Stroke analysis

Karim Saied (karim.saied@unil.ch)

22/12/2017

1. Aim

The present analysis aims to assess the effect of some covariables on patients that have suffered from stroke.

2. Data description

The dataset used (from package ISwR) contains informations about all cases of stroke during the period 1991-1993 in Tartu, Estonia. The follow-up of the study was completed in January 1996.

The dataset contains 829 observations and 10 covariables.

- `sex`: two-levels factor (`Female` and `Male`)
- `died`: date of death (form: YYYY-MM-DD)
- `dstr`: date of stroke (form: YYYY-MM-DD)
- `age`: age at stroke (integer)
- `dgn`: diagnosis (four-levels factor, `ICH`: intracranial haemorrhage, `ID`:unidentified, `INF`:infarction, ischaemic, `SAH`:subarchnoid haemorrhage)
- `coma`: whether the patient was in coma after the stroke (two-levels factor, `Yes` or `No`)
- `diab`: whether the patient suffered from diabetes (two-levels factor,`Yes` or `No`)
- `minf`: whether the patient suffered from myocardial infarction (two-levels factor, `Yes` or `No`)
- `han`: whether the patient suffered from hypertension (two-levels factor, `Yes` or `No`)
- `dead`: date of death during the study
- `obsmonths`: observation time in months

First of all, one has to note that the dataset is biased. As a matter of fact, 62% of the cases were femals whereas only 38% were males.

3. Data visualisation

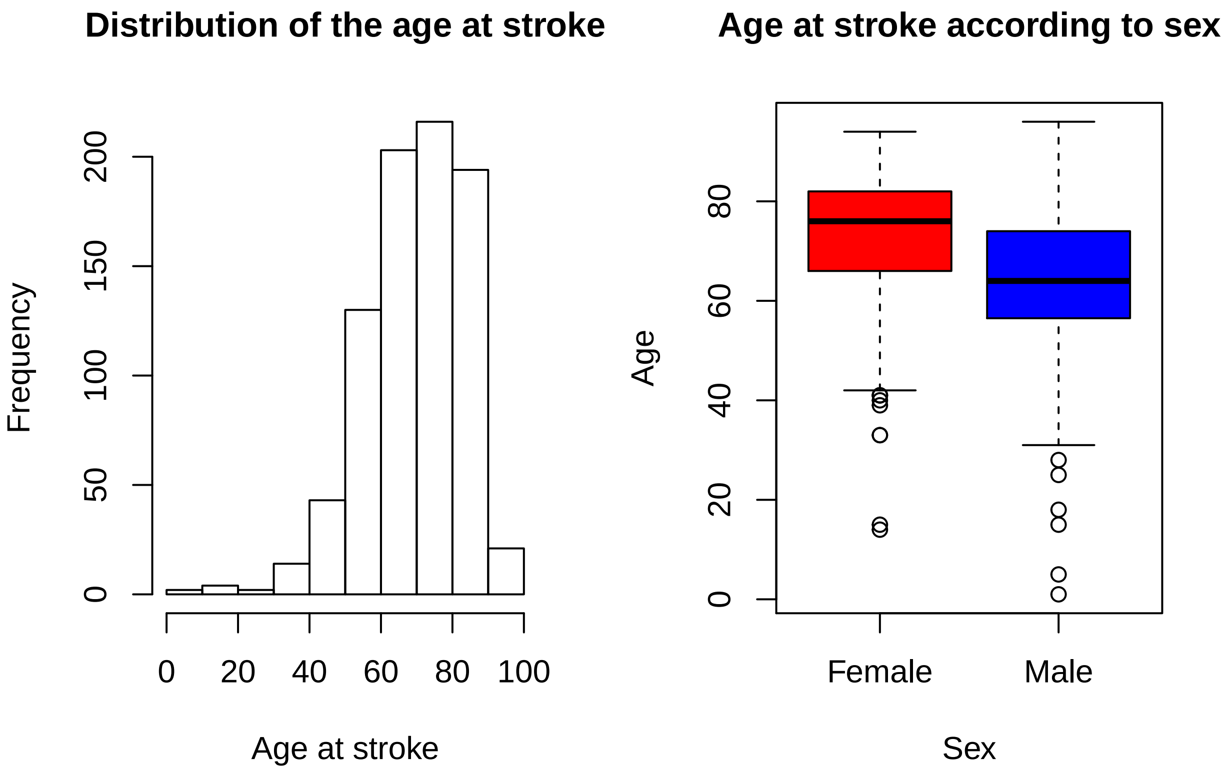

Figure 1. Left panel: distribution of the age at stroke (groups of ten years). Right panel: distribution of the age at stroke according to sex.

