\hat{\Delta}_1}{se(log\hat{\Delta}_1)}\right| \leq z_{\alpha/2} | log\Delta_1 = log\Delta_1^*) \tag{6}$$

- The variance of $log\hat{\Delta}_1$ is given as (George and Desu, 1974; Schoenfeld, 1983),

$$var(log\hat{\Delta}_1) = \frac{1}{m_{1C}} + \frac{1}{m_{1E}} = \frac{2}{n\tilde{\pi}_e}(\frac{1}{p_{1C}} + \frac{1}{p_{1E}}) \tag{7}$$

where m_{1j} = number of events in treatment group j,
p_{1j} = probability of an event in treatment group j

Scenario 1: Testing the Treatment Effect in $M+$ Patients

- For the expected accrual time $T = \frac{n_s}{a}$ and follow-up time τ, p_{1j} can be written as (George and Desu, 1974):

$$p_{1j} = 1 - \frac{a}{n_s \lambda_{1j}} e^{-\tau \lambda_{1j}} (1 - e^{-n_s \lambda_{1j}/a}), \quad j = C, E \tag{8}$$

  where λ_{1j} = hazard function for treatment j for $M+$ patients

- From equation (7), we need to maximize $\tilde{\pi}_e$ to minimize the variance.

$$\Rightarrow \text{optimal } \tilde{\pi}_e = PPV \text{ and optimal } \tilde{\kappa}_0 = 0.$$

Scenario 2: Testing the Overall Treatment Effect

- $log\Delta_{overall} = \pi log\Delta_1 + (1-\pi) log\Delta_0$ where
  $log\Delta_1 = \beta_2 + \beta_3$ is the treatment effect on $M+$, and
  $log\Delta_0 = \beta_2$ is the treatment effect on $M-$,

- $H_0 : log\Delta_{overall} = 0$ v.s. $H_a : log\Delta_{overall} = log\Delta^*_{overall}$

- The power can be written as

$$Pw = 1 - Pr(|\frac{\widehat{log\Delta_{overall}}}{\widehat{se(log\Delta_{overall})}}| \leq z_{\alpha/2} | log\Delta_{overall} = log\Delta^*_{overall}) \quad (9)$$

Scenario 2: Testing the Overall Treatment Effect

- The variance of $\widehat{log\Delta_{overall}}$ is

$$
\begin{aligned}
var(\widehat{log\Delta_{overall}}) &= \pi^2 var(\widehat{log\Delta_1}) + (1-\pi)^2 var(\widehat{log\Delta_0}) \\
&\quad + 2\pi(1-\pi) cov(\widehat{log\Delta_1}, \widehat{log\Delta_0}) \\
&= \pi^2 var(\widehat{log\Delta_1}) + (1-\pi)^2 var(\widehat{log\Delta_0}) \\
&= \pi^2 (\frac{1}{m_{1C}} + \frac{1}{m_{1E}}) + (1-\pi)^2 (\frac{1}{m_{0C}} + \frac{1}{m_{0E}}) \\
&= \frac{2\pi^2}{n\tilde{\pi}_e} (\frac{1}{p_{1C}} + \frac{1}{p_{1E}}) + \frac{2(1-\pi)^2}{n(1-\tilde{\pi}_e)} (\frac{1}{p_{0C}} + \frac{1}{p_{0E}})
\end{aligned}
\tag{10}
$$

$\Rightarrow$ optimal $\tilde{\pi}_e = \pi$ and optimal $\tilde{\kappa}_0 = 1$.

