Statistical Methods in Surrogate Marker Research for Clinical Trials

Jeremy M G Taylor
Yun Li
Roderick J A Little

Department of Biostatistics
University of Michigan

Outline

- Introduction
- Predicting treatment effects using surrogate markers in clinical trials
  - Single Trial
  - Multiple Trials
- A shrinkage approach for estimating a treatment effect using surrogate marker data in clinical trials

Surrogate Markers in a Randomized Trial

- Setting: Two arm randomized clinical trial
  - Z (binary) treatment group
  - S surrogate marker
  - T true endpoint
  - S is measured after Z but before T
- Example 1
  - Z is standard or new therapy for HIV disease
  - S is CD4 counts at 12 weeks after treatment
  - T is death
- Example 2
  - Z is radiation or radiation plus hormone therapy for prostate cancer
  - S is prostate specific antigen at 1 year
  - T is recurrence of prostate cancer

Surrogate Markers in a Randomized Trial

- Surrogate markers (S):
  - Can be measured earlier than the true endpoint (T)
  - Are intermediate physical or laboratory indicators in a disease progression process
  - Surrogate endpoint: May serve as a substitute for the true endpoint
  - Auxiliary variable: May help predict the treatment (Z) effect on T
  - Make trial cheaper and faster

Research Questions

- Can S be a replacement for T?
- How much is the treatment effect on T captured by the treatment effect on S?
- Can we use S to help predict the treatment effect on T when T is completely missing or partially missing?
- Can we use S to improve efficiency in the estimated treatment effect of Z on T?

Landmark Papers

- Prentice's (1989) criteria for perfect surrogacy ($Z \rightarrow S \rightarrow T$):
  - $f(T|Z) \neq f(T)$
  - $f(S|Z) \neq f(S)$
  - $f(T|S) \neq f(T)$
  - $f(T|S,Z) = f(T|S)$
- Then test of null hypothesis, $E(S|Z=0) = E(S|Z=1)$ is a valid test of $E(T|Z=0) = E(T|Z=1)$

Landmark Papers cont.

- Limitations
  - Prentice's criteria are too restrictive
  - Perfect surrogacy unrealistic
  - Don't expect $f(T|S,Z)$ to be exactly equal to $f(T|S)$
  - Conditioning on a post treatment variable (S) is non-causal
  - Not clear how to use S to predict the effect of Z on T

Landmark Papers cont.

- Freedman's (1992) proportion of treatment effect explained (F) for less than perfect surrogacy.
  - Consider: $Z \rightarrow S \rightarrow T$ and $Z \rightarrow T$.
  - e.g. $E(T|Z) = \alpha_0 + \alpha_1 Z$
  - e.g. $E(T|Z,S) = \beta_0 + \beta_1 Z + \beta_2 S$
  - $F = \frac{\textit{Treatment effect on } T \textit{ explained by } S}{\textit{Treatment effect on } T}$.
  - $F = \frac{\alpha_1 - \beta_1}{\alpha_1}$.
- Improvement in F (Wang et al 2003)

Landmark Papers cont.

- Limitations
  - F can be out of range of (0,1)
  - Estimates of F can have wide confidence intervals
  - F does not tell you how to use S to predict the effect of Z on T
  - Conditioning on a post treatment variable (S) is non-causal

Landmark Papers cont

- Multiple Trials (Daniels and Hughes 1997, Buyse et al 2000)
  - Have data on Z, S and T in $n-1$ similar trials
  - Goal estimate treatment effect of Z on T in n^{th} trial
  - Have data on Z, S and partially on T in new trial.
  - Hierarchical models useful
  - Within-trial and between-trial measures of association have been suggested
  - Trials may not be similar

Predicting treatment effect using surrogate markers

- Our objective:
  - Estimate the treatment effect on T in a clinical trial
  - Examine the extent of efficiency gain by using S
- A missing data problem
  - T partially observed

The Setup

The setup: S and T are continuous either in a single trial setting or a multiple-trial setting

x = observed
? = unobserved

Single Trial		
Z	**S**	**T**
x	x	x
x	x	x
x	x	x
x	x	?
x	x	?
x	x	?
x	x	?

x = observed
? = unobserved

1st Trial			- - -	*n*th Trial		
Z	**S**	**T**	- - -	**Z**	**S**	**T**
x	x	x	- - -	x	x	x
x	x	x	- - -	x	x	x
x	x	x	- - -	x	x	x
x	x	x	- - -	x	x	?
x	x	x	- - -	x	x	?
x	x	x	- - -	x	x	?
x	x	x	- - -	x	x	?

Information Recovery from S in a Single Trial

For individual j, S_j and T_j follow:

$$\begin{pmatrix} S_j \\ T_j \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{0Sn} + \delta_{Sn} Z_j \\ \mu_{0Tn} + \delta_{Tn} Z_j \end{pmatrix}, \begin{pmatrix} \sigma_{ss} & \sigma_{st} \\ & \sigma_{tt} \end{pmatrix}\right).$$

