Lung Cancer Survival Analysis

Andrew Kinsman, 9 June 2016

Introduction

This project examines the survival experiences of patients with advanced lung cancer (based on data collected by the North Central Cancer Treatment Group in the US from the late 1980s). The sample comprised 228 individuals (ranging in age from 39 to 82) of whom 138 were men and 90 women. Over the course of the study 165 of these patients died and 63 were right-censored (in other words, their final survival time is unknown, either they were lost to follow-up or survived right through to the end of the study period). Survival times ranged from five to 1022 days.

There are 10 variables in the data: survival time, censoring status (i.e. whether or not the observation is right-censored) and eight covariates (predictor variables). These covariates relate to demographic information about the patient (age, sex and the medical institution they attended), the patient's mealtime calorie intake and recent weight loss, and also some performance status scores that relate to how well the patient can perform usual daily activities (as assessed by both the physician and the patient themselves).

Although data for age and sex were complete, all of the other explanatory variables contained at least one missing value. For the Cox proportional hazards model below, listwise deletion was employed to remove these, which reduced the number of observations to 167.

1. Assess whether the survival experience between males and females are different.

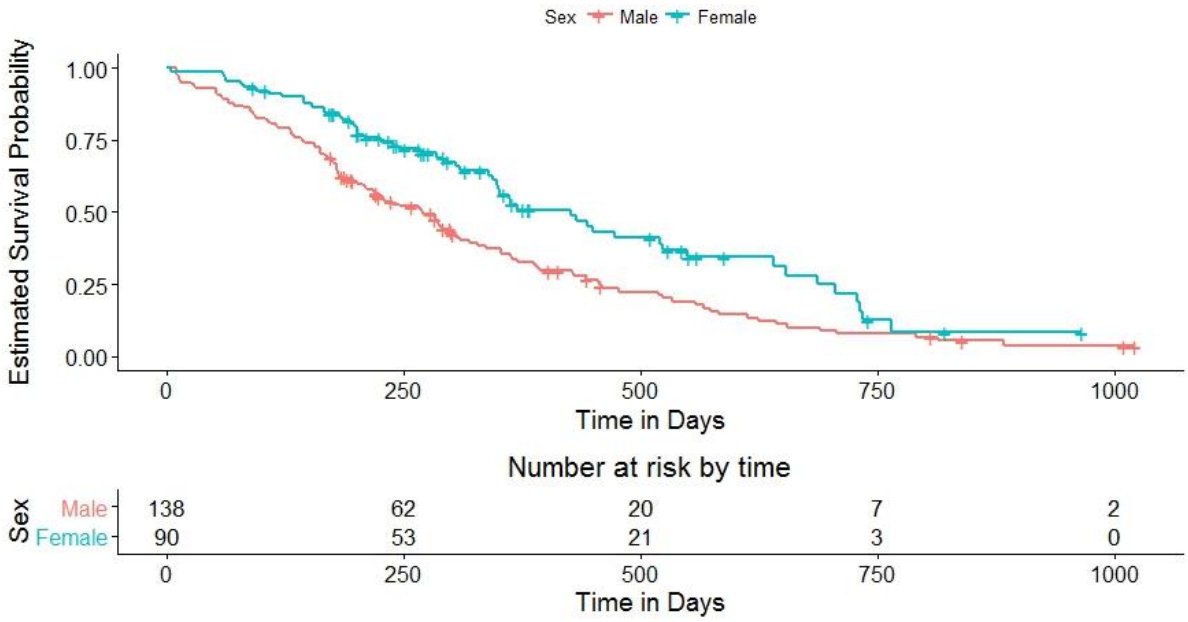
A Kaplan-Meier (product-limit) model was built to compare gender survival experiences. This survival model estimates the proportion of patients who remain alive at each particular point in time after joining the study. Rather than group times into intervals, the Kaplan-Meier approach uses the exact survival times of each individual, with the value of the survival function between successive distinct sampled observations assumed to be constant, so the Kaplan-Meier estimates typically take on a "step" appearance as opposed to a smooth curve. It was also assumed here (and in the Cox model below) that the individuals who were lost to follow-up were not systematically different to those who remained in the study. In other words, that the reason for these observations being censored is purely down to random chance and not something endemic to the study.

Examining the quantiles from this model, it appears that there is a 50% chance that a man with survive for at least 270 days (95% confidence interval of 212, 310 days), but this rises to at least 426 days (95% confidence interval of 348, 550 days) for women. Note that the median rather than the mean is used to estimate these survival times, since it is a more robust metric when the data are skewed, as is invariably the case with survival analysis.

We can also use the model to estimate annual survival rates, since these common benchmarks are easily interpretable by non-statisticians. The survival rate after one year is 33.6% for men and 52.6% for women, while after two years these respective rates are 7.8% and 18.7%.