Scenario 3: Testing the Interaction Effect

- In expression (1), the interaction effect between treatment and biomarker is $log\Delta_I = \beta_3$.
- $H_0 : log\Delta_I = 0$ v.s. $H_a : log\Delta_I = log\Delta_I^*$
- The power can be written as

$$Pw = 1 - Pr(\left| \frac{\widehat{log\Delta_I}}{se(\widehat{log\Delta_I})} \right| \leq z_{\alpha/2} | log\Delta_I = log\Delta_I^*) \tag{11}$$

- The variance of $\widehat{log\Delta_I}$ is (Peterson and George, 1993)

$$\begin{aligned} var(\widehat{log\Delta_I}) &= \frac{1}{m_{1C}} + \frac{1}{m_{1E}} + \frac{1}{m_{0C}} + \frac{1}{m_{0E}} \\ &= \frac{2}{n\tilde{\pi}_e}(\frac{1}{p_{1C}} + \frac{1}{p_{1E}}) + \frac{2}{n(1-\tilde{\pi}_e)}(\frac{1}{p_{0C}} + \frac{1}{p_{0E}}) \end{aligned} \tag{12}$$

Scenario 3: Testing the Interaction Effect

- Unlike Scenarios 1 and 2, there is no simple expression for the optimal $\tilde{\kappa}_0$; a search algorithm is required.
- Let $n_3(\tilde{\kappa}_0)$ represent the required minimum number of randomized patients to achieve the specified power for testing the interaction effect.
- Search all $\tilde{\kappa}_0$s to find the optimal $\tilde{\kappa}_0$, the one yielding the minimum $n_3(\tilde{\kappa}_0)$.

Scenario 4: Testing the Treatment Effect in $M+$ Patients and the Interaction Effect

- In Scenario 4, we jointly test the two hypotheses in Scenarios 1 ($M+$) and 3 (interaction).
- Let $n_1(\tilde{\kappa}_0)$ and $n_3(\tilde{\kappa}_0)$ represent the required minimum number of randomized patients for Scenarios 1 and 3 respectively for a given $\tilde{\kappa}_0$.
- Search all $\tilde{\kappa}_0$s to find the minimum of $max(n_1(\tilde{\kappa}_0), n_3(\tilde{\kappa}_0))$,
- The result is the optimal $\tilde{\kappa}_0$, the one yielding the minimum required number of patients for jointly testing both hypotheses.

Scenario 5: Testing the Treatment Effect in $M+$ Patients and the Overall Treatment Effect

- The approach for Scenario 5 (testing in $M+$ and overall) is analagous to that for Scenario 4 *mutatis mutandis*.
- Let $n_1(\tilde{\kappa}_0)$ and $n_2(\tilde{\kappa}_0)$ represent the minimum required number of randomized patients for Scenarios 1 and 2 respectively for a given $\tilde{\kappa}_0$.
- Search all $\tilde{\kappa}_0$s to find the minimum of $max(n_1(\tilde{\kappa}_0), n_2(\tilde{\kappa}_0))$,
- The result is the optimal $\tilde{\kappa}_0$, the one yielding the minimum required number of patients for simultaneously testing both hypotheses.

Cost Comparison of AEBSD and BSD Designs

- The cost of AEBSD:

$$C_{AEBSD} = (C_D + C_M)n_{AEBSD} + C_f T_{f,AEBSD} + C_s n_s$$

- The cost of BSD:

$$C_{BSD} = (C_D + C_M)n_{BSD} + C_f T_{f,BSD}$$

C_D the treatment cost of each patient
C_M the cost of testing biomarker for each patient
C_f the cost of follow up in unit time
C_s the screening cost for each patient
T_f Expected total follow up time

$$\xi_{patient} = \frac{n_{AEBSD}}{n_{BSD}}, \xi_{cost} = \frac{C_{AEBSD}}{C_{BSD}}, \xi_{screening} = \frac{n_s}{n_{BSD}}$$

