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
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?
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 $E(T|Z = 0) = E(T|Z = 1)$
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 \( T \)
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. $E(T|Z, S) = \beta_0 + \beta_1 Z + \beta_2 S$
- $F = \frac{\text{Treatment effect on } T \text{ explained by } S}{\text{Treatment effect on } T}$.
- $F = \frac{\alpha_1 - \beta_1}{\alpha_1}$.

Improvement in $F$ (Wang et al 2003)
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
**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
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: $S$ and $T$ are continuous either in a single trial setting or a multiple-trial setting.

- **Single Trial**
  - $x = \text{observed}$
  - $? = \text{unobserved}$

- **1st Trial**
  - $x = \text{observed}$
  - $? = \text{unobserved}$

<table>
<thead>
<tr>
<th>Trial</th>
<th>$Z$</th>
<th>$S$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial</th>
<th>$Z$</th>
<th>$S$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>nth</td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial</th>
<th>$Z$</th>
<th>$S$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>nth</td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td></td>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
</tbody>
</table>
Information Recovery from S in a Single Trial

For individual \( j \), \( S_j \) and \( T_j \) follow:

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

We define the individual-level correlation as

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

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

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

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

Single Trial

![Graph showing relative efficiency across different levels of individual-level correlation.]

- 70% Observed
- 40%
- 10%
Multiple Trial Setting

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

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

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:
  \[ R^{2}_{\text{indiv}} = \frac{\sigma_{st}^2}{\sigma_{ss}\sigma_{tt}}. \]

- Trial-Level Correlation:
  \[ R^{2}_{\text{trial}} = \frac{\left( \begin{array}{cc} d_{sr} & d_{ar} \end{array} \right) \left( \begin{array}{cc} d_{ss} & d_{sa} \\ d_{sa} & d_{aa} \end{array} \right)^{-1} \left( \begin{array}{c} d_{sr} \\ d_{ar} \end{array} \right)}{d_{rr}}. \]

- Goal: examine how \( R^{2}_{\text{indiv}} \) and \( R^{2}_{\text{trial}} \) impact the extent of efficiency gain from S in the \( n \)th trial.

- Buyse et al argued that \( R^{2}_{\text{trial}} \) is more important than \( R^{2}_{\text{indiv}} \).
Let $\delta_{Tn}$ be the treatment effect on $T$ in the $n$th 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 $\text{var}(\hat{\delta}_{Tn})$
Prediction and Correlation

- When \( T \) is completely missing in the new trial:
  \[
  \text{var}(\hat{\delta}_{Tn}) = \begin{pmatrix} 0 & 1 \end{pmatrix} \Psi_d \begin{pmatrix} 0 & 1 \end{pmatrix}^T.
  \]

- When \( T \) is partially observed in the new trial:
  \[
  \text{var}(\hat{\delta}_{Tn}) = \begin{pmatrix} 0 & 1 \end{pmatrix} \left( \Psi_d^{-1} + \Phi_e^{-1} \right)^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix},
  \]
  \( \Psi_d \) is a function only of the between-trial covariances (\( D \))
  \( \Phi_e \) is a function only of the within-trial covariances (\( \Sigma \)).

Note: Assume \( \beta, D \) and \( \Sigma \) 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.

\[ n = 40, \ m = 100, \ \beta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)^T = (1, 2, 1, 1)^T, \ d_{ss} = 0.5, \ d_{tt} = 0.2, \ d_{aa} = 3.5, \ d_{rr} = 1.6, \ \sigma_{ss} = 1 \text{ and } \sigma_{tt} = 0.3 \]
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
- $R^2_{trial} = 0.25, R^2_{indiv} = 0.15$
<table>
<thead>
<tr>
<th>Center</th>
<th>n</th>
<th>Medicine Means (S, T)</th>
<th>Surgery Means (S, T)</th>
<th>Correlation Medic</th>
<th>Correlation Surg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>670</td>
<td>(17.6, 16.5)</td>
<td>(13.8, 14.6)</td>
<td>0.37</td>
<td>0.61</td>
</tr>
<tr>
<td>2</td>
<td>352</td>
<td>(17.2, 16.4)</td>
<td>(14.6, 13.0)</td>
<td>-0.46</td>
<td>0.47</td>
</tr>
<tr>
<td>3</td>
<td>770</td>
<td>(19.3, 17.6)</td>
<td>(15.8, 16.2)</td>
<td>0.59</td>
<td>0.55</td>
</tr>
<tr>
<td>4</td>
<td>528</td>
<td>(17.2, 15.5)</td>
<td>(10.9, 12.9)</td>
<td>0.18</td>
<td>0.54</td>
</tr>
<tr>
<td>5</td>
<td>1078</td>
<td>(18.5, 18.7)</td>
<td>(15.0, 15.3)</td>
<td>0.44</td>
<td>0.41</td>
</tr>
<tr>
<td>6</td>
<td>736</td>
<td>(18.6, 18.9)</td>
<td>(15.1, 17.1)</td>
<td>-0.16</td>
<td>-0.01</td>
</tr>
<tr>
<td>7</td>
<td>572</td>
<td>(18.4, 15.3)</td>
<td>(14.6, 14.5)</td>
<td>0.18</td>
<td>0.40</td>
</tr>
<tr>
<td>8</td>
<td>1056</td>
<td>(18.6, 16.2)</td>
<td>(13.6, 13.7)</td>
<td>0.31</td>
<td>0.95</td>
</tr>
<tr>
<td>9</td>
<td>638</td>
<td>(17.6, 16.8)</td>
<td>(14.2, 14.6)</td>
<td>0.04</td>
<td>0.76</td>
</tr>
</tbody>
</table>