One can see that stroke is prevalent for individuals in the 60 to 90 age group (Figure 1, left panel). Moreover, females suffered from stroke at a later age compared to males (right panel). One has to note that data are right-censored. As a matter of fact, some subjects were lost to follow-up either because they left the ongoing study or they were still alive after the end of the follow-up. In both cases, we do not know what happened to the patients after that. Such censoring bias makes linear regression not suitable for the current analysis.

4. Pre-analysis

```
## Call: survfit(formula = stroke_surv ~ 1)
##
##        n  events  median 0.95LCL 0.95UCL
##    829.0   485.0    20.0    13.4    26.6
```

Among 829 patients, 485 of them died during the follow-up (uncensored observations).

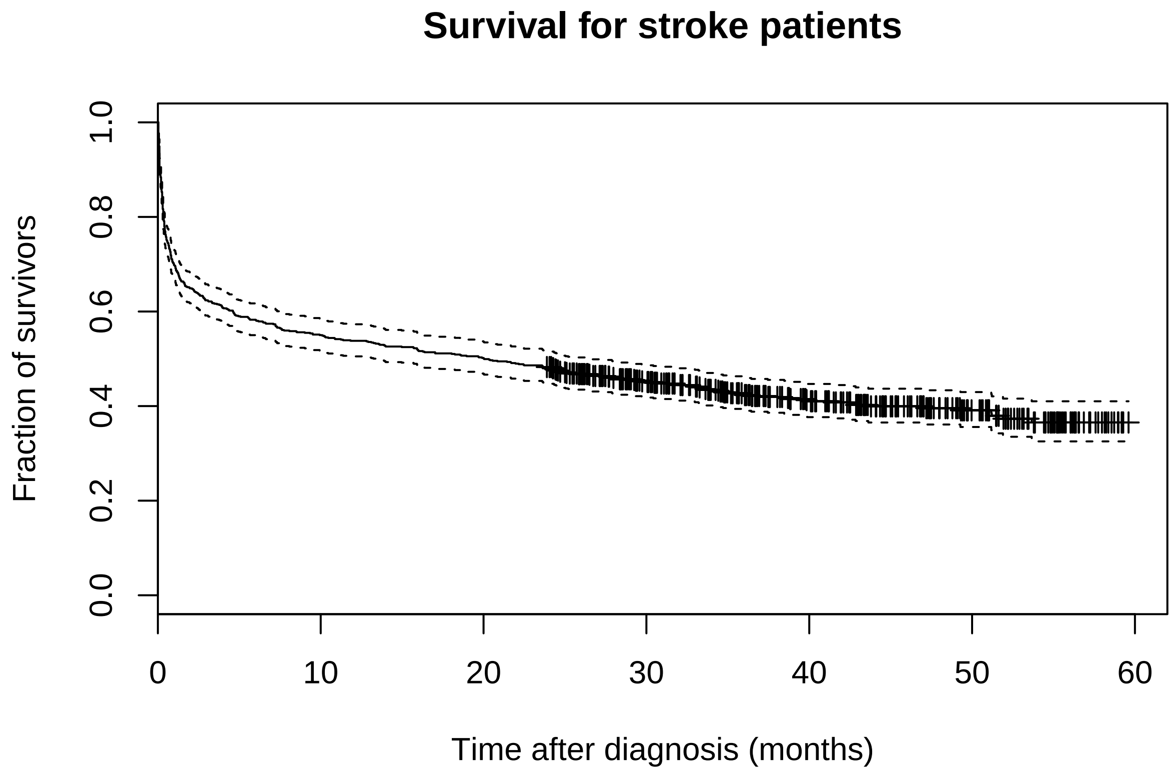

Figure 2. Distribution of the survival according to months after diagnosis. The solid line represents the fraction of survivors at a certain time point. Dashed lines differentiate the 95% confidence interval.

The Kaplan-Meier graph displays the proportion of patients that survived up to a particular time. One can observe that the survival strongly decrease within the first months after diagnosis. After several months, the survival still decrease but at a lower rate. The crosses represent the patients that were lost to follow-up. The population can be splitted according to different variables. First of all, the variable sex is taken into account.