Numerical Example

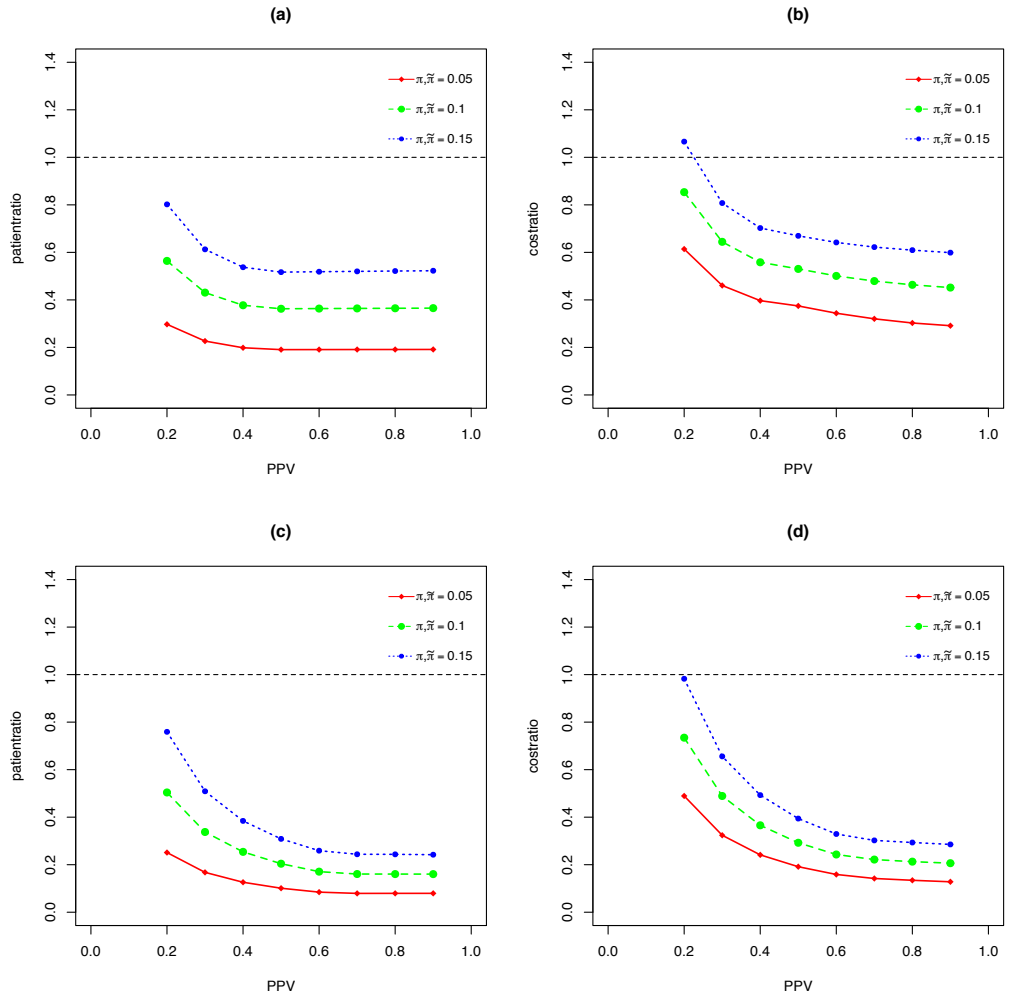
- treatment effect in $M+$ patients: $P_{w1} = 90\%$, $\alpha_1 = 0.025$
- interaction effect: $P_{w2} = 90\%$, $\alpha_2 = 0.025$
- overall treatment effect: $P_{w3} = 90\%$, $\alpha_3 = 0.025$
- $\tau = 1$; $\pi = \tilde{\pi}_e = 0.05, 0.1, 0.15$;
- $C_D = 5000$, $C_f = 200$, $C_M = 1000$, $C_s = 100$;
- λ_{ij} = hazard function for marker i (0,1); treatment j (C,E)

  $\lambda_{0C} = 0.8$, $\lambda_{1C} = 1.6$, $\lambda_{1E} = 1 \Rightarrow \Delta_1 = 0.625$

  $\lambda_{0E} = 0.7 \Rightarrow \Delta_0 = 0.875$ quantitative interaction;