**Table:** Description of Pseudodata in Glaucoma study
<table>
<thead>
<tr>
<th>p</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>center = 8</td>
<td>SIMPLE†</td>
<td>-2.45</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>100% missing</td>
<td>-1.58</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>90% missing</td>
<td>-1.50</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>80% missing</td>
<td>-2.37</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>50% missing</td>
<td>-2.61</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>20% missing</td>
<td>-2.19</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>No missing</td>
<td>-2.33</td>
<td>0.22</td>
</tr>
</tbody>
</table>

**Table:** Estimate treatment effect on IOP. †: Based on complete data before any deletion.
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.
• Explore the extent of efficiency gain with respect to the models that describe the relationship among $T$, $S$ and $Z$

• Propose a shrinkage approach

$x = \text{observed } \ ? = \text{unobserved}$

<table>
<thead>
<tr>
<th>Single Trial</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z$</td>
<td>$S$</td>
<td>$T$</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>$?$</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>$?$</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>$?$</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>$?$</td>
</tr>
</tbody>
</table>
Potential for gain in efficiency from using $S$ is due to

- The use of more data ($S$) gives extra information
- Assumptions in a model
S and $T$ are continuous.
The joint distribution of $S$ and $T$ given $Z$ can be expressed as:

$$T_i = \beta_0 + \beta_1 S_i + \beta_2 Z_i + \beta_3 S_i Z_i + \epsilon_{ti}$$
$$S_i = \alpha_0 + \alpha_1 Z_i + \epsilon_{si}$$

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 $S$

\[ T_i = \beta_0 + \beta_1 S_i + \beta_2 Z_i + \beta_3 S_i Z_i + \epsilon_{ti} \]
\[ S_i = \alpha_0 + \alpha_1 Z_i + \epsilon_{si} \]

- 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 $f(T|S)$

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

- Expect $\beta_2$ and $\beta_3$ 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 \( \beta_2 \).
  - Fully Bayesian version (Ridge-FB): \( \beta_2 \sim N(0, \sigma_{\beta_2}^2) \).
  - Empirical Bayesian version (Ridge-EB)
Proposed Method: Generalized Ridge Regression

The Model:

\[ T_i = \beta_0 + \beta_1 S_i + \beta_2 Z_i + \epsilon_{ti} \]

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

\[
\begin{align*}
E(\hat{\beta} | X_t, T) &= (X_t^T X_t + K)^{-1} X_t^T T, \\
V(\hat{\beta} | X_t, T) &= (X_t^T X_t + K)^{-1} \sigma_t^2.
\end{align*}
\]

where \( \beta^T = (\beta_0, \beta_1, \beta_2) \), \( X_t = (1, S, Z) \), \( K = \text{diag}(0, 0, k_2) \) and \( k_2 = \sigma_t^2 / \sigma_{b_2}^2 \).
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: $\beta_2 = \beta_3 = 0$.

<table>
<thead>
<tr>
<th>Treatment Effect (Q)</th>
<th>PES</th>
<th>APAS</th>
<th>IPAS</th>
<th>Ridge FB</th>
<th>Ridge EB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_2 = 0, Q = 2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimate</td>
<td>2.029</td>
<td>2.042</td>
<td>2.050</td>
<td>2.037</td>
<td>2.040</td>
</tr>
<tr>
<td>MSE</td>
<td>0.133</td>
<td>0.207</td>
<td>0.216</td>
<td>0.156</td>
<td>0.172</td>
</tr>
<tr>
<td>$\beta_2 = 2, Q = 4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimate</td>
<td>3.402</td>
<td>4.042</td>
<td>4.050</td>
<td>3.885</td>
<td>3.963</td>
</tr>
<tr>
<td>MSE</td>
<td>0.527</td>
<td>0.207</td>
<td>0.216</td>
<td>0.236</td>
<td>0.217</td>
</tr>
</tbody>
</table>

Table: $n = 120, \beta_0 = 0.5, \beta_1 = 1, \alpha_0 = 1, \alpha_1 = 2, \sigma_I^2 = 0.1, \sigma_S^2 = 0.5, \text{miss} = 80\%$
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).
Figure: Asymptotic Relative Efficiency Compared with that Obtained from Original Data.
Simulations: Compare Ridge, IPW and Model Selection Regarding MSE

- **n1=60, n2=60**
  - CC
  - Ridge-FB
  - Ridge-EB
  - IPW
  - MdlSel

- **n1=120, n2=120**
  - CC
  - Ridge-FB
  - Ridge-EB
  - IPW
  - MdlSel

- **n1=480, n2=480**
  - CC
  - Ridge-FB
  - Ridge-EB
  - IPW
  - MdlSel
Simulations: Compare Ridge and Model Selection Regarding Coverage Rates

- **n1=60, n2=60**
- **n1=120, n2=120**
- **n1=480, n2=480**
Glaucoma Randomized Trial

- Surgery versus Medicine
- n=600
- $S =$ intra-ocular pressure at 12 months
- $T =$ intra-ocular pressure at 102 months
### Table: Summary Statistics from Glaucoma Trial.

<table>
<thead>
<tr>
<th>Observed at 12th</th>
<th>Medicine</th>
<th>Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing at 102nd Month</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Patients</td>
<td>206</td>
<td>207</td>
</tr>
<tr>
<td>Mean IOP at 12th Month</td>
<td>18.2</td>
<td>14.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observed at 12th</th>
<th>Medicine</th>
<th>Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed at 102nd Month</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Patients</td>
<td>86</td>
<td>74</td>
</tr>
<tr>
<td>Mean IOP at 12th Month</td>
<td>17.9</td>
<td>14.1</td>
</tr>
<tr>
<td>Mean IOP at 102nd Month</td>
<td>17.5</td>
<td>15.1</td>
</tr>
</tbody>
</table>
Glaucoma Randomized Trial

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

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


