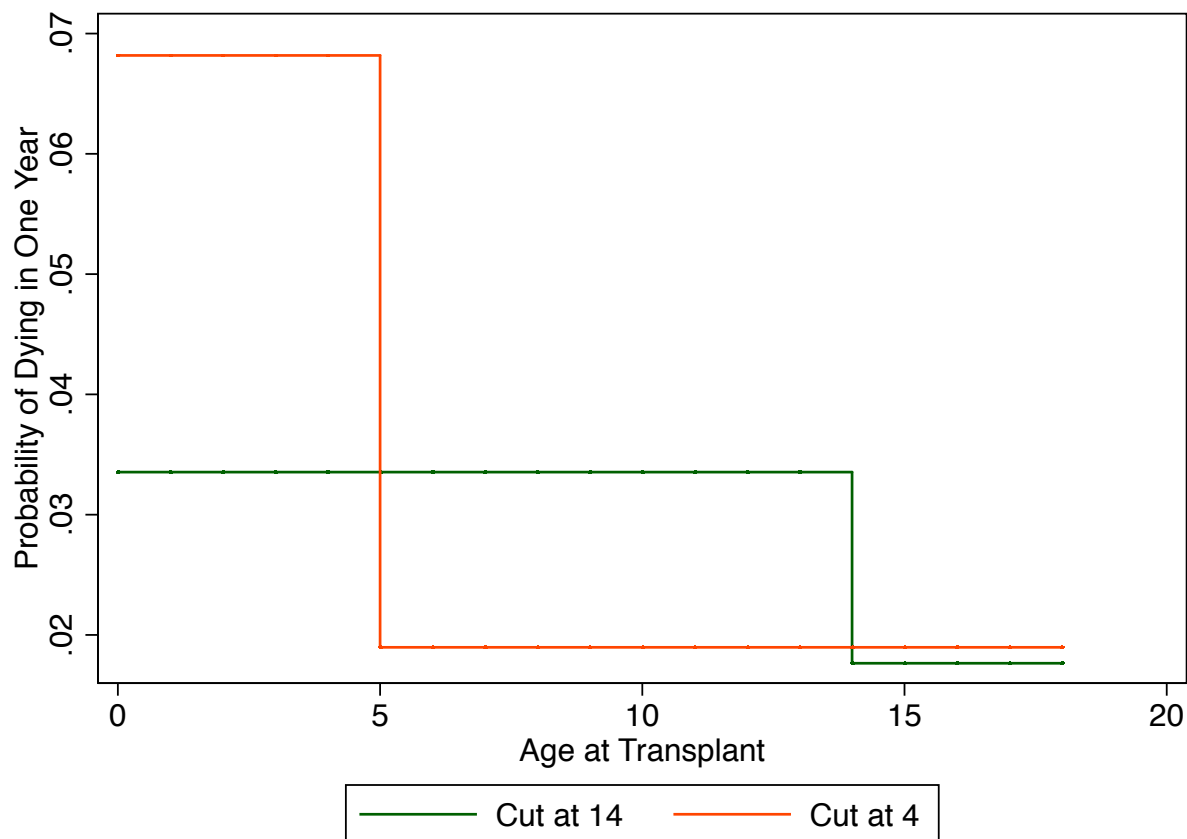


Lab #1 Discussion

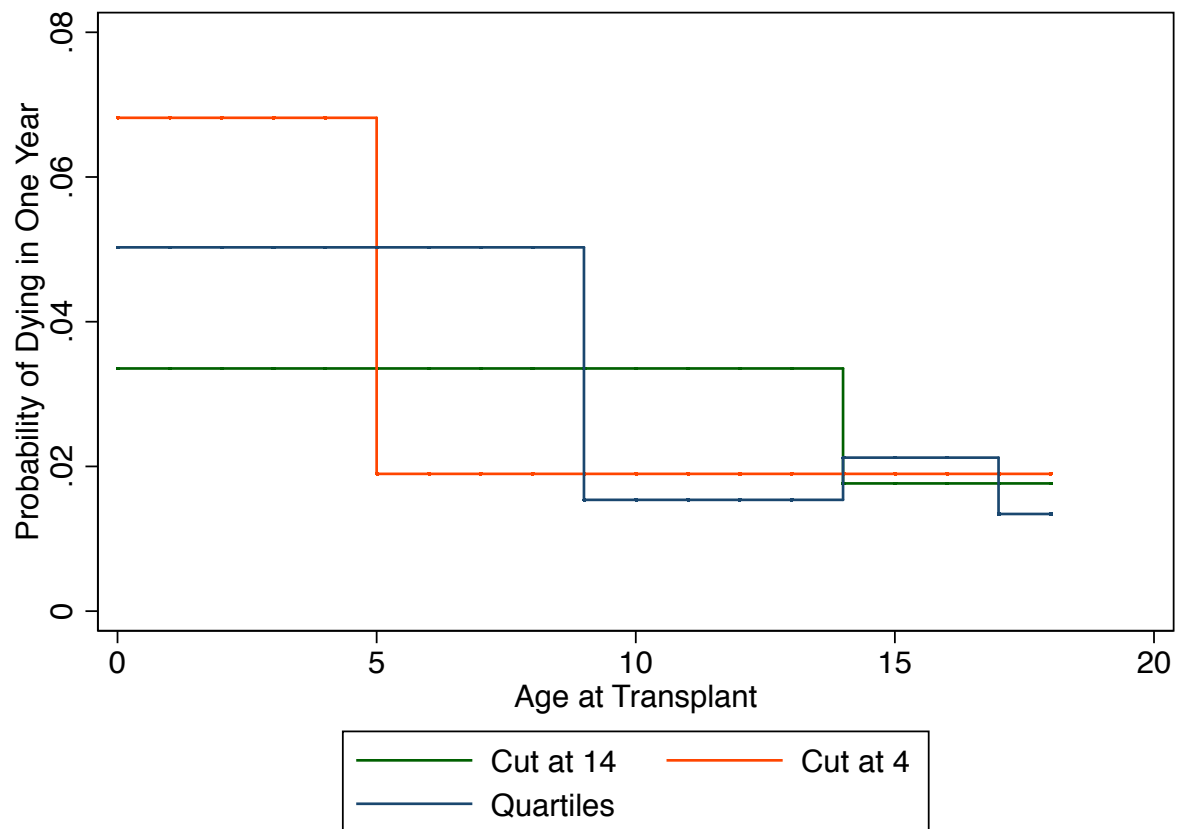
Biostatistics 210

Q1: How do the fits for your two binary models compare graphically?

For this I used the cut at the median (around) 14 years of age and the cut at 4 years of age. The two curves look quite different to me. This underlies my point that the choice of cutpoint has big implications for the results and that is one of the (many) reasons they are to be avoided. Always.