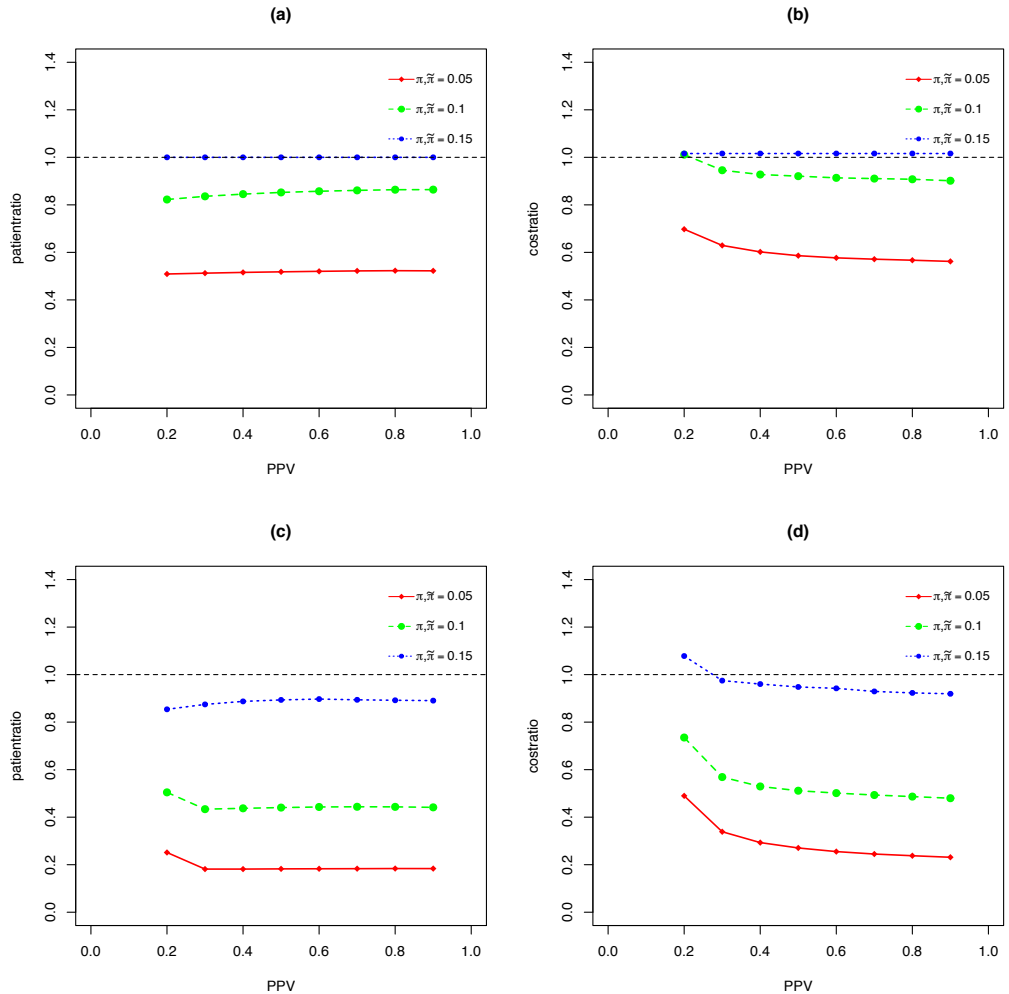
  $\lambda_{0E} = 1.1 \Rightarrow \Delta_0 = 1.375$ qualitative interaction.

Patient and Cost Ratios for Testing Treatment Effect in $M+$ Patients and the Interaction Effect