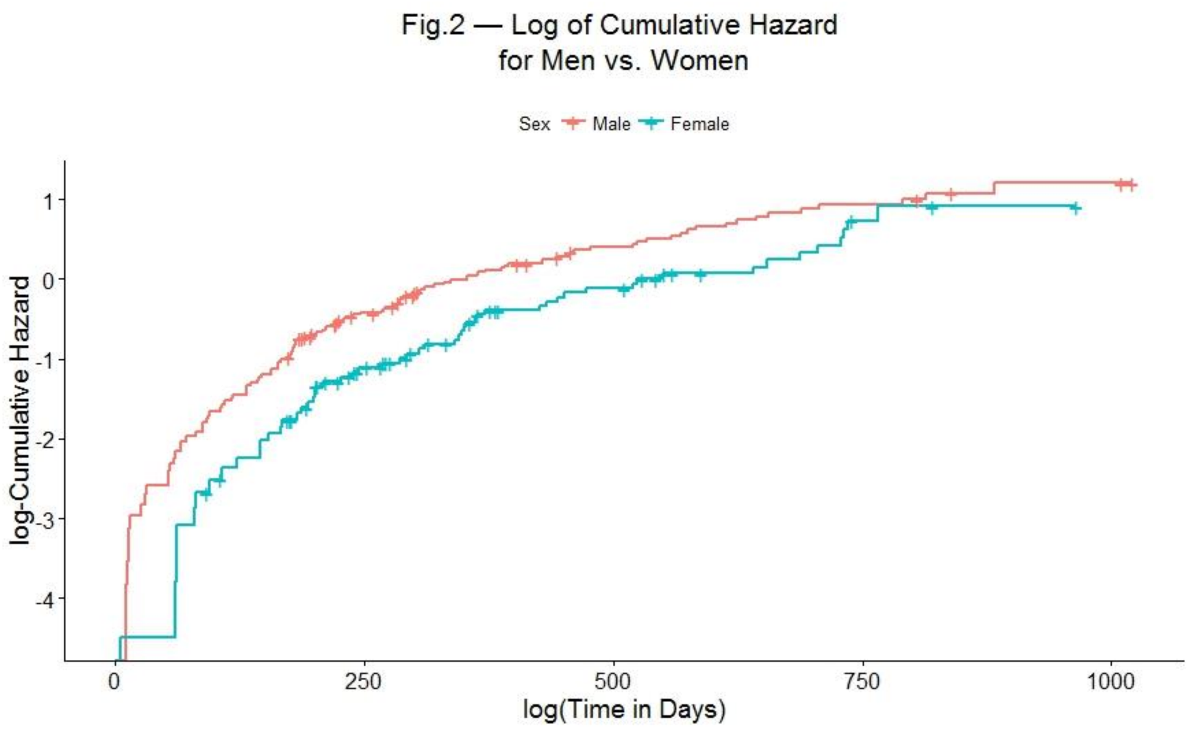
A survival plot for both men and women *(fig.1)* can be used to illustrate the survival function, i.e. the probability of surviving beyond any given point in time. This shows that basically at all times women have a better survival probability than men. (The vertical drops in the plot indicate a death and the crosses represent a censored observation.) The two lines do not cross, so the "proportional hazards" assumption is likely to hold. In other words, the rate at which deaths happen in one group is a constant proportion of the rate at which they happen in the other.

Fig.1 — Survival Function for Men vs. Women



The table underneath the survival plot provides a snapshot of the number at risk (i.e. remaining in the study) at each 250-day interval. These numbers gradually decline over time as people either pass away or are censored due to being lost to follow-up, but at around 800 days the survival functions for men and women do more or less touch each other.

The log-cumulative hazard plot *(fig.2)* shows that the two lines are more or less parallel, again suggesting that hazards are proportional across the two groups, i.e. that the proportional hazards assumption is likely to hold.



Given that these two plots both suggest that the proportional hazards function is likely to hold, we use the log-rank test rather than the Wilcoxon test to assess whether there is a significant difference in the distributions of the survival times for men and women. (The log-rank test is non-parametric and makes no assumptions about the shape of the survival curve.)

The two hypotheses are as follows:
H_0: There is no difference in survival experiences between the two genders.
H_1: There is a difference in survival experiences between the two genders.

The chi-squared statistic is 10.3 with 1 degree of freedom with a p-value of 0.0013. At the 5% significance level we therefore reject the null hypothesis that there is no difference in the survival experiences of the two genders. Men and women have different survival functions — women are expected to survive longer than men (and this applies for any given point in time).

2. Fit a Cox proportional hazards model with all covariates included (treat ECOG as continuous and apply listwise deletion for missing values). Comment on the model and assess whether we can drop the variables Karnofsky.physician and Karnofsky.patient simultaneously.

Whereas the log-rank test can only be used to explore the survival effects of one variable at a time, the purpose of the Cox proportional hazards regression model (Cox model) is to explore the effects of several covariates simultaneously. It is a form of linear regression model, which means that it assumes that basically a single line (or curve) is sufficient to estimate the survival times. The Cox model is semi-parametric — it assumes proportional hazards, meaning that it is therefore possible to estimate the effect parameters without any regard for the hazard function. However, the flexibility offered by the fact that it makes no assumptions of any particular distribution of the survival times is very desirable in survival analysis, for which one would not expect the data to follow (say) a normal distribution.

After fitting a Cox model, the likelihood ratio test p-value is very close to zero (p < 0.001) so the global test indicates that at least one variable in the model has a significant effect on survival times, although further analysis is required to determine which variable(s) is(are) important. The confidence intervals for age, pat.karno and meal.cal all contain one, suggesting that they are not useful for the model. However the other five explanatory variables may all potentially be of use.

