An alternative model for diagnosis probability

Rachael Caelie (Rocky) Aikens

7/14/2020

Setup

A model for Longitudinal Disease Progression

Now let's imagine we're dealing with a disease that begins at a low severity and progresses every year. We'll assume that severity progresses from 0 to 1 according to a sigmoid function with the following parameters:

$$R_i(t) = \frac{1}{1 + e^{-\beta_i(t-T_i)}}.$$

Here, β_i and T_i are intrinsic to the individual:

- β_i is the slope of progression. A larger β_i indicates a quickly progressing disease.
- T_i controlls the time of onset. T_i is the time when severity reaches 0.5

We'll simulate β_i and T_i from the following distributions:

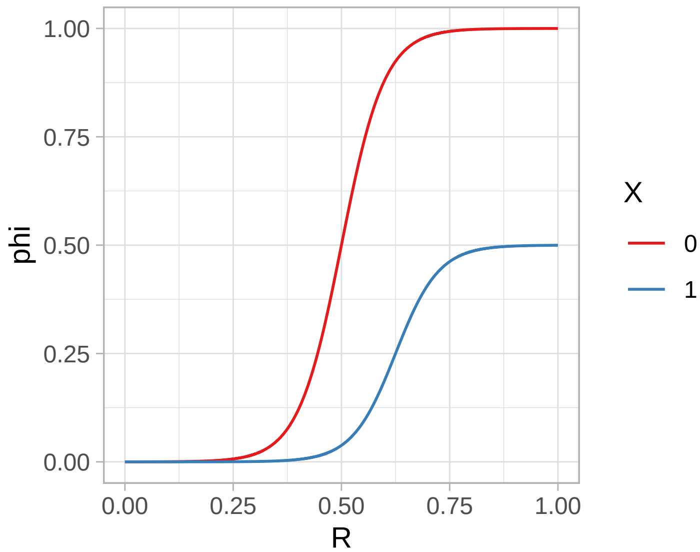
$$\beta_i \sim_{iid} \text{Beta}(\alpha = 2, \beta = 6)$$
$$T_i \sim_{iid} \text{Normal}(\mu = 50, \sigma = 5)$$

Model: Diagnosis rate depends on severity and baseline covariates

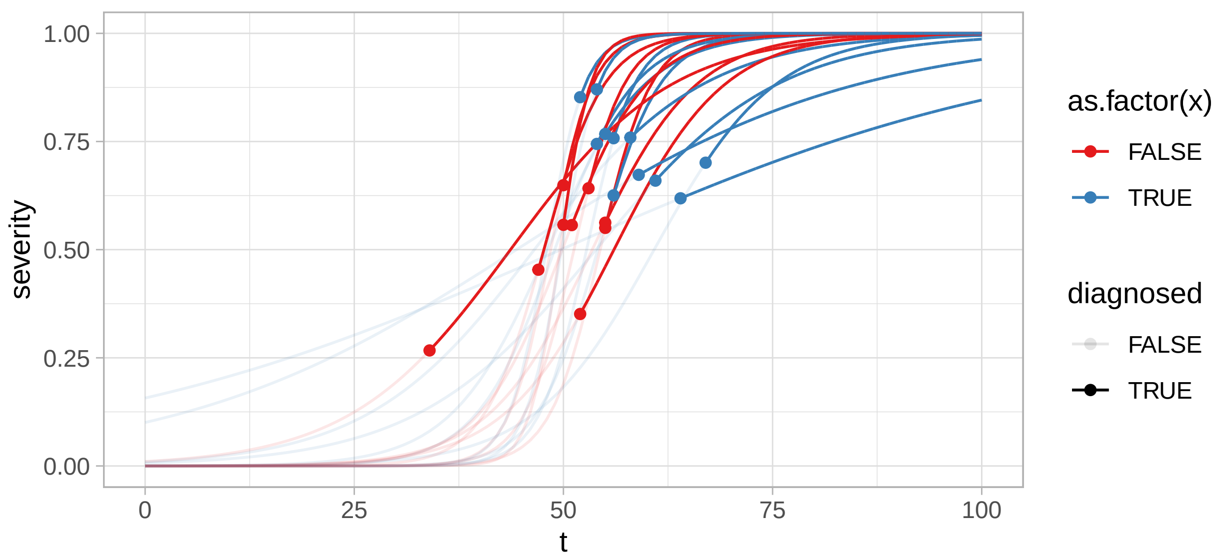
Now, let's suppose that the probability of diagnosis depends on severity and baseline characteristics:

