Increasing efficiency for estimating treatment-biomarker interactions with historical data

Philip S. Boonstra

University of Michigan
20 May 2014
Predictive biomarkers in randomized phase II trials

Targeted therapies are, by design, expected not to work equally well for all patients. Biomarker(s) may be predictive for treatment response, e.g. higher levels mean higher treatment sensitivity.
Can characterize “predictiveness” through treatment-biomarker interaction in statistical model for response, but modest size of typical phase II trials makes estimating interaction difficult.
Randomized Phase II trial at Michigan

- Overexpression of ETS transcription factors \textit{ERG} and \textit{ETV1} is characterized by a fusion of the androgen-sensitive promoter from a prostate-specific gene and an ETS gene [Tomlins et al., 2005].
- Advanced prostate cancer patients evaluated for ETS transcription factors in the metastatic lesion.
Randomized Phase II trial at Michigan
Randomized Phase II trial at Michigan

- ETS-mediated oncogenic features such as metastasis and tumor growth depend on PARP1 [Brenner et al., 2011].
- Thus a PARP1-inhibitor can specifically target ETS-positive prostate cancers
Randomized Phase II trial at Michigan

The trial will evaluate the addition of PARP1-inhibitor to a standard treatment regime (Abiraterone).
Randomized Phase II trial at Michigan

- Advanced prostate cancer
- ETS fusion +: Randomize Abiraterone +/- PARP1-inhibitor
- ETS fusion -: Randomize Abiraterone +/- PARP1-inhibitor
- $n = 120$
- Need large sample sizes to estimate interactions
Historical data may increase statistical efficiency for estimating treatment-biomarker interactions in randomized phase II trials. Two such types consist of

- patients who all received the standard of care but for which biomarker measurements are available
- patients who received either the standard of care or experimental treatment but for which biomarker measurements are missing
Historical data to increase efficiency

- Strategies exist to estimate *main* effect of treatment [e.g., Thall and Simon, 1990, Neuenschwander et al., 2010]
- Bias in treatment effect estimates is possible if historical and prospective data are incompatible [DuMouchel and Harris, 1983, Cuffe, 2011]. Will discuss this more at the end.
1. Define prospective data and 2 types of historical data
2. Continuous response/biomarker setting
   2.1 Formulate response model
   2.2 Propose a relative efficiency (RE) metric based on Fisher Information to quantify gain from historical data
   2.3 Evaluate RE both numerically and analytically
3. Binary response/biomarker setting – Preliminary results
4. Interpretation and caveats
Data collected

- **Response** $Y$ (continuous, binary)
- **Prognostic Covariates** $W \in \mathbb{R}^q$
- **Treatment** $T \in \{0, 1\}$
- **Biomarker** $V$ (continuous, binary)
Sources of data

Group 1  Prospective randomized trial: $n_1$ observations consisting of $\{Y, W, T, V\}$.

Group 2  Previous biomarker study, lacking the experimental therapy: $n_2$ observations consisting of $\{Y, W, T \equiv 0, V\}$.

Group 3  Previous observational or clinical trial of experimental therapy: $n_3$ observations consisting of $\{Y, W, T\}$.
Response model – continuous setting

- **Response** $Y \in \mathbb{R}$
- **Prognostic Covariates** $W \in \mathbb{R}^q$
- **Treatment** $T \in \{0, 1\}$
- **Biomarker** $V \in \mathbb{R}$

A Gaussian-linear model for the response $Y$ is

$$Y|X \sim N(X^\top \beta, \sigma^2),$$

with $X = \{1, W^\top, T, V, T \times V\}^\top$, $\beta = \{\beta_0, \beta_W^\top, \beta_T, \beta_V, \theta\}^\top$ and $\beta_W = \{\beta_{W1}, \beta_{W2}, \ldots, \beta_{Wq}\}^\top$.

$$E(Y|X) = \beta_0 + \beta_W^\top W + \beta_T T + \beta_V V + \theta TV$$
Response model – continuous setting

- **Response** $Y \in \mathbb{R}$
- **Prognostic Covariates** $W \in \mathbb{R}^q$
- **Treatment** $T \in \{0, 1\}$
- **Biomarker** $V \in \mathbb{R}$