We define the individual-level correlation as

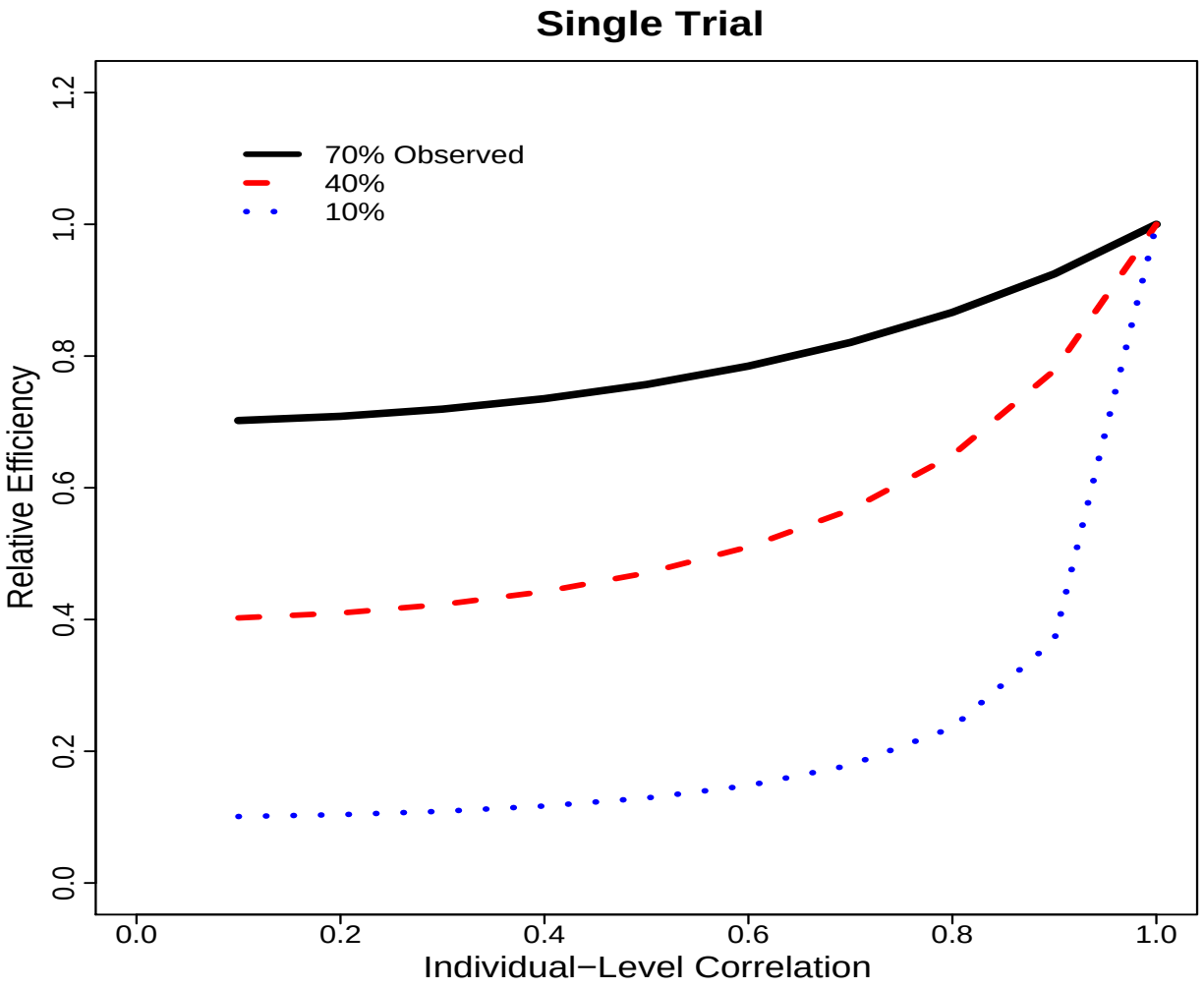
$$R^2_{indiv} = \frac{\sigma^2_{st}}{\sigma_{ss}\sigma_{tt}}.$$

Comparing the estimator using S with that without the use of S and T fully observed, we have the relative efficiency (RE) as

$$RE = \frac{(1 - (\%missing)))}{(1 - R^2_{indiv} \times (\%missing))}.$$

Note: assume the same correlation, patient size and observed proportion of T are the same between two groups.

Information Recovery from S in a Single Trial



Multiple Trial Setting

Suppose we have n randomized trials, $i = 1, \ldots, n$, where the nth trial is new. m patients per trial. For individual j in the ith trial,

$$S_{ij} = \alpha_0 + \alpha_1 Z_{ij} + a_{0i} + a_{1i} Z_{ij} + \varepsilon_{Sij}$$
$$T_{ij} = \gamma_0 + \gamma_1 Z_{ij} + r_{0i} + r_{1i} Z_{ij} + \varepsilon_{Tij}$$

where

$$\begin{pmatrix} \varepsilon_{Sij} \\ \varepsilon_{Tij} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{ss} & \sigma_{st} \\ & \sigma_{tt} \end{pmatrix} \right),$$

and

$$\begin{pmatrix} a_{0i} \\ r_{0i} \\ a_{1i} \\ r_{1i} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, D = \begin{pmatrix} d_{ss} & d_{st} & d_{sa} & d_{sr} \\ & d_{tt} & d_{ta} & d_{tr} \\ & & d_{aa} & d_{ar} \\ & & & d_{rr} \end{pmatrix} \right).$$

Correlation (Buyse *et al*, 2000)

- Individual-Level Correlation:

$$R^2_{indiv} = \frac{\sigma^2_{st}}{\sigma_{ss}\sigma_{tt}}.$$

- Trial-Level Correlation:

$$R^2_{trial} = \frac{\begin{pmatrix} d_{sr} & d_{ar} \end{pmatrix} \begin{pmatrix} d_{ss} & d_{sa} \\ d_{sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{sr} \\ d_{ar} \end{pmatrix}}{d_{rr}}.$$