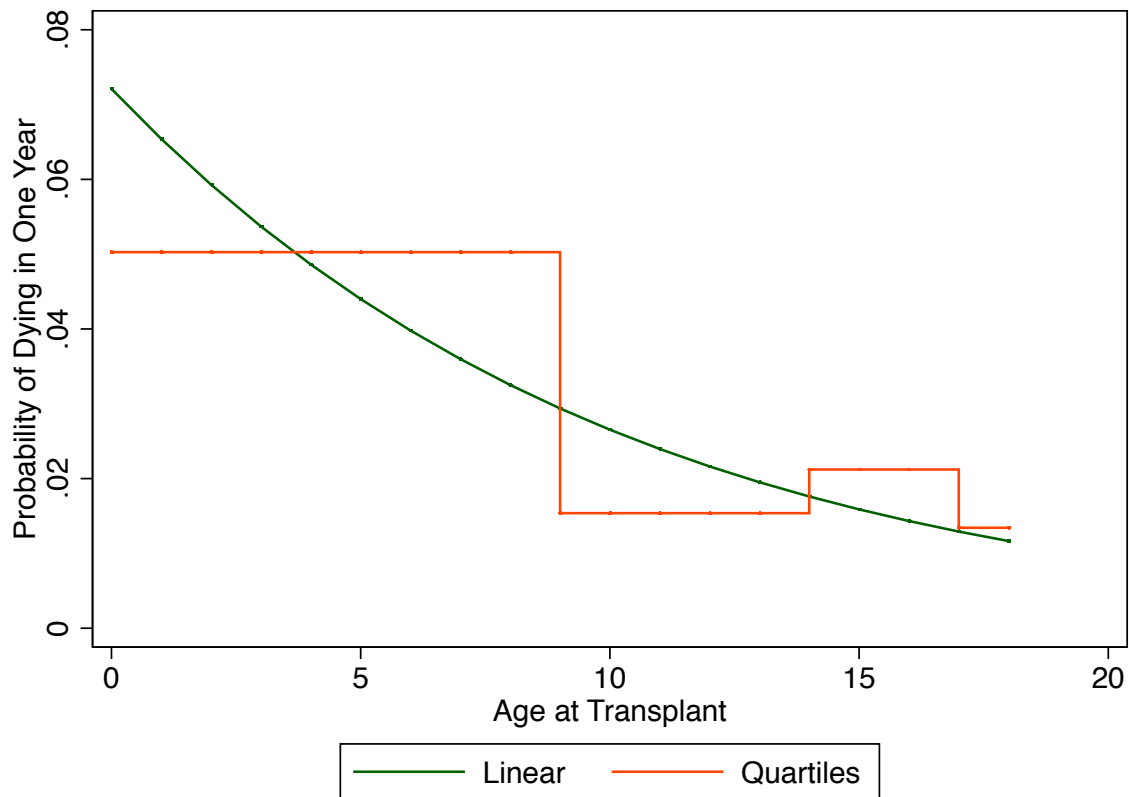


Q2: How does this fit compare graphically with your binary fits?



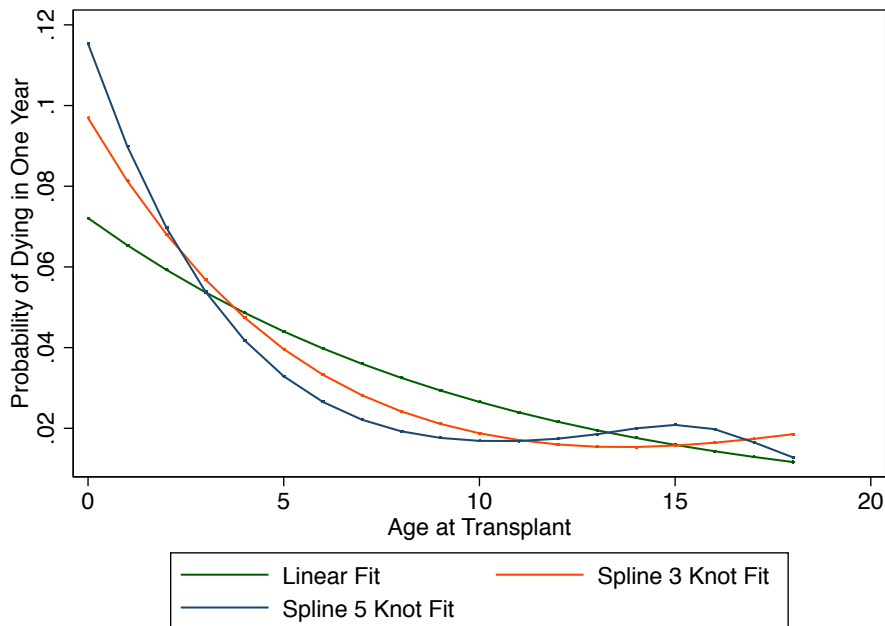
The quartile fit is graphed on the next page and we see it gives a much richer picture of what is going on. It gives us the sense that there may be high mortality below the age of 10 which appears reduced and largely stable after age 10.

Q3: How do the categorical and linear fits compare, do you think the linear assumption appears to be fitting the data well?



It doesn't appear to fit the data well. There is some suggestion of a systematic departure. It appears that the linear fit underestimates the effect of age at younger ages and overestimates the effects at older ages. This is worth investigating. Splines will be a real help.

Q4: How do the spline and linear fits compare, do you think the linear assumption appears to be fitting the data well?



I graph a 3 knot and 5 knot spline fit. They both resemble the pattern we saw with the cutpoints. Fairly high and quickly declining mortality for children less than 5 years of age which then levels off around age 10. The question we then have to address is whether this is substantively (clinically) important. If the primary purpose of our analysis was prediction, we might gain very little from adding in the splines. The cross-validated area under the ROC curves are actually not so different for the 3 fits. If our purpose was only to focus on other variables and control for confounding, the linear age might also be adequate. However, if age were one the major predictors in the study, then we might consider presenting these results using the spline fit. However, I would not have a problem with someone presenting a quartile cutpoint for this data because they would simply be finding an alternative way of presenting the pattern seen with the splines. I think it is important to bring a heavy dose of external information to this in judging whether such a pattern is plausible. It is also important to consider if this pattern is driven by confounding which is why I have you adjust for some confounders in the next set of commands. It is clear that the departure from linearity is statistically significant -- the LR test rejects the hypothesis that there is no difference between a spline and linear model:

```
.lrtest A
```

```
Likelihood-ratio test
```

```
LR chi2(1) = 12.33
```

```
(Assumption: . nested in A)
```

```
Prob > chi2 = 0.0004
```

Q5: Does this change your conclusions?

Nope.

```
. lrtest B
```

Likelihood-ratio test

LR chi2(1) = 6.44

(Assumption: . nested in B)

Prob > chi2 = 0.0111