First we examine what happens when ph.karno and pat.karno are removed simultaneously from the full Cox model. We have already seen that pat.karno does not seem to be significant, while a correlation matrix of the six continuous explanatory variables suggests that ph.karno is highly correlated with ph.ecog, with a correlation coefficient of 0.82. This is hardly surprising, since the physician's Karnofsky and ECOG performance scores are bound to be heavily related because they are two almost identical performance assessments undertaken by the physician, just with somewhat different scales (although differences will arise because physicians may not convert these scores in a common fashion; translating an ECOG score into an equivalent Karnofsky score is a subjective process).

With the pat.karno and ph.karno variables removed, the r-squared is reduced from 0.264 to 0.238. However, that in itself does not indicate whether or not the second model is actually "worse" than the first, since the r-squared for the Cox model is a generalised version that is highly sensitive to the proportion of censored observations. Ideally one would prefer a parsimonious model that relies on few variables, so the trade-off between explanatory performance and the number of variables needs to be considered.

We can use a log-likelihood ratio test to compare the full model to the reduced model. This shows that at the 5% significance level the second model (with ph.karno and pat.karno removed) is not (quite) statistically significantly worse than the full model (p-value 0.0568 for chi-squared statistic of 5.7365 with 2 degrees of freedom). The second, more parsimonious, model is therefore preferred.

3. Find the "best" Cox proportional hazards model using a variable selection procedure of your choice. Call this model bfit.

The next task is to decide which of the eight covariates should be retained in the final model, and which can be set aside. One very common means of variable selection is to use a stepwise regression process. Using the Akaike Information Criterion, stepwise regression on this dataset reduces the model down to five explanatory variables: sex, ph.ecog, ph.karno, pat.karno and wt.loss.

Here the r-squared is only 0.15, but using the log-likelihood test the difference between this model and the full model is not found to be significant at the 5% significance level (p-value of 0.1593 on a chi-squared statistic pf 23.872 with 18 degrees of freedom). Thus the five-covariate model derived from stepwise regression is preferred.

However, examination of the remaining coefficients suggests that ph.karno, pat.karno and wt.loss may not be useful (the confidence intervals for each of these contains one and the p-values are not significant at the 5% level). If we try removing those three variables the R-squared drops to 0.11, but using the log-likelihood test the new bivariate model is not found to be significantly worse than the full model at the 5% significance level (p-value of 0.0637 on a chi-squared statistic of 31.637 with 21 degrees of freedom).

No significant interactions were found between sex and ph.ecog, so no interaction component is included in the final model. Our "best" model therefore contained only two covariates, gender and the physician's ECOG score. However, it is important to try and interpret this model and also to assess the model fit.

4. Assess the goodness of fit of bfit and interpret the model.

Looking at the exponent of the coefficient, the estimated hazard ratio for women vs. men is 0.60 (confidence interval at 95% of 0.41, 0.88) with a p-value of < 0.01. Thus for women the hazard of death is 0.6 times that of a man, at any given time.

In addition, the estimated hazard ratio for ph.ecog is 1.62 (confidence interval at 95% of 1.25, 2.10) with a p-value of < 0.001. The hazard of death increases by 62% for each additional unit of ph.ecog, at any given time.

These hazard ratios are a combination of two effects:

1. The difference in the number of people who die in the two groups.

2. The difference in the survival times of people who die in the two groups.

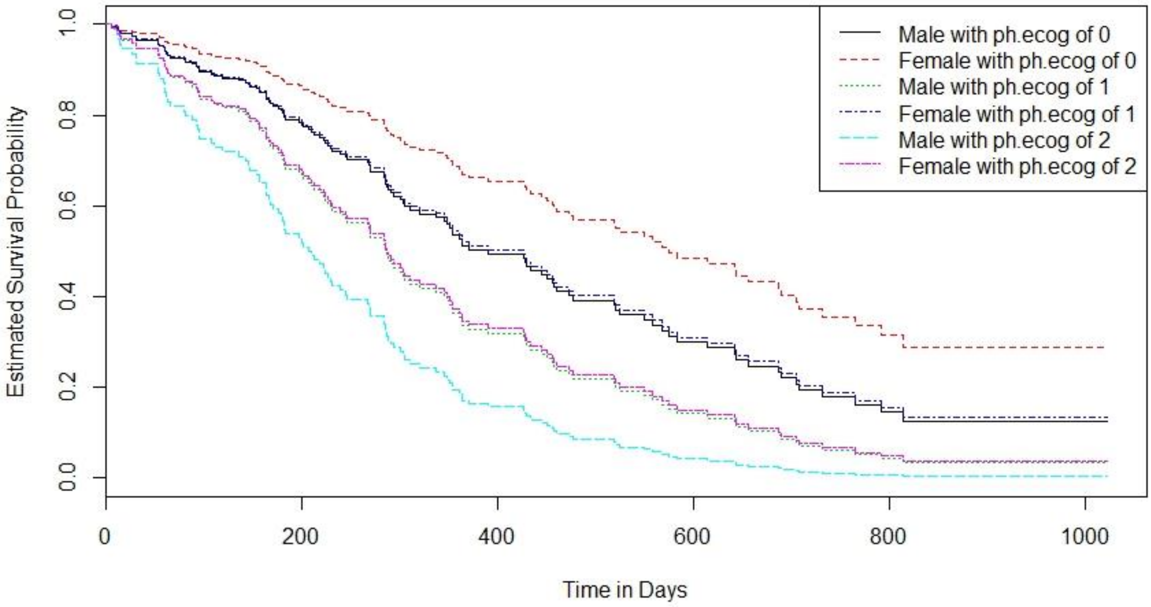
Broadly speaking, the hazard ratio is a kind of average relative risk over time. However, it can only really be used as an estimate for relative risk when the relative risk is more or less constant over time (in other words, when the proportional hazards assumption holds, as it does here).