A Gaussian-linear model for the response $Y$ is

$$Y|X \sim N(X^\top \beta, \sigma^2),$$

with $X = \{1, W^\top, T, V, T \times V\}^\top$, $\beta = \{\beta_0, \beta_w^\top, \beta_T, \beta_V, \theta\}^\top$ and

$$\beta_w = \{\beta_{w1}, \beta_{w2}, \ldots, \beta_{wq}\}^\top.$$

$$E(Y|X) = \beta_0 + \beta_w^\top W + \beta_T T + \beta_V V + \theta TV$$

$\theta$ is parameter of interest
To account for missing data in group 3, we assume a biomarker model

\[ V | W \sim N(\gamma_0 + \gamma^\top W, \tau^2) \]

and marginalize the response model over \( V \). Conditional independence: \( T \perp V | W \).
Biomarker model – continuous setting

To account for missing data in group 3, we assume a biomarker model

\[ V|W \sim N(\gamma_0 + \gamma^\top W, \tau^2) \]

and marginalize the response model over \( V \). Conditional independence: \( T \perp V|W \).

Note: Groups 2 and 3 by themselves provide no information about \( \theta \) itself.
Differences between groups

Two assumptions allow for population-level differences in groups:

1. Prognostic covariates \((W)\), which are available in all groups, explain some differences between populations.

2. For differences not accounted for by \(W\), we allow intercepts, \(\beta_0\) and \(\gamma_0\), to differ between populations.
Variance estimate of $\hat{\theta}$

1. Log-likelihood contributions from each group: $\ell_1, \ell_2, \ell_3$

2. $I_k = E \begin{bmatrix} \frac{\partial \ell_k}{\partial \beta} \frac{\partial \ell_k}{\partial \beta^\top} & \frac{\partial \ell_k}{\partial \beta} \frac{\partial \ell_k}{\partial \gamma^\top} \\ \frac{\partial \ell_k}{\partial \gamma} \frac{\partial \ell_k}{\partial \beta^\top} & \frac{\partial \ell_k}{\partial \gamma} \frac{\partial \ell_k}{\partial \gamma^\top} \end{bmatrix}$

3. $I$ obtained from linear combination of $I_1, I_2$ and $I_3$

4. $\text{Var}(\hat{\theta})$ obtained from diagonal element of $(I^{-1})$
Fisher information: group 1

Expected information contribution from one observation in group 1:

\[
I_1 = \frac{1}{\sigma^2} \begin{pmatrix}
1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
W & WW^\top & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
T & TW^\top & T & \cdot & \cdot & \cdot & \cdot & \cdot \\
V & VW^\top & TV & V^2 & \cdot & \cdot & \cdot & \cdot \\
TV & TVW^\top & TV & TV^2 & TV^2 & \cdot & \cdot & \cdot \\
0 & 0_q^\top & 0 & 0 & 0 & 1/(2\sigma^2) & \cdot & \cdot \\
0 & 0_q^\top & 0 & 0 & 0 & 0 & \sigma^2/\tau^2 & \cdot \\
0 & 0_q^\top & 0 & 0 & 0 & 0 & 0 & \sigma^2/(2\tau^4) \\
0_q & 0_q & 0_q^\top & 0_q & 0_q & 0_q & (\sigma^2/\tau^2)W & (\sigma^2/\tau^2)WW^\top \\
0 & 0_q^\top & 0 & 0 & 0 & 0 & 0 & 0_q^\top & \sigma^2/(2\tau^4)
\end{pmatrix}
\]

\[
\begin{pmatrix}
\beta_0 & \beta_W & \beta_T & \beta_V & \theta & \sigma^2 & \gamma_0 & \gamma & \tau^2
\end{pmatrix}
\]
Fisher information: group 2

Expected information contribution from one observation in group 2:

\[ I_2 = \frac{1}{\sigma^2} \begin{pmatrix} 1 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ W & WW^T & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0_q^T & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ V & VW^T & 0 & V^2 & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0_q^T & 0 & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & 0_q^T & 0 & 0 & 0 & 1/(2\sigma^2) & \cdots & \cdots & \cdots \\ 0 & 0_q^T & 0 & 0 & 0 & \sigma^2/\tau^2 & \cdots & \cdots & \cdots \\ 0_q & 0_q & 0_q^T & 0_q & 0_q & 0_q & (\sigma^2/\tau^2)W & (\sigma^2/\tau^2)WW^T & \cdot \\ 0_q & 0_q & 0_q^T & 0_q & 0_q & 0_q & 0 & 0 & \sigma^2/(2\tau^4) \\ \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_w \\ \beta_T \\ \beta_V \\ \theta \\ \sigma^2 \\ \gamma_0 \\ \gamma \\ \tau^2 \end{pmatrix} \]
Fisher information: group 3

