Applied example

Rachael Caelie (Rocky) Aikens

12/2/2020

Set-up

We'd like to ask the question: Do patients have better outcomes when their identities match their own? To do so, we use a dataset of 1155903 coronary artery bypass grafting (cabg) surgeries on Medicare patients from 1998 to 2016. We compare 30 day mortality for patients whose identities are matched to their primary surgeons with those whose identities are not.

Provider race is not recorded in the dataset, so we focus on sex/gender. Unfortunately, patients are characterized only by "sex" and providers are characterized only by "gender," and the data dictionaries don't really clarify what is meant by this. The remainder of this analysis will treat patient sex and provider gender match as "treated" and mismatch as "untreated."

Data cleaning

The original dataset contained 1155903 surgeries. 58 patients had a missing date for their surgery. 247686 had a recorded date, but no recorded gender information for their provider. One observation had sex recorded as unknown and was removed from the analysis. Otherwise, no patients had any recorded sex besides male or female, and no providers had any gender codes other than male or female. 908159 surgeries were included.

Most pre-treatment covariates are more than 20% missing. Only ischemia, hyperlipidimia, and hypertension are less than 80% missing. I recode admissions codes to be "Emergency", "Urgent", "Elective" or "Other", since codes besides emergency, urgent, or elective are quite rare in the dataset.

Analysis

Pooled

First, I reserved 10% of the control data as a pilot set to train a prognostic model. Then I trained a propensity score model on the entire dataset. The plot below shows an assignment-control plot from the fitted propensity and prognostic scores. In order to avoid overplotting, only a subsample of 400 observations is shown.
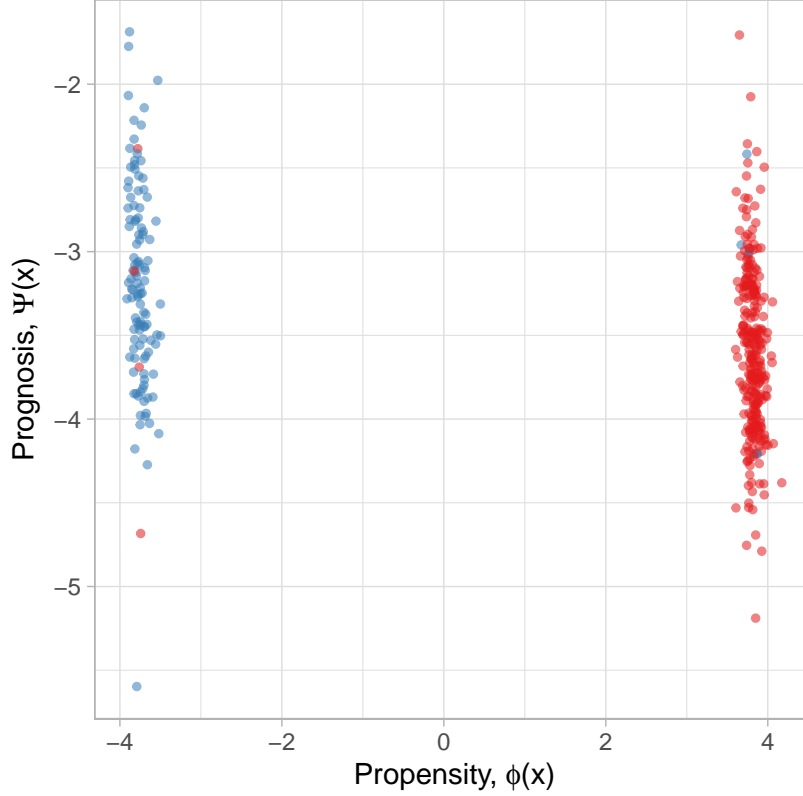
Right away something is obviously amiss. Female patients are much more likely to be mismatched with their primary surgeon than male patients (see table below). Because of this stark contrast in propensity for treatment between these two groups, I stratify by patient gender and make separate plots for each of them. In order to do so, I split both the pilot and analysis sets by patient sex and I estimate new prognostic and propensity score models for each of the two groups. Note that there is a large imbalance in pilot set size between the two groups because male patients are almost all "treated" and female patients are almost all "control." In fact, there are ~1300 male controls in the pilot set and ~30000 female controls in the pilot set. This is a tough position to be in because that means a prognostic model fit on the whole pilot set will be much better fit to females, but if we split the pilot set and fit two sex-specific prognostic models, the male prognostic model will probably be much worse fit than the female prognostic model.

BENE_MALE	SEX_GENDER_MATCH	n()
FALSE	FALSE	294494
FALSE	TRUE	6944
TRUE	FALSE	13082
TRUE	TRUE	593638

Shown another way, here are the fraction of treated individuals bamong men and women:

BENE_MALE	Percent Treated
FALSE	0.0230362
TRUE	0.9784382

2

Males

Here's an assignment control plot for males. In this plot, both the prognostic and propensity score models were fit on males only. The propensity score model is fit on all males, while the prognostic score model is fit on only the males in the pilot set. Again, I'm downsampling the data to avoid overplotting.
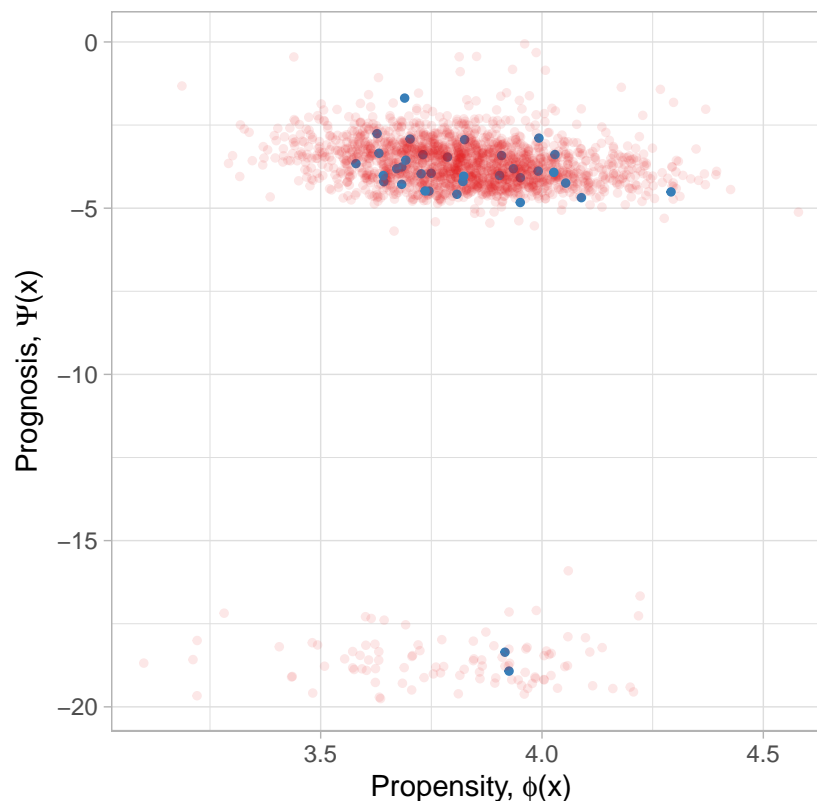

Why the big striations in prognostic score? Sometimes this is normal. When there are categorical covariates that are highly predictive of the outcome under the control assignment or the treatment assignment, this is reflected by striae or subgroups in the assignment control plot. Sometimes this is an important indication because it suggests a variable or variables that may be important candidates for exact matching or stratification.

However, in this specific instance, it's more likely that the prognostic model is just overfit - attributing large weight to categorical covariates that may not actually be meaningful. Part of the problem is probably that the outcome (30 day mortality) is quite rare, meaning that the model is difficult to fit, even with over 1,000 examples. The table below shows the outcomes among men in the pilot set. Since only 42 individuals in the pilot set had the outcome of interest, it is perhaps unsurprising that we have a hard time obtaining a clear picture of which variation is most important to the outcome. We will see that the prognostic score fit on the female pilot set sample is much smoother – probably because overfitting is less of a concern with such a large pilot set.

MORT_30_DAY	n()
FALSE	1262
TRUE	41

Females

Here's the same plot for the female patients, with prognostic and propensity models fit only for female patients.
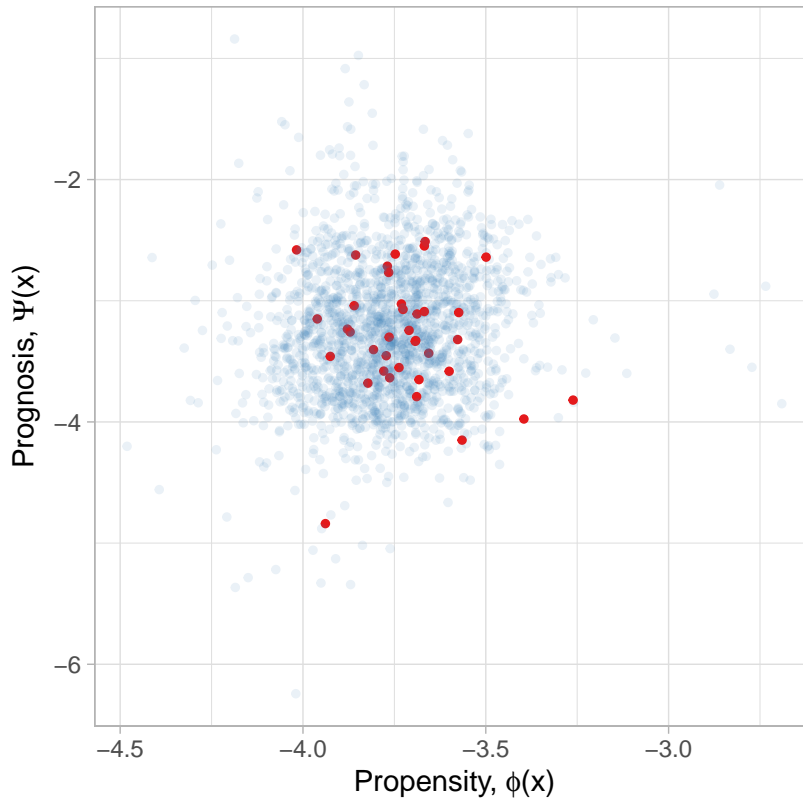
The plot for female patients looks a lot better. This is probably because there are many more female control observations in the pilot set on which to fit a prognostic model. The table below shows the outcomes of female patients in the pilot set. Notice that there are many more observations with the outcome, helping with model fitting.

MORT_30_DAY	n()
FALSE	28196
TRUE	1252