We can also use the bfit model to make risk predictions for different survival times using the two levels of sex (male/female) and the three standard values of ph.ecog (0, 1 and 2; ph.ecog=3 is omitted here as there is only a single person who meets that criteria).

For a survival time of one year and a ph.ecog of 2, the risk of death for men is 82.8% and for women 65.3%, but for a ph.ecog of 0 those risks are reduced to 48.9% and 33.2% respectively. Naturally, all of those risks of death are much higher when looking at survival chances after two years rather than one, and after the end of the study at 1022 days the chances of survival for the ph.ecog=2 group are slim for both sexes, but women still have 28.8% survival chance if they are in the ph.ecog=0 group.

These six survival curves can also be plotted for the purposes of comparison *(fig.3)*. Interestingly, the survival curve for men with ph.ecog=0 appears to be almost identical to that of the women with ph.ecog=1, and likewise the survival curve for men with ph.ecog=1 is almost identical to that of the women with ph.ecog=2. It seem like there are effectively four distinct groups, and being male bumps a person up a risk group compared to an equivalent female, with men in ph.ecog=2 having by far the worst prognosis.

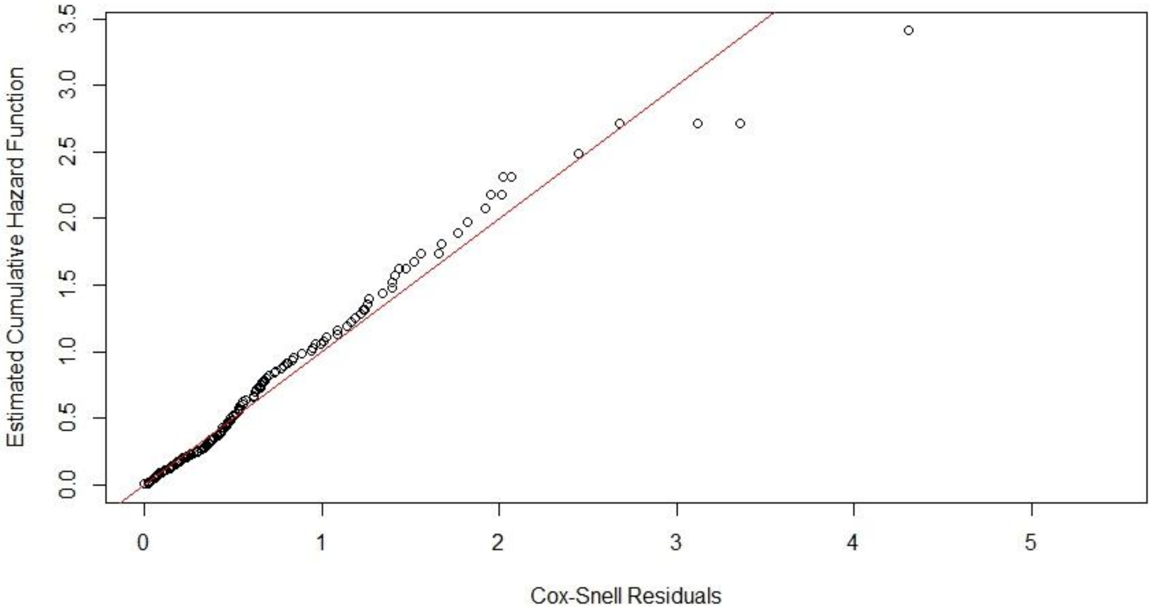
Fig.3 — Predicted Survival Rates



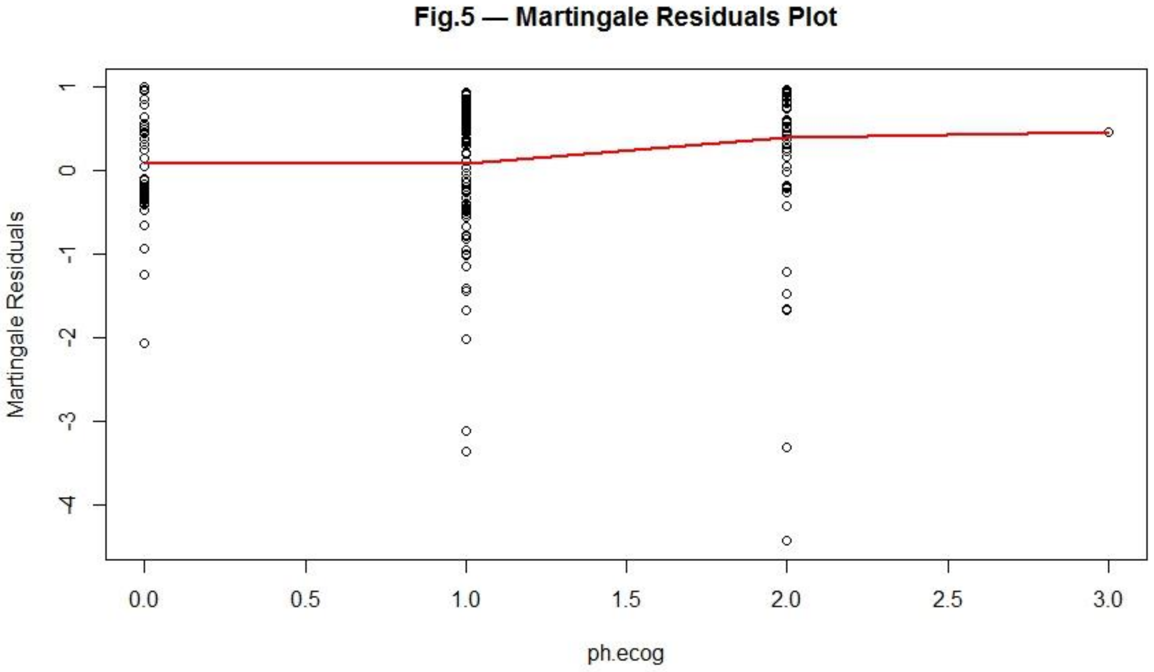
It is also important to run some diagnostics on the model to assess its "goodness of fit". How well does the model match the actual data?