Expected information contribution from one observation in group 3:

\[
I_3 = \frac{1}{\delta^2} \begin{pmatrix}
1 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
W & WW^\top & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
T & TW^\top & T & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\tilde{V} & \tilde{V}W^\top & T\tilde{V} & \tilde{V}^2 + 2\tau^4 D^2 / \delta^2 & \cdots & \cdots & \cdots & \cdots & \cdots \\
T\tilde{V} & T\tilde{V}W^\top & T\tilde{V} & T\tilde{V}^2 + 2T\tau^4 D^2 / \delta^2 & T\tilde{V}^2 + 2T\tau^4 D^2 / \delta^2 & \cdots & \cdots & \cdots & \cdots \\
0 & 0^\top & 0 & \tau^2 D / \delta^2 & T\tau^2 D / \delta^2 & 1/(2\delta^2) & \cdots & \cdots & \cdots \\
D & DW^\top & TD & \tilde{V}D & T\tilde{V}D & 0 & D^2 & \cdots & \cdots \\
DW & DWW^\top & TDW & \tilde{V}DW & T\tilde{V}DW & 0_q & D^2 W & D^2 WW^\top & \cdots \\
0 & 0^\top & 0 & \tau^2 D^3 / \delta^2 & T\tau^2 D^3 / \delta^2 & D^2 / (2\delta^2) & 0 & 0 & D^4 / (2\delta^2) \\
\beta_0 & \beta_W & \beta_T & \beta_V & \theta & \sigma^2 & \gamma_0 & \gamma & \tau^2 \\
\end{pmatrix}
\]

with \(D = \beta_V + T\theta\), \(\tilde{V} = \mathbb{E}[V|W] = \gamma_0 + W^\top \gamma\), and \(\delta = (\tau^2 D^2 + \sigma^2)^{1/2}\). The additional off-diagonals result from marginalizing over the missing \(V\).
Numerical study

$q = 5$ prognostic covariates. For each group, 100,000 draws of

1. $W \sim N(0_q, I_q)$,

2. $T|W \sim \text{Bin}(1/2)$ (group 1)
   or $T|W \sim \text{Bin}(\expit\{\alpha_0 + \alpha^\top W\})$ (group 3),

3. $V|W \sim N(\gamma_0 + \gamma^\top W, \tau^2)$ (for groups 1 and 2),

and calculate the empirical average values of $I_1, I_2,$ and $I_3$.

Choose $\alpha_0$ and $\gamma_0$ so that $E[T] = 0$ and $E[V] = 0$. Modify the signal with coefficients of determination, $R^2_\sigma$ and $R^2_\tau$. 
Define the following measures of relative efficiency for estimating $\theta$, given some non-negative number $k$:

\[
RE_2(k) = \{(I_1 + kI_2)^{-1}\}_\theta / \{I_1^{-1}\}_\theta,
\]
\[
RE_3(k) = \{(I_1 + kI_3)^{-1}\}_\theta / \{I_1^{-1}\}_\theta,
\]
\[
RE_{23}(k) = \{(I_1 + 0.5k[I_2 + I_3])^{-1}\}_\theta / \{I_1^{-1}\}_\theta,
\]
\[
RE_{\text{max}}(k) = \{(I_1 + kI_1)^{-1}\}_\theta / \{I_1^{-1}\}_\theta = 1/(1 + k).
\]

$k$ is the sample size ratio, e.g. $n_2/n_1$ or $n_3/n_1$, for the historical data compared to the prospective data.
Relative Efficiencies

Efficiency gains from historical data under varying $R^2$'s for response (row, $R^2_{\sigma}$) and biomarker (column, $R^2_{\tau}$) models; partial-$R^2$ from $\theta$ is in upper corner of each panel.
Relative Efficiencies