- Goal: examine how R^2_{indiv} and R^2_{trial} impact the extent of efficiency gain from S in the nth trial.
- Buyse et al argued that R^2_{trial} is more important than R^2_{indiv}

Prediction and Correlation

Let δ_{Tn} be the treatment effect on T in the nth trial.
Let $\hat{\delta}_{Tn}$ be the empirical Bayes estimate

- $$S_{ij} = \alpha_0 + \alpha_1 Z_{ij} + a_{0i} + a_{1i} Z_{ij} + \varepsilon_{Sij}$$
  $$T_{ij} = \gamma_0 + \gamma_1 Z_{ij} + r_{0i} + r_{1i} Z_{ij} + \varepsilon_{Tij}$$
- $\delta_{Tn} = \gamma_1 + r_{1n}$
- Use mixed model algebra (Henderson) to get $var(\hat{\delta}_{Tn})$

Prediction and Correlation

- when T is completely missing in the new trial:

$$var(\hat{\delta}_{Tn}) = \begin{pmatrix} 0 & 1 \end{pmatrix} \Psi_d \begin{pmatrix} 0 & 1 \end{pmatrix}^T.$$

- when T is partially observed in the new trial:

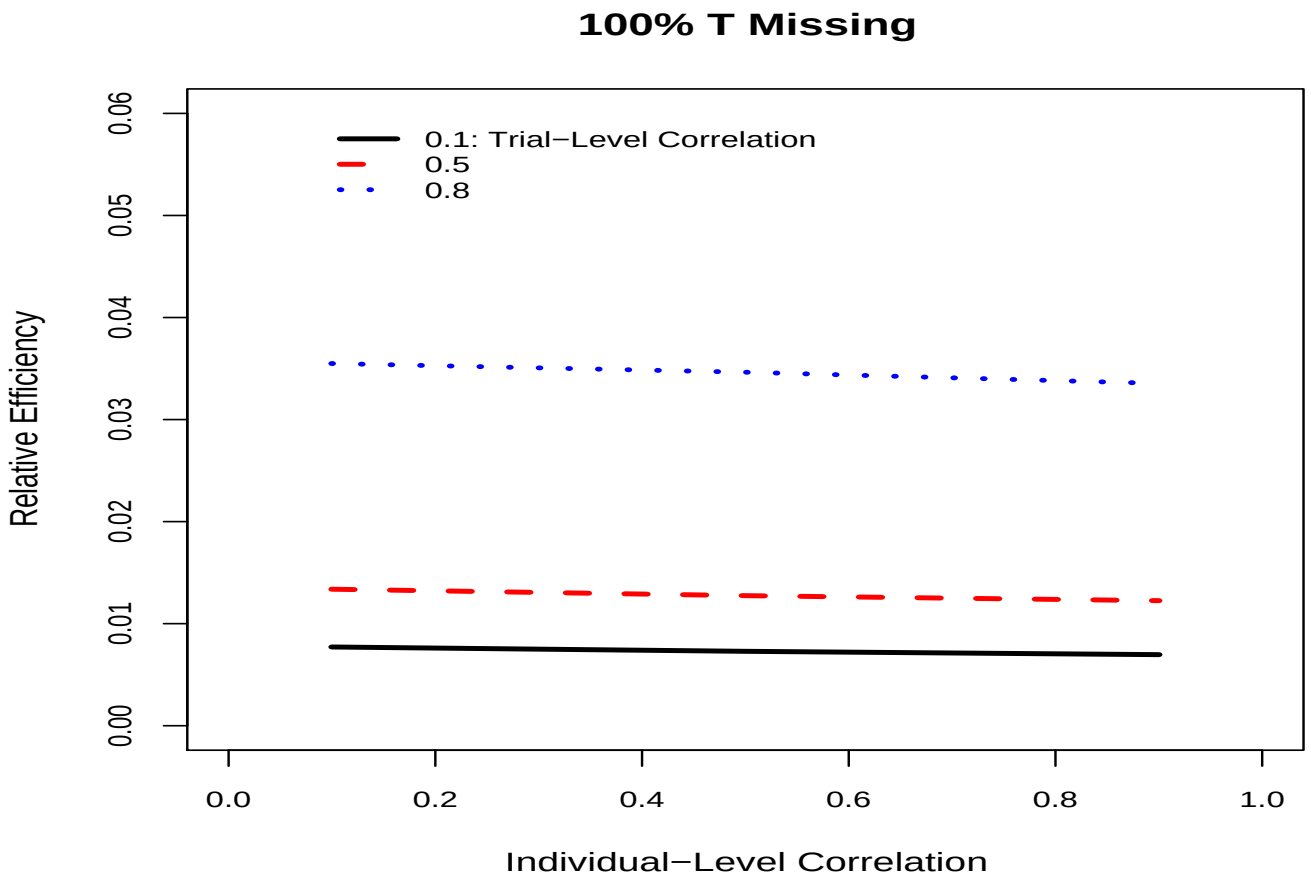
$$var(\hat{\delta}_{Tn}) = \begin{pmatrix} 0 & 1 \end{pmatrix} \left(\Psi_d^{-1} + \Phi_e^{-1} \right)^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

- Ψ_d is a function only of the between-trial covariances (D)
- Φ_e is a function only of the within-trial covariances (Σ).

Note: assume β, D and Σ are known quantities.