The Cox-Snell residuals plot *(fig.4)* is one, rather unreliable, means of checking the fit of the model. Here, apart from a few values on the right with higher residuals, the points fall approximately on a straight line with unit slope, suggesting that the model is probably adequate.

Fig.4 — Cox-Snell Residuals Plot



The Martingale residuals plot *(fig.5)* can be used to ensure that the one variable that is being treated as continuous here, ph.ecog, is in the correct functional form (i.e. does not require transformation). The red line in this plot is approximately a straight line, suggesting that that the function form for ph.ecog is appropriate.



The score residuals plots (figs.6 & 7) assess the proportional hazards assumption. Although the sex plot seems fine, the ph.ecog has some extreme values as time increases. A statistical test of the proportional hazards assumption (using the scaled Schoenfeld residuals rather than the score residuals, which are a modification of the Schoenfeld residuals) results in a p-value of 0.018 for ph.ecog (chi-squared statistic of 5.58), indicating that at the 5% significance level we should reject the null hypothesis that the hazards are proportional over time. The variance of the score residuals for ph.ecog appears to be increasing over time, so proportional hazards cannot be assumed for that variable.

Fig.6 — Score Residuals Plot for Sex

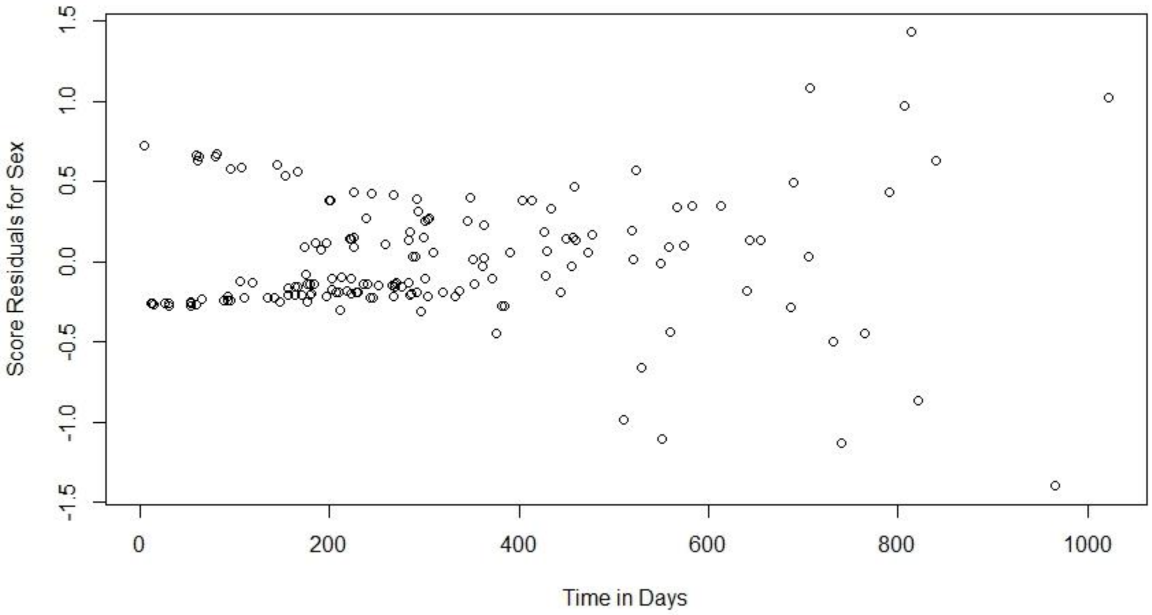
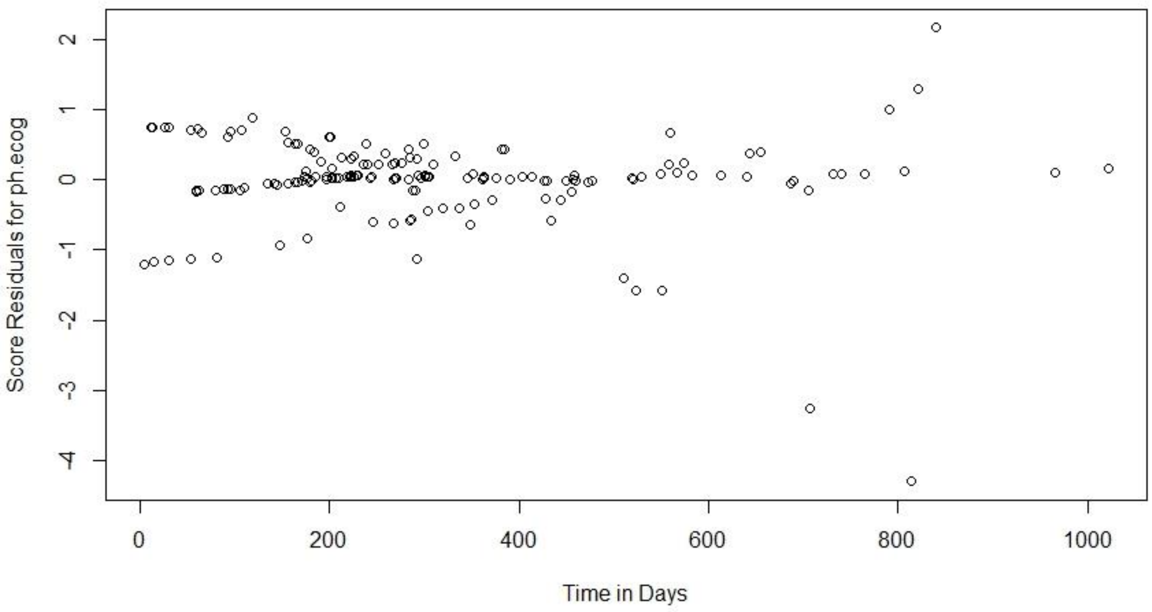


