Predictive Biomarkers with Observational Data

Noah Simon and Richard Simon

July 2016

Observational Data

Often do not have a randomized clinical trial to build/evaluate a predictive biomarker.

On non-randomized observational data, may be open to confounding

Two Parts

There are two issues we will need to tackle; Combating confounding in:

- Constructing a predictive biomarker
- Evaluating our predictive biomarker

Two Parts

There are two issues we will need to tackle; Combating confounding in:

- Constructing a predictive biomarker
- Evaluating our predictive biomarker

We will begin with a discussion of evaluation; assuming a fixed biomarker.

Confounding - I

What is confounding in this context?

It is actually almost identical to the classical setting; testing for average treatment effect (without biomarker):

Suppose we have a fixed binary biomarker z...

and are interested in evaluating

$$E_{tr=T}[y|z=1] - E_{tr=C}[y|z=1]$$

Just average treatment effect in the fixed subpopulation $\{z=1\}$.

Confounding - II

Suppose also have w, a confounding variable...

ie. w is correlated with both tr and y in our $z = 1$ subgroup.

In particular, assume that there is no unobserved confounding

Need to adjust for w!

Confounding - II

Suppose also have w, a confounding variable...

ie. w is correlated with both tr and y in our $z = 1$ subgroup.

In particular, assume that there is no unobserved confounding

Need to adjust for w!

Example: w may be some observable health metric used by doctors to decide aggressiveness of therapy.

Combating Confounding in Evaluation

Two basic methods to correct:

- Regression-Based Correction
- Inverse Prob Weighted/Propensity Score-based Correction

Regression-Based Correction

One form of correction can be done via regression:

Using only biomarker+ patients, fit a model

$$\mathsf{E}\left[y|tr, w, z = 1\right] = \beta_0 + \beta_1 I\left\{tr = T\right\} + \gamma^\top w$$

and test β_1...

Can use logistic/cox equivalent for binary/survival response

Regression-Based Correction

One form of correction can be done via regression:

Using only biomarker+ patients, fit a model

$$\mathsf{E}\left[y|tr, w, z = 1\right] = \beta_0 + \beta_1 I\left\{tr = T\right\} + \gamma^\top w$$

and test β_1...

Can use logistic/cox equivalent for binary/survival response

Warning! Strong assumptions about form of confounding

Regression-Based Correction

Could use more flexible model

$$\mathrm{E}\left[y|tr, w, z = 1\right] = \beta_0 + \beta_1 I\left\{tr = T\right\} + h(w)$$

with general h.

Regression-Based Correction

Could use more flexible model

$$\mathrm{E}\left[y|tr, w, z = 1\right] = \beta_0 + \beta_1 I\left\{tr = T\right\} + h(w)$$

with general h. Or even

$$\mathrm{E}\left[y|tr, w, z = 1\right] = G\left(I\left\{tr = T\right\}, w\right)$$

Here though; non-homogenous treatment effect; so must average treatment effect over w

General form known as *G-computation*; requires some semi-parametric tomfoolery.

Propensity Score-Based Correction

Alternative correction done using propensity scores:

$$p_T(w) = \mathrm{P}(tr = T | w, z = 1)$$
$$p_C(w) = \mathrm{P}(tr = C | w, z = 1)$$

We estimate *biomarker+ average treatment effect* using inverse probability weighting (IPW):

$$\hat{ATE} = \frac{1}{n+} \left[\sum_{z_i=1, tr=T} \frac{y_i}{p_T(w_i)} - \sum_{z_i=1, tr=C} \frac{y_i}{p_C(w_i)} \right]$$

where $n+$ is the number of biomarker+ patients.