T is Completely Missing in the New Trial

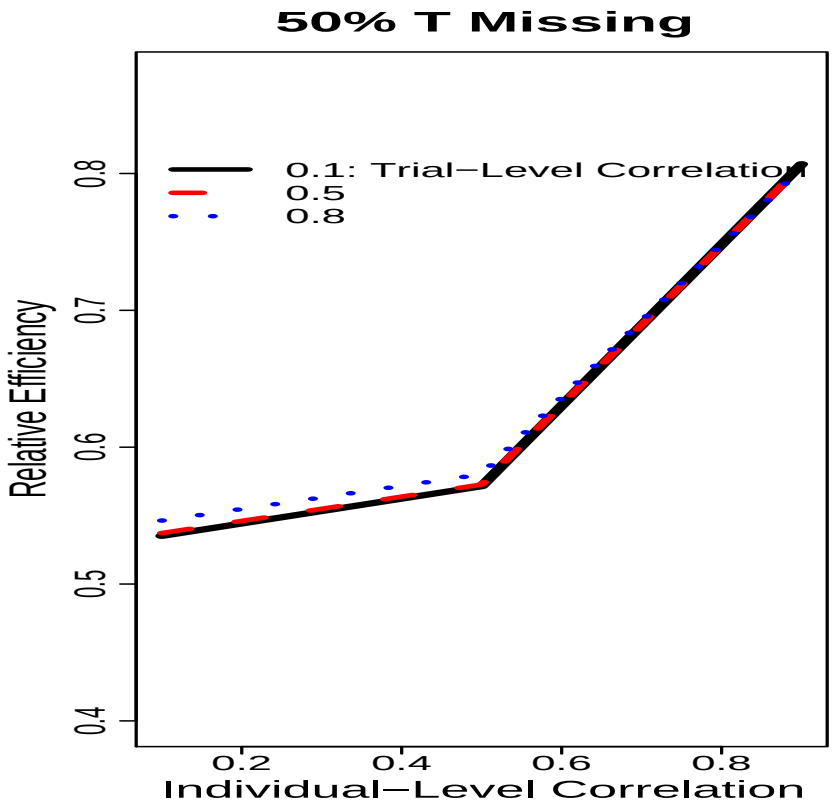
Relative efficiency: comparing estimates using S when T is completely missing in the new trial to that when T is completely observed.

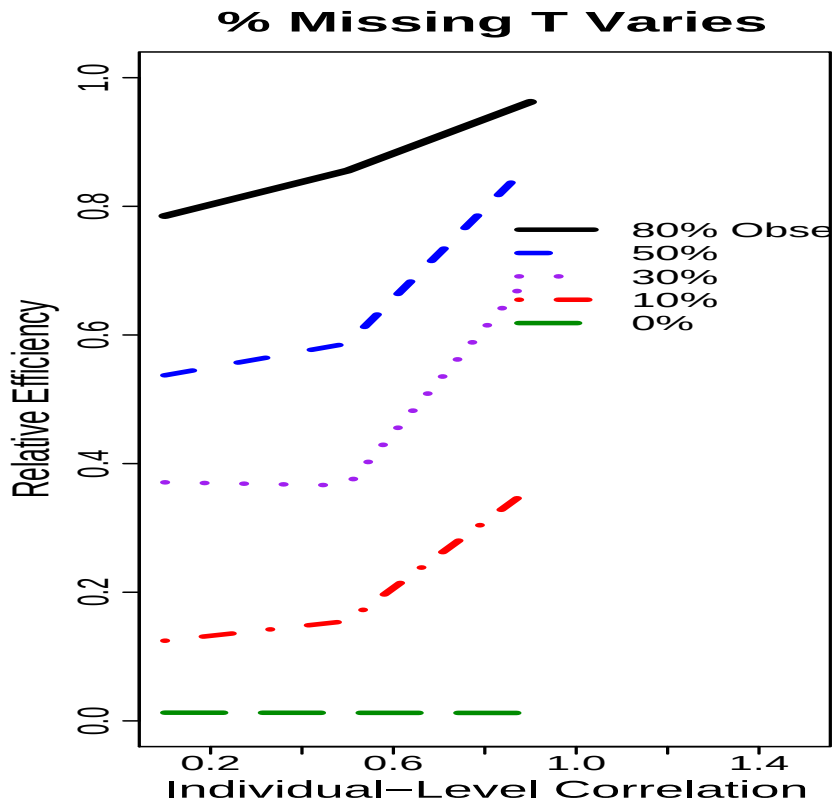


$n = 40$, $m = 100$, $\beta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)^T = (1,2,1,1)^T$, $d_{ss} = 0.5$, $d_{tt} = 0.2$, $d_{aa} = 3.5$, $d_{rr} = 1.6$, $\sigma_{ss} = 1$ and $\sigma_{tt} = 0.3$

T is Partially Missing in the New Trial

Relative efficiency: comparing estimates using S when T is partially observed in the new trial to that when T is completely observed.





$n = 40$, $m = 100$, $\beta^T = (1,2,1,1)$, $d_{ss} = 0.5$, $d_{tt} = 0.2$, $d_{aa} = 3.5$, $d_{rr} = 1.6$, $\sigma_{ss} = 1$ and $\sigma_{tt} = 0.3$.