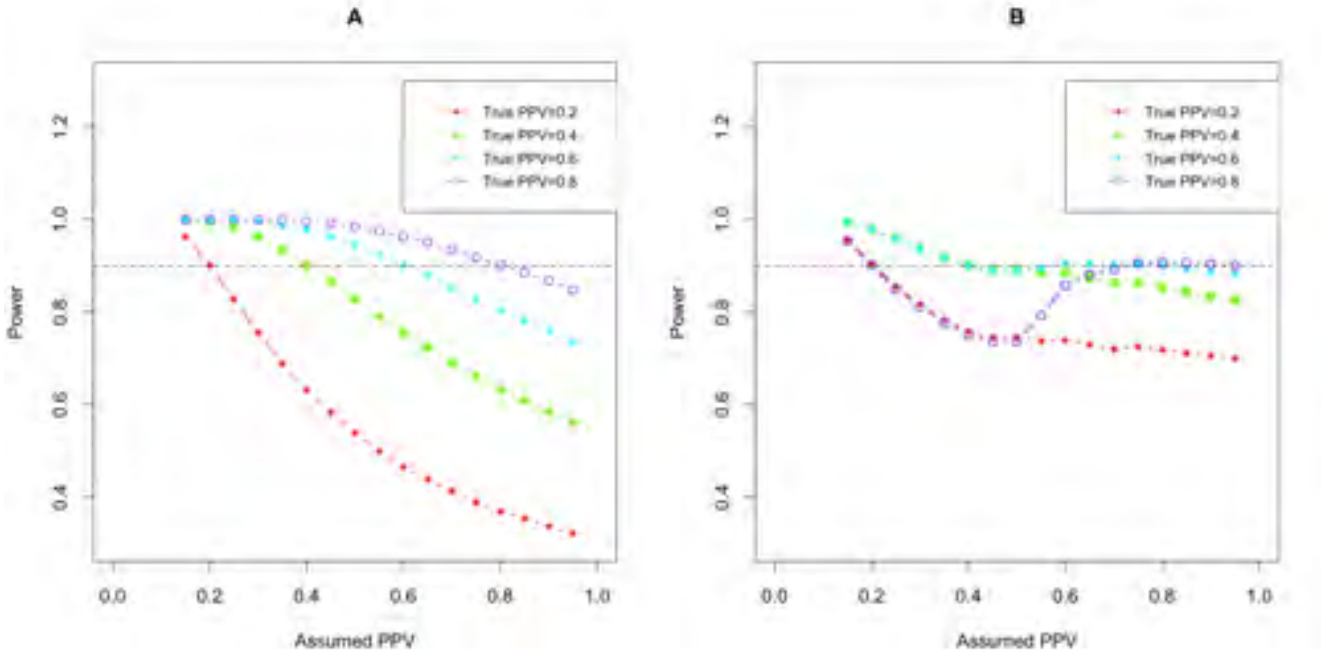
Under Scenario 4, for different *PPV* and different pairs of $\pi = \tilde{\pi}_e$, (a) Patient ratio when $\Delta_0 = 0.875$; (b) Cost ratio when $\Delta_0 = 0.875$; (c) Patient ratio when $\Delta_0 = 1.375$; (d) Cost ratio when $\Delta_0 = 1.375$.

Patient and Cost Ratios for Testing Treatment Effect in $M+$ Patients and Overall



Under Scenario 5, for different *PPV* and different pairs of $\pi = \tilde{\pi}_e$, (a) Patient ratio when $\Delta_0 = 0.875$; (b) Cost ratio when $\Delta_0 = 0.875$; (c) Patient ratio when $\Delta_0 = 1.375$; (d) Cost ratio when $\Delta_0 = 1.375$.

Power as a Function of the Specified *PPV* for Different True *PPV* Values



(a) Power of test of treatment effect in $M+$ patients for different true *PPV*s as a function of the specified *PPV* when $\pi = \tilde{\pi} = 0.15$; (b) Power of test of interaction effect for different true *PPV*s as a function of the specified *PPV* when $\pi = \tilde{\pi} = 0.15$