Propensity Score-Based Correction

Alternative correction done using propensity scores:

$$p_T(w) = \mathrm{P}(tr = T | w, z = 1)$$
$$p_C(w) = \mathrm{P}(tr = C | w, z = 1)$$

We estimate *biomarker+ average treatment effect* using inverse probability weighting (IPW):

$$\hat{ATE} = \frac{1}{n+}\left[\sum_{z_i=1, tr=T} \frac{y_i}{p_T(w_i)} - \sum_{z_i=1, tr=C} \frac{y_i}{p_C(w_i)}\right]$$

where $n+$ is the number of biomarker+ patients.

This has a simple normal limit; so easy to test if ATE non-zero.

Propensity Score-Based Correction

$$\hat{ATE} = \frac{1}{n+}\left[\sum_{z_i=1, tr=T} \frac{y_i}{p_T(w_i)} - \sum_{z_i=1, tr=C} \frac{y_i}{p_C(w_i)}\right]$$

In practice $p_T(w)$, $p_C(w)$ unknown, so must estimate.

- Adds variance (which can be accounted for)
- If estimated flexibly, can mess up normality

(requires augmented IPW)

Comparison of Corrections

G-computation

- requires a joint model for y, tr, and w ✗
- only takes differences of estimated quantities ✓

IPW estimator

- only requires joint model for tr and w ✓
- uses estimates in the denominator ✗

IPW issues

If propensity scores in denominator get sufficiently close to 0; estimator explodes.

If propensity score model is overfit this will happen;

May also happen under model misspecification (or from outliers).

Summary of Evaluation

Evaluation of a predictive biomarker is just a test of average treatment effect (in biomarker+)

Many tools already created for this...

Just apply those tools!

Combating Confounding in Biomarker Construction

In a randomized trial we consider

$$\mathrm{E}_{T}\left[y|x\right] - \mathrm{E}_{C}\left[y|x\right]$$

This ignores the confounder; and can make incorrect decisions in its presence

Simple Example

Suppose w binary (eg. disease severity),

$$\mathrm{E}_T\left[y|x, w=1\right] = \mathrm{E}_T\left[y|x, w=0\right] - 2\delta$$
$$\mathrm{E}_C\left[y|x, w=1\right] = \mathrm{E}_C\left[y|x, w=0\right] - 2\delta$$

