Objectives

The aim is to investigate which variables are associated with survival in women diagnosed and treated for breast cancer.
A full description of the survival outcomes and the variables are shown in table 1.

Method

Strategy of selecting appropriate model

The first task was to decide which survival model was appropriate for my analysis. This was done by first seeing if an exponential model will be suitable, I conducted a weibull regression on the data to produce table 2, as shown, none of the gamma estimate confidence intervals included 1, hence it is statistically significant that an exponential model will not be suitable. I then categorised the continuous data in the model (age, estrogen, progesterone, number of nodes and size of tumours) into discrete groups by using their corresponding 25%, 50% and 75% quantile, I then produced log log plots of all the predictors, as seen in figure 1. From this we can see that none of the graphs show a convincing set of parallel lines with most of them intersecting hence violating the proportional hazards assumption, as a result I decided to use a cox proportional hazards model rather than a weibull model as the former uses less assumptions than the latter.

Some of the predictors in the study did not explicitly show a linear relationship with the associated risk of death or tumour recurrence, hence a spline model was used for these variables (see results section for more details).

To obtain the final multivariable model we used backwards elimination to fit the most appropriate multivariable model.
Survival analysis approach is going to be taken rather than logistic regression to take into account of censored data in the study.

Results

Individually analysing predictors on outcome

A statistically significantly increased risk of death or tumour recurrence was observed for the tumour grade, size of tumour and number of positive lymph nodes involved. Specifically, having a tumour grade at medium compared to a low grade tumour has more than double the associated risk from reoccurring a tumour or dying at 125% increase in risk (hazard ratio 2.254, 95% confidence interval 1.37 to 3.71; P=0.001). The high grade tumour had an even greater effect compared to the low grade with more than triple the associated risk at a 203% increase in risk (hazard ratio 3.027, 95% confidence interval 1.777 to 5.155 ; P<0.001). The size of tumour was associated with a 1.6% increase in risk for every mm in size (hazard ratio 1.016, 95% confidence interval 1.009 to 1.023; P<0.001).The number of positive lymph nodes had a 6.1% increase in associated risk (hazard ratio 1.061, 95% confidence interval 1.046 to 1.076; P<0.001) for every additional positive lymph node. See table 3 for a summary.

In contrast there was a statistically significant decreased risk of death or tumour recurrence observed for patients that receive hormone therapy and having a higher progesterone receptor status. Specifically, patients that received hormone therapy had a decreased risk of 34.1% compared to patients who did not receive hormone therapy (hazard ratio 0.659, 95% confidence interval 0.508 to 0.857; P=0.002). For every additional femtomole of progesterone, there was also a 0.2% decrease in risk (hazard ratio 0.998, 95% confidence interval 0.996 to 0.999; P<0.001).

There was no linear relationship established with estrogen and age against the outcome variable, hence a spline model was used to see if this could provide a better fit, the log ratio test shows that the model with splines on estrogen and age were a better fit, specifically the spline estrogen model had a P value of 0.005 for the log ratio test showing that it is a significantly better fit than the old model. The spline model of age provided a P value of <0.001 demonstrating it is also a much better fit than the current . The plots are shown in figure 2 demonstrate the spline relationship, however the relationships are unclear between the outcome and predictors of estrogen and age hence these covariates will be omitted from the multivariable model.

Multivariable model

From performing a multiple regression cox model with the following equation:

$$h(t) = h(0)exp(\beta_h hormone + \beta_p progesterone + \beta_{g2} grade2 + \beta_{g3} grade3 + \beta_{size} size + \beta_{nodes} nodes)$$

We can perform backwards elimination, we omit tumour size as our variable (because P>0.2 for size).

As a result we obtain the final multivariable model:

$$h(t) = h(0)exp(\beta_h hormone + \beta_p progesterone + \beta_{g2} grade2 + \beta_{g3} grade3 + \beta_n nodes)$$

(Where h(t) = hazard ratio of died/tumour recurrence, h(0) = base hazard ratio, hormone= Hormone treatment status; progesterone=Progesterone receptor status (fmol) ;grade=Tumour grade, size=Tumour size (mm), nodes=Number of positive lymph nodes involved)

This gives the results as shown in table 3. After adjusting for hormone therapy, progesterone and nodes there was a significant decrease in the hazard ratio of grade medium and especially high grade, with a decrease from triple the associated risk increase to just over twice the associated risk, this can highlight that tumour grade can be a confounding variable. All of the other predictors did not change drastically and hence should be contained in the final model.

Conclusion

The study exposes the elevated risks a patient could have from recurring a tumour or dying - most notably from the grade of tumour, how many positive lymph nodes involved and the size of the tumour, however when considered in a multivariable model the risk exposed in the size of the tumour is no longer statistically significant. The grade of the tumour in particular stood out, as a high grade tumour resulted in triple the risk exposed compared to a low grade tumour, however in the multivariable model this risk was reduced to twice the associated risk compared to the low grade tumour whilst holding other predictors constant, hence further analysis is required to investigate other potential variables for the model as this can be classified as a confounding variable. In contrast a person who received hormone therapy had substantially reduced risk compared to someone who didn't receive hormone therapy. A higher progesterone receptor status were also associated with decreased risk of tumour recurrence/death. Age and estrogen receptor status did not have a clear relationship associated with risk of tumour recurrence/death and has hence been omitted from the multivariable model. These findings have important implications for developing potential tumour recurrence prevention strategies (such as advocating hormone therapy) and having a detailed understanding of the risks an individual patient has after being diagnosed and treated for breast cancer.