Adaptive Bayesian Estimate of *PPV*

Let $PPV \sim Beta(\alpha_0, \beta_0)$ so that the initial estimate of PPV is

$$PPV_0 = \frac{\alpha_0}{\alpha_0 + \beta_0}$$

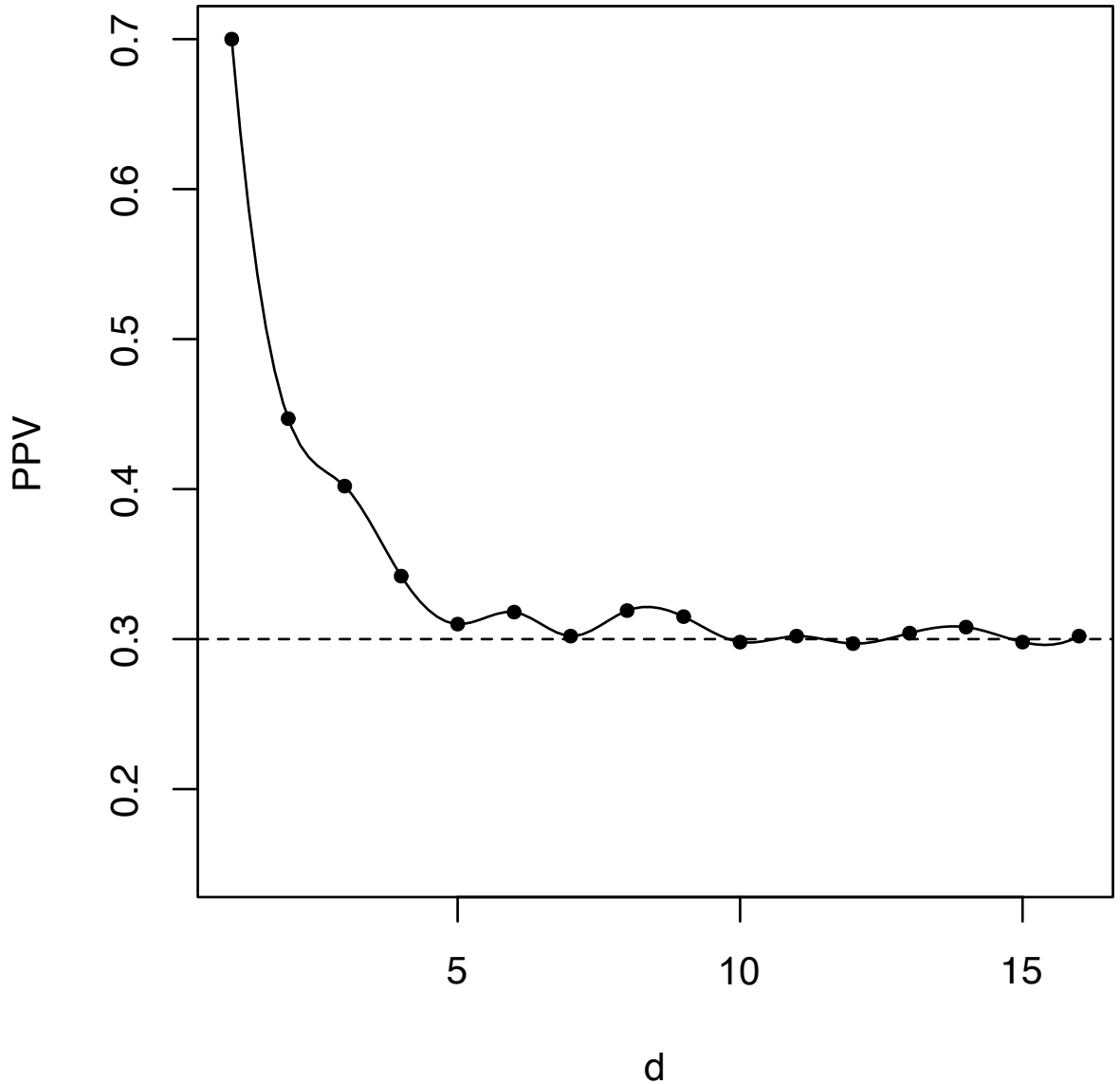
Update the estimate of PPV after each group of k $\tilde{M}+$ patients as

$$PPV_d = \frac{\alpha_0 + m_d}{\alpha_0 + \beta_0 + dk}$$

where $d = 1, 2, \ldots$ and
m_d = total number of the dk patients with both $\tilde{M}+$ and $M+$.

Recalculate $n, n_s, \tilde{\kappa}_0$ after each update.

Bayesian Estimate of *PPV*



AEBSD Case Study: Gefitinib or Carboplatin-Paclitaxel in NSCLC

Consider a hypothetical study of Gefitinib or Carboplatin-Paclitaxel in pulmonary adenocarcinoma in North America.