Efficiency gains from historical data under varying $R^2$’s for response (row, $R^2_\sigma$) and biomarker (column, $R^2_\tau$) models; partial-$R^2$ from $\theta$ is in upper corner of each panel.
Relative Efficiencies

Efficiency gains from historical data under varying $R^2$’s for response (row, $R^2_\sigma$) and biomarker (column, $R^2_\tau$) models; partial-$R^2$ from $\theta$ is in upper corner of each panel.
Analytical evaluation of group 2 contribution

Group 2: Previous biomarker study, lacking experimental therapy: $n_2$ observations consisting of $\{Y, T \equiv 0, V, W\}$. 
Analytical evaluation of group 2 contribution

Group 2: Previous biomarker study, lacking experimental therapy: \( n_2 \) observations consisting of \( \{Y, T \equiv 0, V, W\} \).

Write variance of \( \hat{\theta} \) from group 1 + group 2 as

\[
v_{\hat{\theta}}(\mu_T, n_1, k, \Omega) \\
\equiv (1/n_1)\{(I_1 + kI_2)^{-1}\}_\theta \\
= \ldots \\
= \left(\frac{1}{n_1}\right) \left(\frac{\sigma^2}{\gamma^\top \Sigma_W \gamma + \tau^2}\right) \left(\frac{1 + k}{\mu_T(1 + k) - \mu_T^2}\right),
\]

\( \mu_T \equiv E[T] \in (0, 1) \) (group 1), \( k = n_2/n_1 \), and \( \Omega \) denotes all other parameters.
Maximum possible efficiency gain from group 2

\[
\lim_{k \to \infty} \frac{v_\hat{\theta}(\mu_T, n_1, k, \Omega)}{v_\hat{\theta}(\mu_T, n_1, 0, \Omega)} = \lim_{k \to \infty} \frac{(1 + k)(\mu_T - \mu_T^2)}{\mu_T(1 + k) - \mu_T^2} = 1 - \mu_T.
\]

When \( \mu_T = 1/2 \) (an equal-arm prospective trial), group 2 can reduce the variance of \( \hat{\theta} \) by no more than 1/2.
Approaching a single-arm trial

\( v(\mu_T, n_1, k, \Omega) \) is minimized with respect to \( \mu_T \) by using
\[
\mu_T^{\text{opt}} = \frac{1}{2} + \frac{k}{2} = \frac{1}{2} + \frac{n_2}{2n_1}.
\]
If \( n_2 > n_1 \), then prospective trial should have single arm: the experimental therapy. Group 2 provides all the controls.
Approaching a single-arm trial

\(v_\theta(\mu_T, n_1, k, \Omega)\) is minimized with respect to \(\mu_T\) by using
\[
\mu_T^{\text{opt}} = \frac{1}{2} + \frac{k}{2} = \frac{1}{2} + \frac{n_2}{2n_1}.
\]
If \(n_2 > n_1\), then prospective trial should have single arm: the experimental therapy. Group 2 provides all the controls.

Theoretical result – caveats at end of talk.
Response model – binary setting

- **Response** \( Y \in \{0, 1\} \)
- **Prognostic Covariates** \( W \in \mathbb{R}^q \)
- **Treatment** \( T \in \{0, 1\} \)
- **Biomarker** \( V \in \{0, 1\} \)

A logistic model for the response \( Y \) is

\[
\Pr(Y = 1|X) = \text{expit}\{X^\top \beta\},
\]

with \( X = \{1, W^\top, T, V, T \times V\}^\top \), \( \beta = \{\beta_0, \beta_W^\top, \beta_T, \beta_V, \theta\}^\top \) and \( \beta_W = \{\beta_{W1}, \beta_{W2}, \ldots, \beta_{Wq}\}^\top \).
To account for missing data in group 3, we assume a biomarker model

$$
\Pr(V = 1|\mathbf{W}) = \expit\{[1, \mathbf{W}^\top] \mathbf{\gamma}\}
$$
Fisher Information