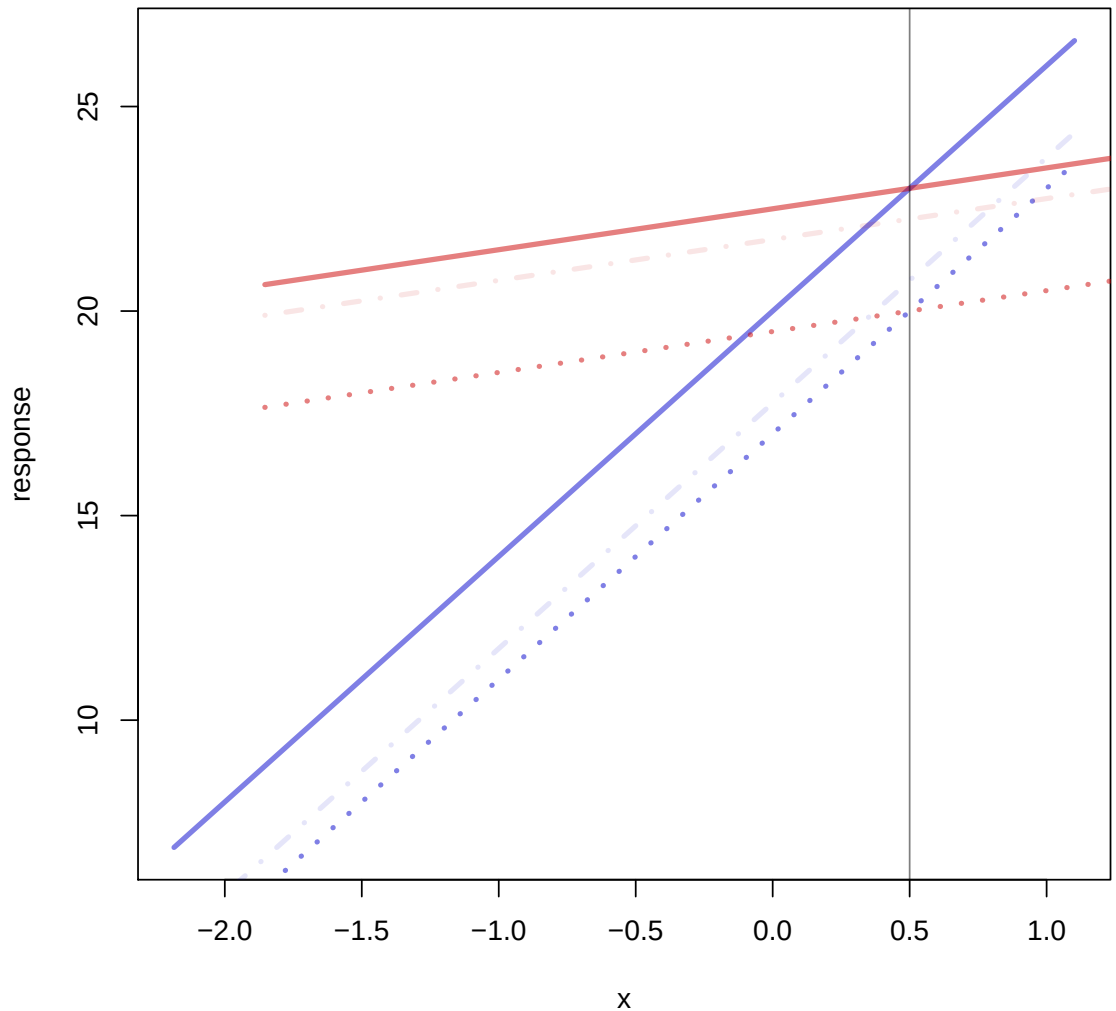
and for all x

$$P_T\left(w=1|x\right) = 0.75 \qquad P_C\left(w=1|x\right) = 0.25$$

Even in this simple scenario, ignoring w leads to incorrect cutpoint.