- M: EGFR mutation
- $\pi \approx 0.10$ in North America
- $\tilde{M}$: Predictive score considering adenocarcinomas, no history of smoking, females and Asian descent
- $\tilde{\pi} \approx 0.15$, $PPV \approx 0.60$
- $a = 10$ patients/month, $\tau = 2$ years
- Median PFS for (M, D) group (Mok et al. (2009)):

  9.8 mos in (1,1) 2.0 mos in (0,1)
  4.7 mos in (1,0) 5.7 mos in (0,0)

AEBSD Case Study: Gefitinib or Carboplatin-Paclitaxel in NSCLC

- H_0:$log\,\Delta_1 = 0$ vs.H_a:$log\,\Delta_1 \neq 0$ and
  H_0:$log\,\Delta_I = 0$ vs.H_a:$log\,\Delta_I \neq 0$

- Results:
  - AEBSD: $n = 157$, $n_+ = 94$, $n_- = 63$, $n_s = 1047$, $\tilde{\pi}_e = 0.60$
  - BSD: $n = 926$, $n_+ = 94$, $n_- = 832$, $\pi = 0.10$
  - Comparison: $\xi_{patient} = 0.17$, $\xi_{cost} = 0.42$, $\xi_{screening} = 1.13$

Conclusions

- Enrichment strategies can be a powerful tool to improve trial efficiency while retaining the full features of a BSD design.
- Compared to a BSD design, enrichment designs:
  - reduce the number of randomized patients
  - reduce the total trial cost
- The AEBSD design allows immediate randomization without waiting for the result of the true biomarker test
- However, enrichment designs may increase the total trial time

Additional Reading I

B. Freidlin and E. L. Korn.

Biomarker enrichment strategies: matching trial design to biomarker credentials.
Nature Reviews Clinical Oncology, 11(2):81–90, 2014.

B. Freidlin, L. M. McShane, and E. L. Korn.

Randomized clinical trials with biomarkers: design issues.
Journal of the National Cancer Institute, 102:152–160, 2010.

S. L. George and M. Desu.

Planning the size and duration of a clinical trial studying the time to some critical event.
Journal of chronic diseases, 27(1-2):15–24, 1974.

S. L. George and X. Wang.

Targeted clinical trials.
In D. Harrington, editor, *Designs for Clinical Trials: Perspectives on Current Issues*, pages 157–177. Springer, 2012.

O. Gevaert, S. Echegaray, A. Khuong, C. D. Hoang, J. B. Shrager, K. C. Jensen, G. J. Berry, H. H. Guo, C. Lau, S. K. Plevritis, D. L. Rubin, S. Napel, and A. N. Leung.

Predictive radiogenomics modeling of egfr mutation status in lung cancer.
Scientific Reports, 7:41674, 2017.

E. L. Korn and B. Freidlin.

Biomarker-based clinical trials.
In S. L. George, X. Wang, and H. Pang, editors, *Cancer Clinical Trials: Current and Controversial Issues in Design and Analysis*, pages 333–364. Chapman and Hall/CRC, 2016.

Additional Reading II

T. S. Mok, Y.-L. Wu, S. Thongprasert, C.-H. Yang, D.-T. Chu, N. Saijo, P. Sunpaweravong, B. Han, B. Margono, Y. Ichinose, et al.
Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma.
New England Journal of Medicine, 361(10):947–957, 2009.

B. Peterson and S. L. George.
Sample size requirements and length of study for testing interaction in a 2 $\times$ k factorial design when time-to-failure is the outcome.
Controlled clinical trials, 14(6):511–522, 1993.

C. N. Prabhakar.
Epidermal growth factor receptor in non-small cell lung cancer.
Translational lung cancer research, 4(2):110–118, Apr 2015.

D. A. Schoenfeld.
Sample-size formula for the proportional-hazards regression model.
Biometrics, 39:499–503, 1983.

R. Simon and A. Maitournam.
Evaluating the efficiency of targeted designs for randomized clinical trials.
Clinical Cancer Research, 10(20):6759–6763, 2004.