Glaucoma Randomized Trial

- Surgery versus Medicine
- Multicenter study
- Treat data from each center as a different trial
- n=600
  - Pseudo-data
  - Looks like original data
  - Expanded to n=6000
- S = intra-ocular pressure at 12 months
- T = intra-ocular pressure at 96 months
- $R^2_{trial} = 0.25, R^2_{indiv} = 0.15$

Center	n	Medicine	Surgery	**Correlation**	
		Means (S,T)	Means (S,T)	Medic	Surg
1	670	(17.6, 16.5)	(13.8, 14.6)	0.37	0.61
2	352	(17.2, 16.4)	(14.6, 13.0)	-0.46	0.47
3	770	(19.3, 17.6)	(15.8, 16.2)	0.59	0.55
4	528	(17.2, 15.5)	(10.9, 12.9)	0.18	0.54
5	1078	(18.5, 18.7)	(15.0, 15.3)	0.44	0.41
6	736	(18.6, 18.9)	(15.1, 17.1)	-0.16	-0.01
7	572	(18.4, 15.3)	(14.6, 14.5)	0.18	0.40
8	1056	(18.6, 16.2)	(13.6, 13.7)	0.31	0.95
9	638	(17.6, 16.8)	(14.2, 14.6)	0.04	0.76

Table: Description of Pseudodata in Glaucoma study

p	Estimate	Standard Error	p-value
center = 8			
SIMPLE†	-2.45	0.29	$< .0001$
100% missing	-1.58	0.79	0.063
90% missing	-1.50	0.47	0.006
80% missing	-2.37	0.39	$< .0001$
50% missing	-2.61	0.29	$< .0001$
20% missing	-2.19	0.23	$< .0001$
No missing	-2.33	0.22	$< .0001$

Table: Estimate treatment effect on IOP. †: Based on complete data before any deletion.

Summary

- The use of S as an auxiliary variable can be beneficial.
- In a single trial, the efficiency gain using S is only mildly encouraging because extremely high correlation is necessary.
- In a multiple trial setting:
  - when T is completely missing, efficiency gain is associated with R^2_{trial} but negligible.
  - when T is partially missing, efficiency gain is mostly associated with R^2_{indiv} and can be very big.
- Recommendation: collecting S that has high individual-level correlation with T and some amount of T is essential.

A Shrinkage Approach for Estimating a Treatment Effect Using Surrogate Marker Data in Clinical Trials

- Explore the extent of efficiency gain with respect to the models that describe the relationship among T, S and Z
- Propose a shrinkage approach

x = observed ? = unobserved

Single Trial		
Z	**S**	**T**
x	x	x
x	x	x
x	x	x
x	x	?
x	x	?
x	x	?

Estimating Treatment Effect Using S

- Potential for gain in efficiency from using S is due to
  - The use of more data (S) gives extra information
  - Assumptions in a model

Estimating Treatment Effect Using S

- S and T are continuous.
- The joint distribution of S and T given Z can be expressed as:

$$\begin{aligned} T_i &= \beta_0 + \beta_1 S_i + \beta_2 Z_i + \beta_3 S_i Z_i + \varepsilon_{ti} \\ S_i &= \alpha_0 + \alpha_1 Z_i + \varepsilon_{si} \end{aligned}$$

- Treatment effect $Q = E(T|Z=1) - E(T|Z=0)$
- $Q = \beta_1 \alpha_1 + \beta_2 + \beta_3 \alpha_0 + \beta_3 \alpha_1$

Estimating Treatment Effect Using S

$$\begin{aligned} T_i &= \beta_0 + \beta_1 S_i + \beta_2 Z_i + \beta_3 S_i Z_i + \varepsilon_{ti} \\ S_i &= \alpha_0 + \alpha_1 Z_i + \varepsilon_{si} \end{aligned}$$

- The treatment effect on T (Q) and model assumptions:
  - Perfect Surrogacy Model (PES): $\beta_2 = \beta_3 = 0$

$$Q = \beta_1 \alpha_1 .$$

  - Additive Partial Surrogacy Model (APAS): $\beta_3 = 0$

$$Q = \beta_2 + \beta_1 \alpha_0 .$$

  - Interactive Partial Surrogacy Model (IPAS): $\beta_2 \neq 0$; $\beta_3 \neq 0$

$$Q = \beta_1 \alpha_1 + \beta_2 + \beta_3 \alpha_0 + \beta_3 \alpha_1 .$$

Numerical Results

- Z and S fully observed, T partially observed
- Extent of information gain depends on the true model and the assumptions in the fitted model.
- Substantial efficiency gain if perfect surrogacy (PES) is assumed correctly.
- Estimates are biased if the perfect surrogacy (PES) is assumed incorrectly.
- Limited gain in efficiency, but no bias, by assuming full model (IPAS)

A Naive Proposed Method - Model Selection

- Model Selection
  - First test if $\beta_2 = 0$ and $\beta_3 = 0$
  - If accept, then use PES model for estimation
  - If reject, then use IPAS model for estimation

Prentice Criteria

- Prentice criteria $f(T|S,Z) = f(T|S)$
- For many plausible S, $f(T|S,Z)$ will be close to $f(T|S)$
-

$$T_i = \beta_0 + \beta_1 S_i + \beta_2 Z_i + \beta_3 S_i Z_i + \varepsilon_{ti} \tag{1}$$

- Expect β_2 and β_3 to be close to zero, but not exactly zero.