Fig.7 — Score Residuals Plot for ph.ecog



The proportional hazards assumption for ph.ecog can also be checked with survival and log-cumulative hazard plots *(figs.8 & 9).* The red and green lines cross in the survival function and also do not run parallel in the hazard function, both of which suggest that the proportional hazards assumption does not hold for ph.ecog. In other words, there appears to be an interaction between ph.ecog and time, the effect of ph.ecog on survival changes over time. One possible approach for dealing with this would be to incorporate time-dependent covariates into the model.

Fig.8 — Survival Function
for Physician's ECOG Score

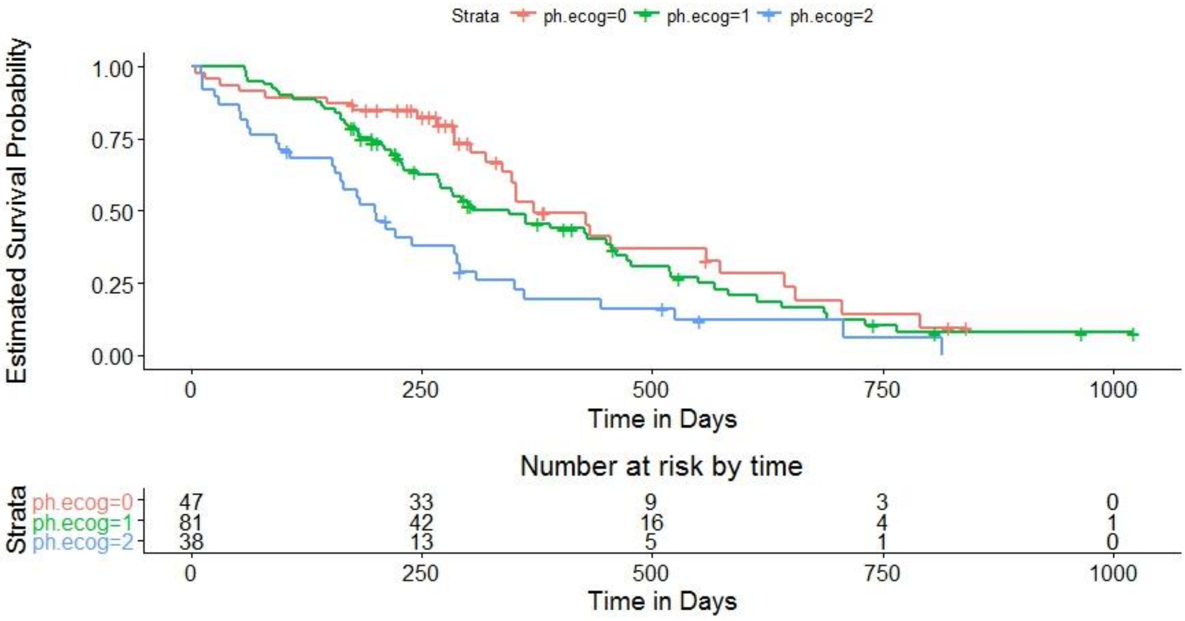
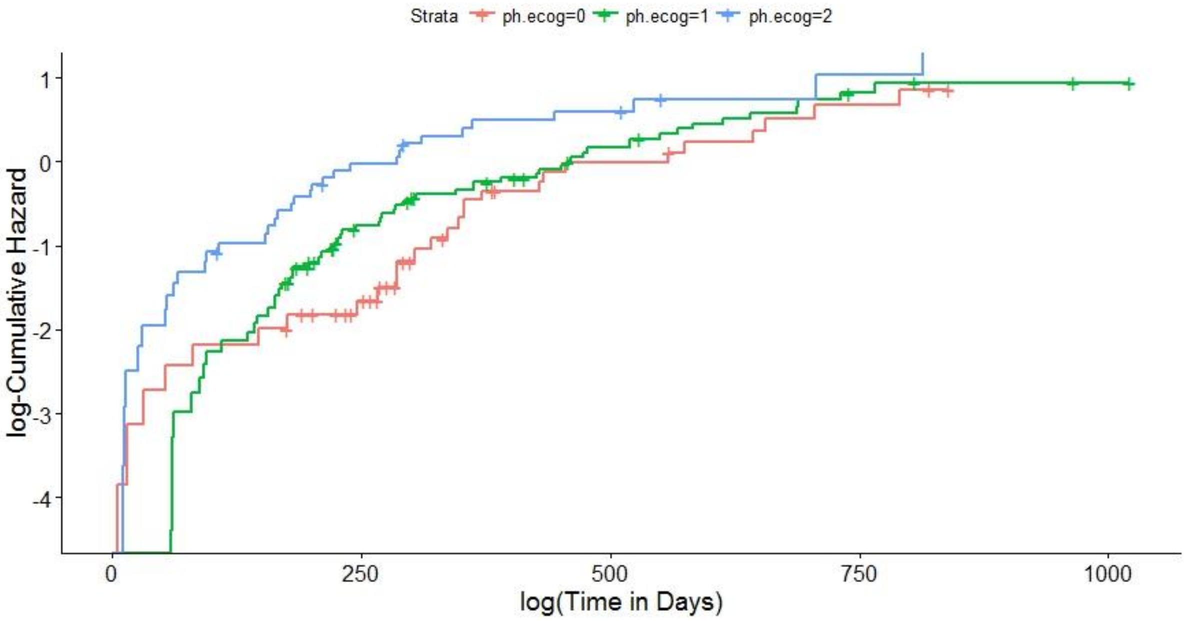
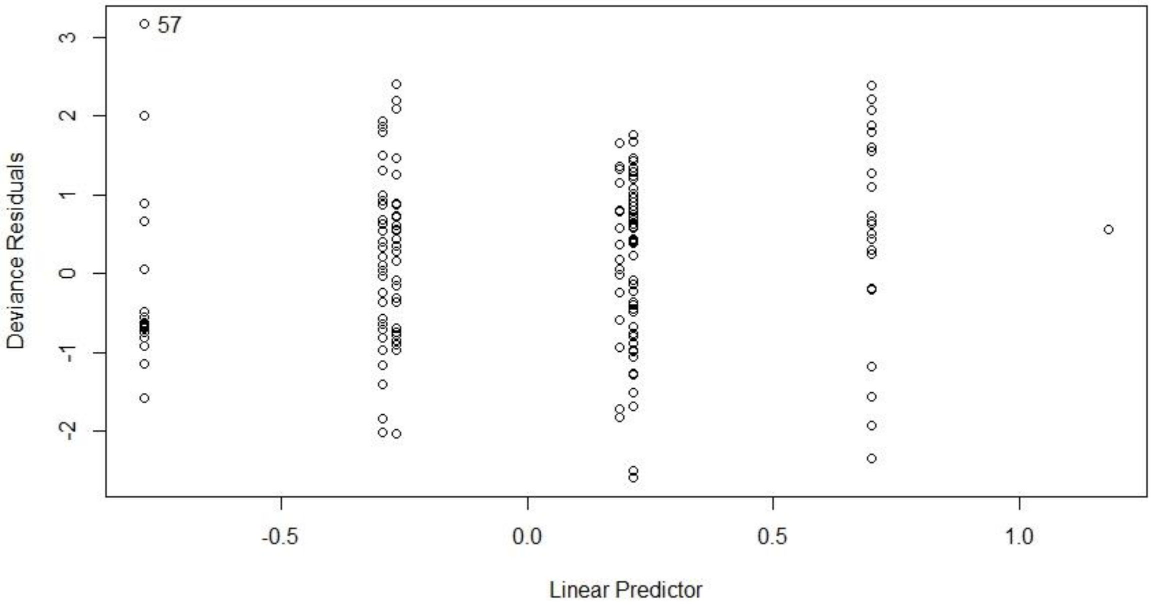