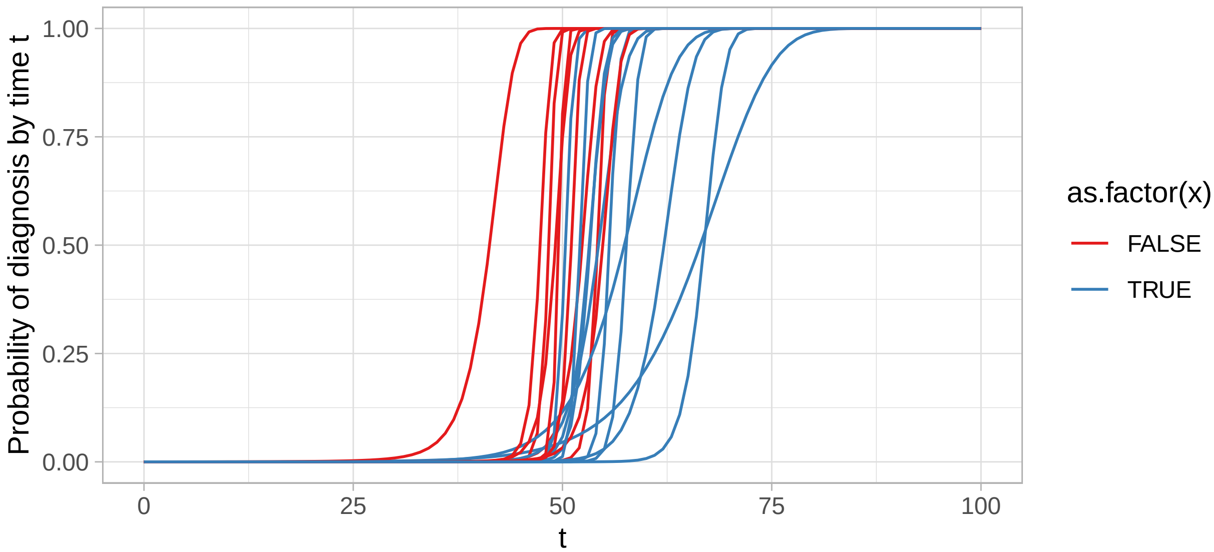
$$\phi(R_i, X_i) = \frac{1 + \beta_2 X_i}{1 + exp(-(\beta_0 + \beta_1 R_i))} - c_{x_i},$$

where again, $c_{x_i} = \frac{1}{1+exp(-(\beta_0+\beta_2 X_i))}$ is a corrective constant to ensure that $\phi(0, X_i) = 0$. Also, let $\beta_2 = -\frac{1}{2}$. Inessence, this means that people with $X_i = 0$ are diagnosed in the same way as above, but people with $X_i = 1$ are diagnosed with a lower probability, shown below:



An example of how this plays out is shown below on a subset of the data. People with $X = 0$ tend to be diagnosed earlier in their disease progression than people with $X = 1$.





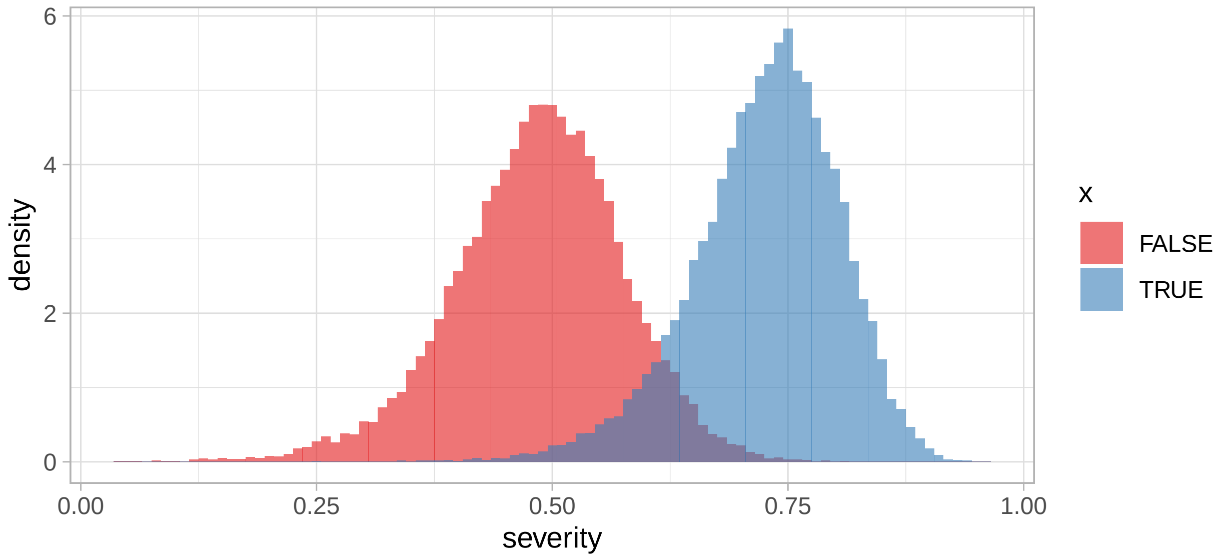
Under this model (where diagnosis probability depends on X), let's see what we glean from a naive study of these data.

Let's suppose there are 5000 individuals with this disease in the EHR. We don't observe their underlying disease state; just their diagnosis. We also don't observe the severity of their illness prior to diagnosis.

```
## # A tibble: 2 x 2
##   x         n
##   <lgl> <int>
## 1 FALSE 25295
## 2 TRUE  24676
```

```
## # A tibble: 2 x 2
##   x         n
##   <lgl> <int>
## 1 FALSE 25295
## 2 TRUE  24705
```

In this set-up, 49971 individuals out of 5000 with the disease are diagnosed. All of the undiagnosed individuals have $X = 1$. The plot below shows the severity at onset for individuals with $x = 0$ compared to $x = 1$. This probably is indicative of the fact that the people with $x = 0$ tend to be diagnosed at the start of their disease onset, while the people with $x = 1$ tend to be diagnosed later in their disease progression.



6
4
2
0
density
0.00
0.25
0.50
0.75
1.00
severity
x
FALSE
TRUE