Proposed Method: Generalized Ridge Regression

The Model:

$$T_i = \beta_0 + \beta_1 S_i + \beta_2 Z_i + \varepsilon_{ti}$$

- Apply shrinkage to β_2.
  - Fully Bayesian version (Ridge-FB): $\beta_2 \sim N(0, \sigma^2_{b_2})$.
  - Empirical Bayesian version (Ridge-EB)

Proposed Method: Generalized Ridge Regression

The Model:

$$T_i = \beta_0 + \beta_1 S_i + \beta_2 Z_i + \varepsilon_{ti}$$

- Fully Bayesian version: $\beta_2 \sim N(0, \sigma^2_{b_2})$ and diffuse informative priors for all other parameters.
- Empirical Bayesian version: $\hat{\sigma}^2_{b_2} = \hat{\beta}^2_2$.

$$\mathrm{E}(\hat{\beta}|X_t, T) = (X_t^T X_t + K)^{-1} X_t^T T,$$
$$\mathrm{V}(\hat{\beta}|X_t, T) = (X_t^T X_t + K)^{-1} \sigma_t^2.$$

where $\beta^T = (\beta_0, \beta_1, \beta_2)$, $X_t = (1, S, Z)$, $K = \mathrm{diag}(0, 0, k_2)$ and $k_2 = \sigma_t^2 / \sigma^2_{b_2}$.

Simulations: Compare Ridge, IPAS, APAS and PES

1) IPAS: $\beta_2 \neq 0$ and $\beta_3 \neq 0$; 2) APAS: $\beta_3 = 0$; 3) PES: $\beta_2 = \beta_3 = 0$.

	Fitted Models				
Treatment Effect (Q)	PES	APAS	IPAS	Ridge FB	Ridge EB
$\beta_2 = 0, Q = 2$					
Estimate	2.029	2.042	2.050	2.037	2.040
MSE	0.133	0.207	0.216	0.156	0.172
$\beta_2 = 2, Q = 4$					
Estimate	3.402	4.042	4.050	3.885	3.963
MSE	0.527	0.207	0.216	0.236	0.217

Table: $n = 120$, $\beta_0 = 0.5$, $\beta_1 = 1$, $\alpha_0 = 1$, $\alpha_1 = 2$, $\sigma_t^2 = 0.1$, $\sigma_s^2 = 0.5$, $\%miss = 80\%$

Numerical Study of Extent of Information Recovery

- True model: perfect surrogacy model ($\beta_2 = \beta_3 = 0$).
- Fitted model:
  - 1) full model, (IPAS);
  - 2) no interaction model, ($\beta_3 = 0$), (APAS);
  - 3) perfect surrogacy model, ($\beta_2 = \beta_3 = 0$), (PES),
  - 4) simple estimator based on complete cases, (CC).