Similar strategy as before: numerically determine $I_k$, invert to obtain large-sample variance of $\theta$. When $V$ is measured, we have

$$I(\beta, \gamma) = \begin{pmatrix} r_Y^2 XX^\top & r_Y r_V X [1, W^\top] \\ r_Y^2 [1, W^\top]^\top [1, W^\top] & r_V^2 \end{pmatrix},$$

where $r_Y = (Y - \expit\{X^\top \beta\})$ and $r_V = (V - \expit\{[1, W^\top] \gamma\})$; Fisher information is now a function of both the response and the parameters.
Fisher Information

Similar strategy as before: numerically determine $I_k$, invert to obtain large-sample variance of $\theta$. When $V$ is measured, we have

$$I(\beta, \gamma) = \begin{pmatrix} r_Y^2 XX^\top & r_Yr_VX[1, W^\top] \\ r_Yr_VX[1, W^\top] & r_V^2 [1, W^\top]^\top [1, W^\top] \end{pmatrix},$$

where $r_Y = (Y - \expit\{X^\top \beta\})$ and $r_V = (V - \expit\{[1, W^\top] \gamma\})$; Fisher information is now a function of both the response and the parameters.

When $V$ is unmeasured, required to first marginalize over missing $V$. 
numerical study

\( q = 5 \) prognostic covariates. For each group, 100,000 draws of

1. \( W \sim N(0_q, I_q) \),

2. \( T|W \sim \text{Bin}(1/2) \) (group 1)
   or \( T|W \sim \text{Bin}(\expit\{\alpha_0 + \alpha^\top W\}) \) (group 3),

3. \( V|W \sim \text{Bin}(\expit\{[1, W^\top]\gamma\}) \) (for groups 1 and 2),

4. \( Y|X \sim \text{Bin}(\expit\{X^\top \beta\}) \)

and calculate the empirical average values of \( I_1, I_2, \) and \( I_3 \).

Choose \( \alpha_0 \) and \( \gamma_0 \) so that \( \mathbb{E}[T] = 0 \) and \( \mathbb{E}[V] = 0.2 \).
Relative Efficiencies

Sample size ratio (k)

$$RE \begin{cases} 0.3 \\ 0.5 \\ 0.7 \\ 0.9 \end{cases} = \theta \begin{cases} 0.5 \\ \beta V^{0.5} \end{cases} \begin{cases} 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 \end{cases} = \theta \begin{cases} 1.5 \\ \beta V^{0.5} \end{cases} \begin{cases} 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 \end{cases}$$

Sample size ratio (k)

$$RE_{2}$$ $$RE_{3}$$ $$RE_{23}$$ $$RE_{\text{max}}$$

$0$ $1$ $2$ $3$

$\theta = 0.5$ $\theta = 1.5$

$\beta_v = 0.5$

$\beta_v = 0$

Sample size ratio (k)
Interpretation, caveats

1. Initial efficiency gain observed from both types of historical data.
Interpretation, caveats

1. Initial efficiency gain observed from both types of historical data.
2. Efficiency gains depend on models being “similar enough”*, although not identical.

[*“similar enough” means that $\beta$ and $\theta$ are equal between groups. However, may be able to relax this through a Bayesian analysis, i.e. commensurate or power priors.*]
Interpretation, caveats

1. Initial efficiency gain observed from both types of historical data.

2. Efficiency gains depend on models being “similar enough”*, although not identical.

*“similar enough” means that $\beta_V$ and $\theta$ are equal between groups. However, may be able to relax this through a Bayesian analysis, i.e. commensurate or power priors.
Interpretation, caveats

3. Not always realistic or desirable to prescribe a single-arm trial; interactions are not be-all/end-all
Interpretation, caveats

3. Not always realistic or desirable to prescribe a single-arm trial; interactions are not be-all/end-all

4. A multiplicative interaction parameter is neither sufficient (may not be *clinically* meaningful) nor necessary (rather a more complicated subgroup effect) for a predictive biomarker
5. Large-sample investigation demonstrates a proof of principle: more data can increase efficiency.
Interpretation, caveats

5. Large-sample investigation demonstrates a proof of principle: more data can increase efficiency.

6. Next step is to investigate small-sample properties of estimators.
5. Large-sample investigation demonstrates a proof of principle: more data can increase efficiency.
6. Next step is to investigate small-sample properties of estimators.
7. Also worthwhile to extend to censored outcomes.
Acknowledgments

Collaborators: Jeremy M G Taylor, Bhramar Mukherjee

Funding: Eli Lilly and Company