Fig.9 — Log of Cumulative Hazard
for Physician's ECOG Score



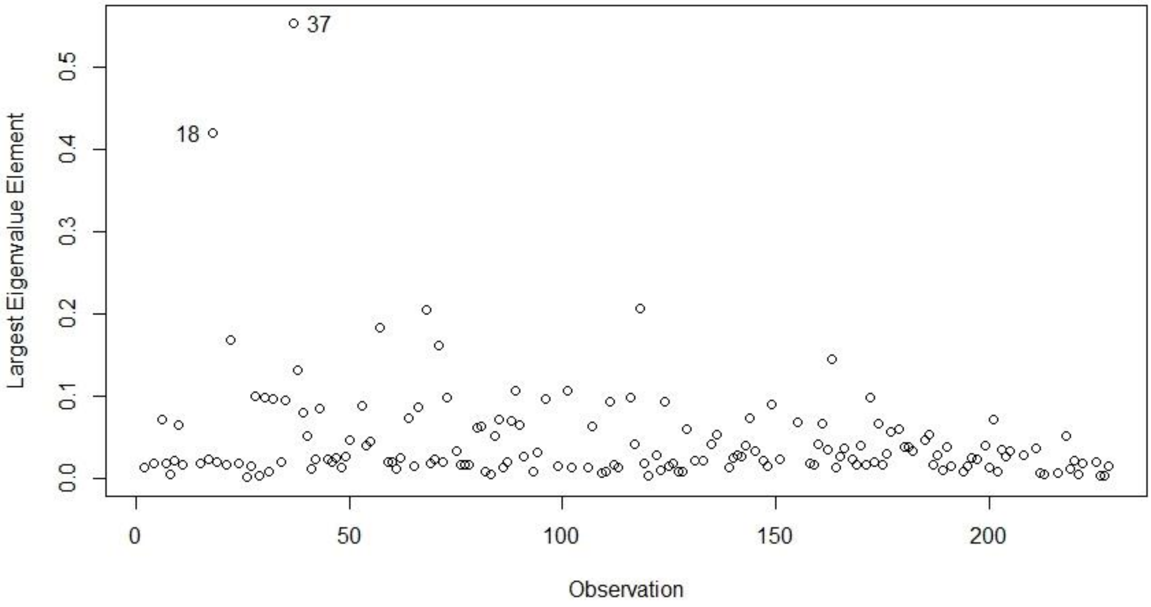
The deviance residuals plot *(fig.10)* is used to identify outliers, i.e. individuals whose survival times are not well fitted by the model. There is one outlier who has a deviance residual in excess of 3 — observation 41 in the complete cases dataset (which is equivalent to 57 in the full dataset, the identification number shown on the plot). This individual has an estimated large negative risk score (the linear predictor on the x-axis) and thus a relatively high median expected survival time (female, ph.ecog of 0, physician's Karnofsky rating at the maximum 100 etc.) of 574 days. However, they passed away in just five days.

Fig.10 — Deviance Residuals Plot



The influential observations plot *(fig.11)* highlights any observations that have a particularly large effect on the parameter estimates. Observations 11 and 25 in the complete cases dataset (which are equivalent to 18 and 37 in the full dataset) may have an undue effect on the parameter estimates.

Fig.11 — Influential Observations Plot



Both of these observations are men have a relatively high ph.ecog, a relatively low ph.karno/pat.karno, a relatively high weight loss and a long survival time. According to the bfit model, their expected median survival time is 210 days (95% confidence interval of 167, 285 days), so survival times of 707 and 814 days respectively greatly exceeded expectations. It appears that these patients may have rather "beaten the odds" with regards their expected survival duration, and their relative "success" might possibly provide some useful insights into improving survival times for other individuals.

5. Summarise your analysis based on bfit and present targeting to a non-statistical educated audience. Present your software code, output and plots in the appendix.

This project analyses lung cancer survival times based on a sample of 228 patients, 165 of whom died during the course of the study and 63 of whom have no recorded death date, either because they were lost to follow-up or survived right through to the end.

Our first task was to see whether there were any difference between the survival experiences of men and women. As shown in *fig.1*, women have a higher survival rate than men at all points in time. In other words, regardless of how long ago they joined the study, a woman has a comparatively better life expectancy than an equivalent man who joined the study on the same date. On average (using the median, i.e. the midpoint of all the observations when they are ordered by survival times) women were expected to live for at least 426 days compared to just 270 for men. The survival rate after one year is 33.6% for men and 52.6% for women, while after two years their respective rates are 7.8% and 18.7%. Using a standard statistical test, we were able to conclude that there is a statistically significant difference in the survival experiences between the two genders.

In all, there were data for eight “explanatory” variables, i.e. potential contributors to variations in survival times. Unfortunately, some records had missing data for one or more variables, so these were removed, leaving a final sample of 167 patients.

Our statistical modelling approach (known as the Cox proportional hazards model) whittled the explanatory variables down from eight to just two key ones: sex, and the physician’s ECOG score (which ranges from zero to five, with zero being the most positive assessment of a patient’s current situation, indicating that the person is able to carry out a basically normal, fully active lifestyle). Women with a low ECOG score had the best prognosis and men with a high ECOG score had the worst. There were strong similarities between the survival experiences of women with an ECOG score of 0 and men with an ECOG score of 1, and likewise between women with an ECOG score of 1 and men with an ECOG score of 2.

Interestingly, there were a pair of men who greatly exceeded their predicted survival times. Whereas they each would have been anticipated to survive 210 days according to the model, they actually lived for 707 and 814 days respectively. In addition, there was also a woman who was expected to survive at least 574 days, but died five days into the study. It is possible that a close examination of these three unusual cases might possibly shed some light on how survival rates might be improved in future.

However, one of the main assumptions for applying the Cox model was violated (known as “proportional hazards”). A physician’s ECOG score of 2 would seem to suggest worse outcomes *at all times* than an ECOG score of 1. Counter-intuitively, however, it appears that for the first 150 or so days, those with a physician’s ECOG score of 2 survived at a better rate than those with a score of 1. It may be worth closely examining the deaths for these two groups during the first six months to see whether any insights can be gained into this discrepancy.