Numerical Study of Extent of Information Recovery

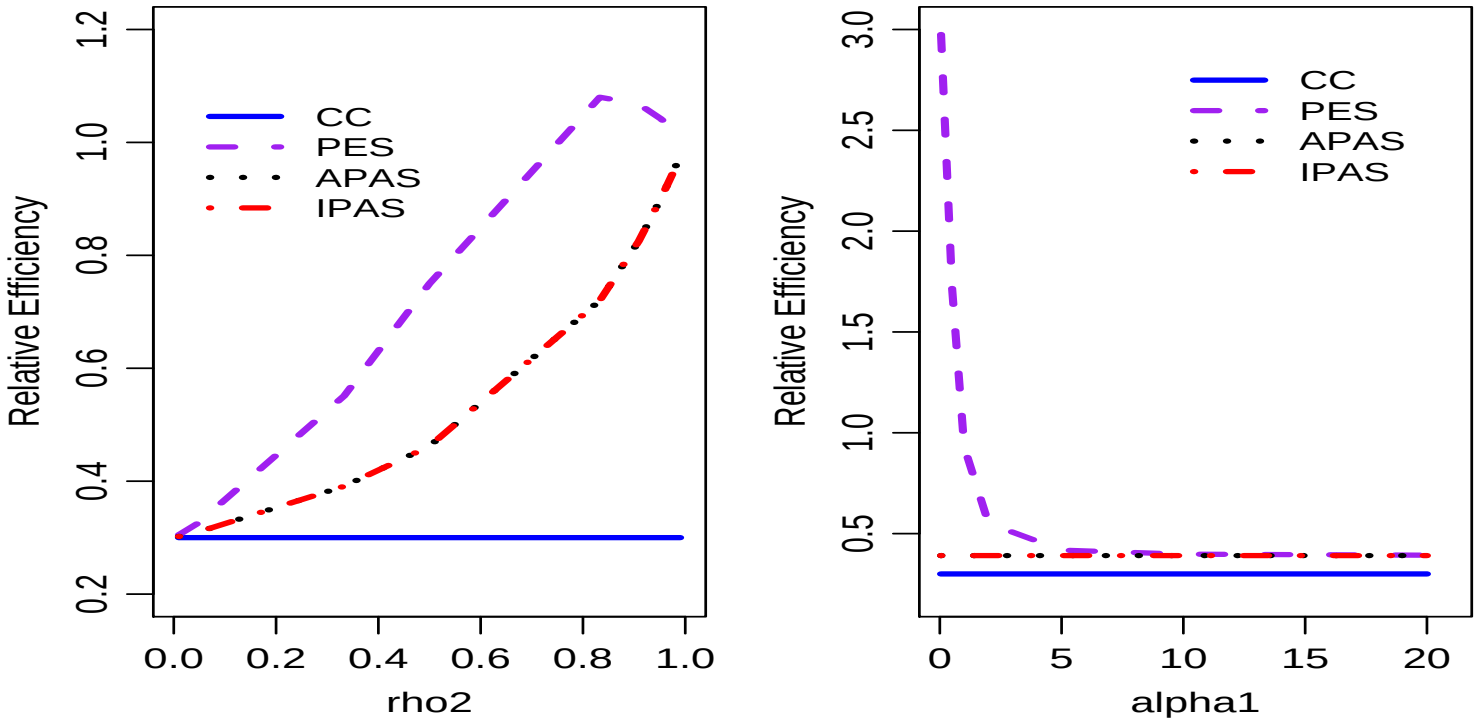
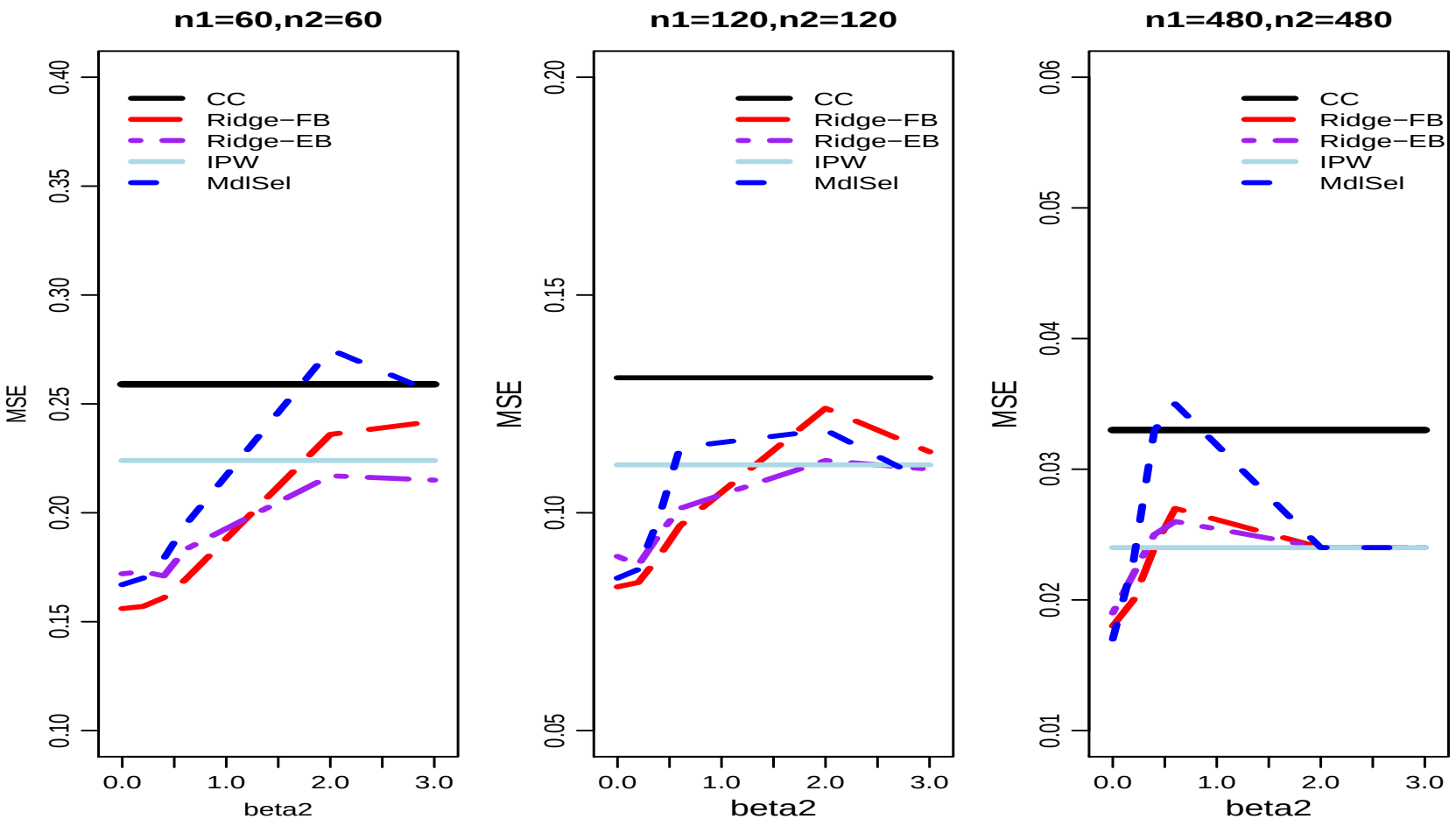
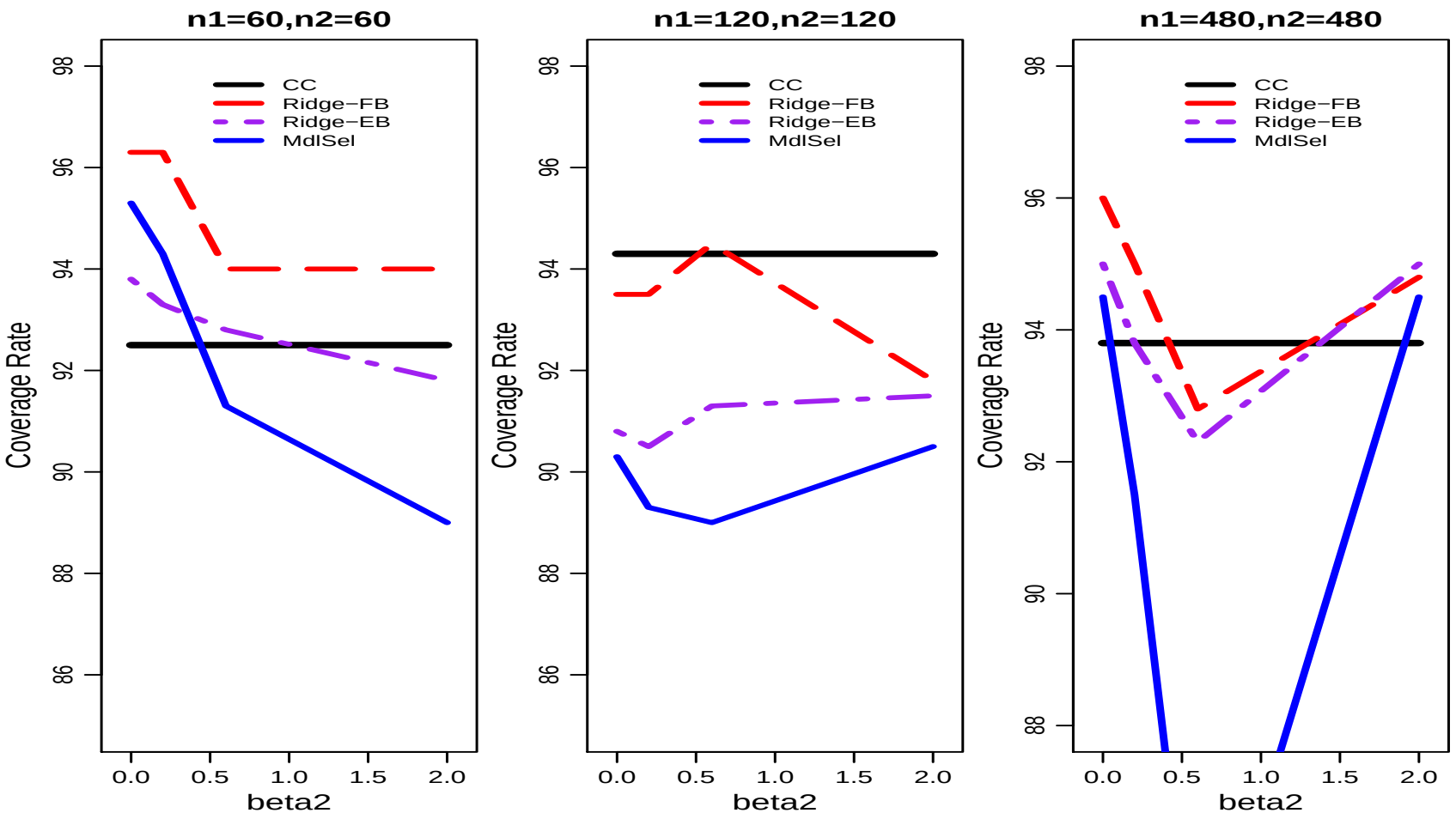