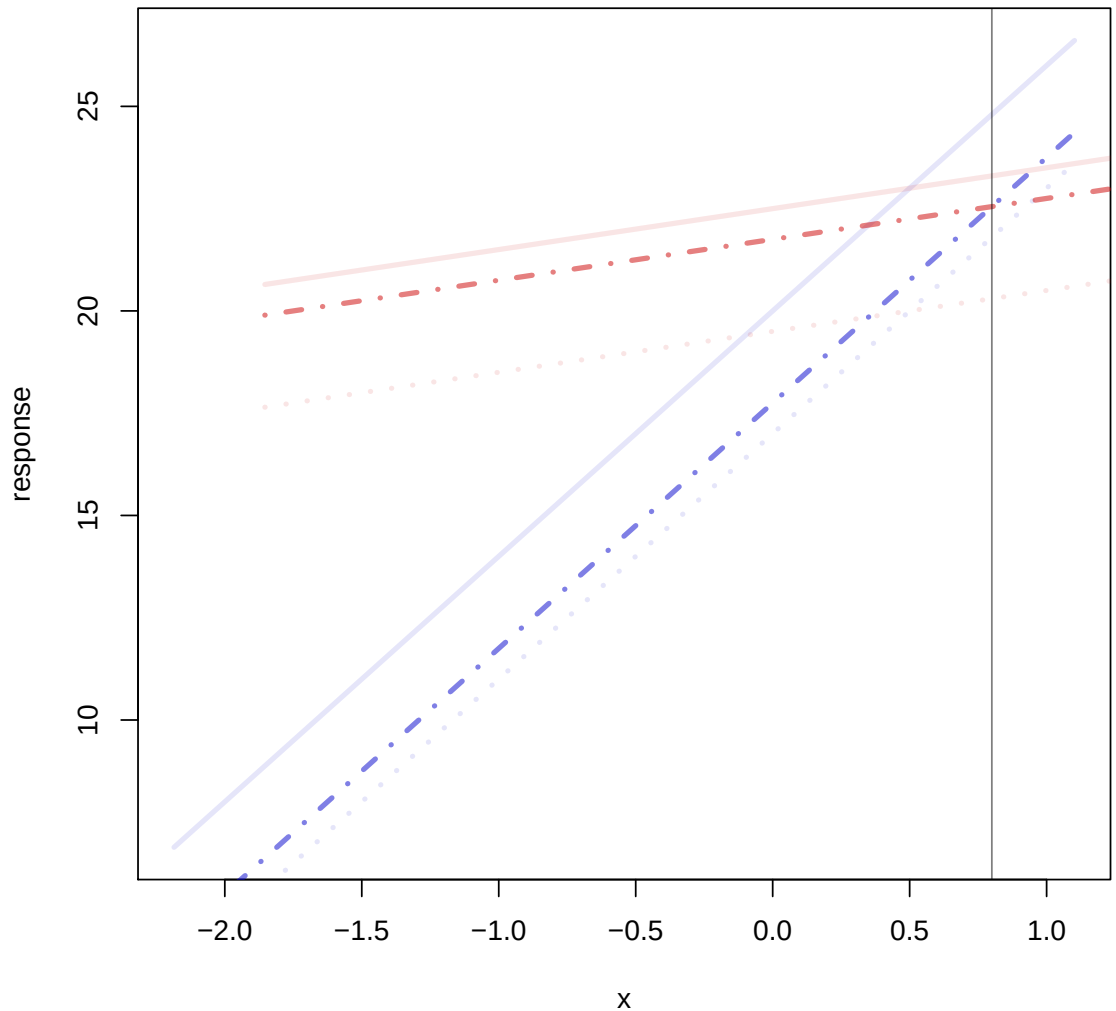
Simple Example

Conditional on w we see



Simple Example

Averaging over w we see



Simple Example

This issue occurs because we are averaging over w differently for treatment and control

We can fix this by

- Not averaging over w (ie. conditioning on w)
- Standardizing how we average over w.

Conditioning on w

To fix things we combine x and w and form our biomarker using both

ie. Find (x, w) with

$$\mathrm{E}_T\left[y|x,w\right] - \mathrm{E}_C\left[y|x,w\right]$$

Standardizing how we average over w

We leave this as an exercise for the student.

The recipe (propensity scores)

1. Split our data into two subsets: Training and Evaluation
2. On Training data:
   2.1 Estimate $\mathrm{E}_T[y|x,w]$ and $\mathrm{E}_C[y|x,w]$
   2.2 Define our binary classifier $z(x,w)$ by

$$z(x,w) = I\left\{\hat{\mathrm{E}}_T[y|x,w] - \hat{\mathrm{E}}_C[y|x,w] > 0\right\}$$

3. On full data, estimate propensity scores
4. Test using IPW for samples with $z(x_i, w_i) = 1$ in Evaluation subset

The recipe (regression)

1. Split our data into two subsets: Training and Evaluation
2. On Training data:
   - 2.1 Estimate $\mathrm{E}_T[y|x,w]$ and $\mathrm{E}_C[y|x,w]$
   - 2.2 Define our binary classifier $z(x,w)$ by

$$z(x,w) = I\left\{\hat{\mathrm{E}}_T[y|x,w] - \hat{\mathrm{E}}_C[y|x,w] > 0\right\}$$

3. On Evaluation subset, using only observations with $z(x_i, w_i) = 1$, fit the model (or equivalent)

$$y = \beta_0 + \beta_1 tr + \gamma^\top w + \epsilon$$

4. Test if β_1 is non-zero

Caution!

This is a *better* way to analyze observational data than ignoring confounding...

But there is always unmeasured confounding and model misspecification!

Findings here still need to be validated in RCT!