Figure: Asymptotic Relative Efficiency Compared with that Obtained from Original Data.

Simulations: Compare Ridge, IPW and Model Selection Regarding MSE



Simulations: Compare Ridge and Model Selection Regarding Coverage Rates



Glaucoma Randomized Trial

- Surgery versus Medicine
- n=600
- S = intra-ocular pressure at 12 months
- T = intra-ocular pressure at 102 months

Glaucoma Randomized Trial

	Medicine	Surgery
Observed at 12th Missing at 102nd Month		
Number of Patients	206	207
Mean IOP at 12th Month	18.2	14.3
Observed at 12th Observed at 102nd Month		
Number of Patients	86	74
Mean IOP at 12th Month	17.9	14.1
Mean IOP at 102nd Month	17.5	15.1

Table: Summary Statistics from Glaucoma Trial.

Glaucoma Randomized Trial

Estimation Method	Estimate	95% CI	CI Width
CC	-2.39	$(-3.84, -0.94)$	2.91
IPAS	-2.42	$(-3.79, -1.05)$	2.75
APAS	-2.40	$(-3.77, -1.03)$	2.73
PES	-1.83	$(-2.49, -1.18)$	1.32
Model-Selection	-1.83	$(-2.49, -1.18)$	1.32
Ridge-EB	-2.09	$(-3.14, -1.05)$	2.09
Ridge-FB	-2.02	$(-3.03, -1.01)$	2.03
IPW	-2.39	$(-3.73, -1.05)$	2.68

Table: Quantity of Interest: Difference in the IOP Reduction at the 102nd Month between Medicine and Surgery Treatments.

Summary

- Ridge method can strike a balance between efficiency gain and bias reduction depending on the evidence from the data regarding the validity of the surrogacy assumption.
- Ridge method has better MSE and coverage rate properties than the competing methods, particularly in small samples and/or when S is close to being a perfect surrogate.

Selected References

1. Buyse M., et al (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*. **1**, 49-67..
2. Daniels M.J. and Hughes M.D. (1997). Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine*, **16**, 1515-1527.
3. Freedman L.S., et al (1992). Statistical validation of intermediate endpoints for chronic disease. *Statistics in Medicine*, **11**, 167-178.
4. Henderson C.R. (1984). Applications of linear models in animal breeding. *University of Guelph*
5. Musch D.C., et al (1999): The Collaborative Initial Glaucoma Treatment Study (CIGTS): Study design, methods, and baseline characteristics of enrolled patients. *Ophthalmology*. **106**: 653-62.
6. Prentice R.L. (1989). Surrogate endpoints in clinical trials, definition and operational criteria. *Statistics in Medicine*, **8**, 431-440.
7. Wang Y. and Taylor J.M.G. (2003). A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics*, **58**, 803-812.