4.1. Survival according to sex

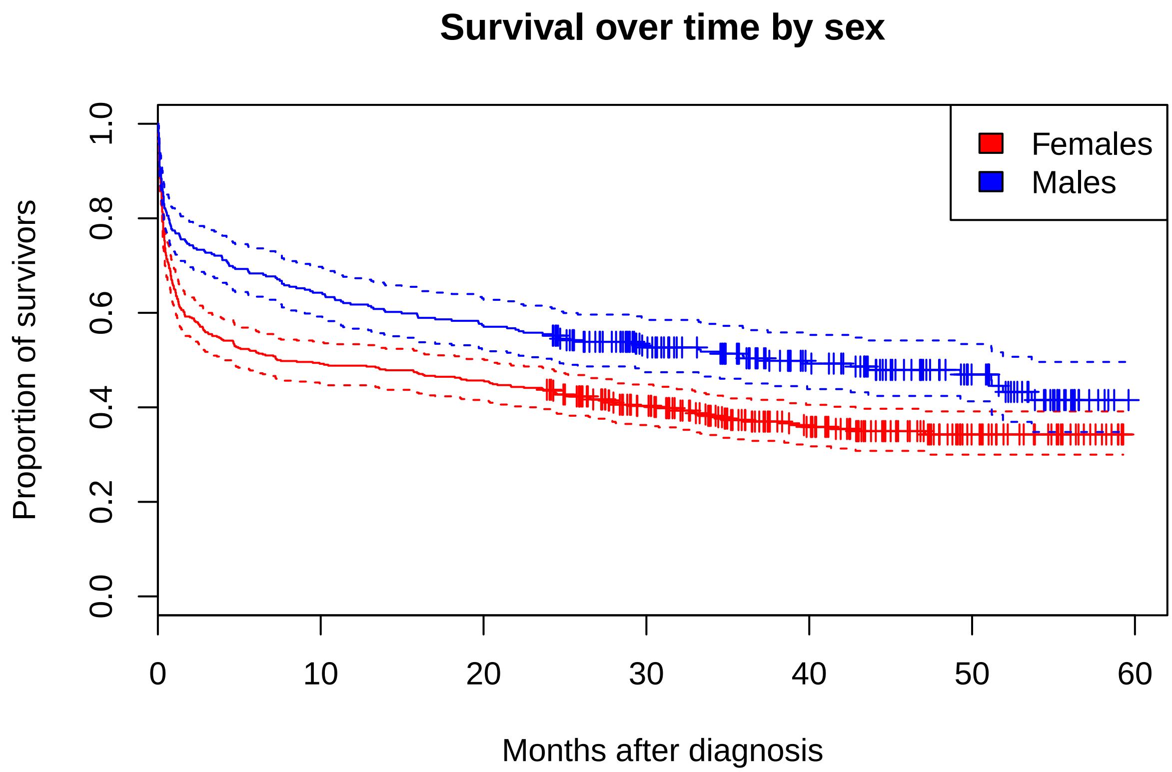

Figure 3. Distribution of the survival according to months after diagnosis for males and females separately. The solid line represents the fraction of survivors at a certain time point for males and females. Dashed lines differentiate the 95% confidence interval.

One can observe that the survival of females decreases more rapidly. Such difference is evaluated statistically using a log-rank test:

```
## Call:
## survdiff(formula = stroke_surv ~ stroke$sex)
##
##                     N Observed Expected (O-E)^2/E (O-E)^2/V
## stroke$sex=Female 510      321      280      6.11      14.6
## stroke$sex=Male   319      164      205      8.32      14.6
##
##  Chisq= 14.6  on 1 degrees of freedom, p= 0.000132
```

The log-rank test shows that, following a stroke episode, males have a significantly higher survival compared to females (p-value < 0.001). Among 510 females, 321 died whereas only 164 males died among 319. Then, the survival between females and males was evaluated after adjusting for age. For that purpose, a Cox Proportional-Hazards model is used:

```
## Call:
## coxph(formula = stroke_surv ~ stroke$sex + stroke$age)
##
##                   coef exp(coef) se(coef)     z      p
## stroke$sexMale 0.02233   1.02258  0.10072  0.22   0.82
## stroke$age     0.04929   1.05052  0.00437 11.27 <2e-16
##
## Likelihood ratio test=159  on 2 df, p=0
```

```
## n= 829, number of events= 485
```

The results show that the age and the risk of death are significantly related (p-value < 0.001) as shown by the estimated hazard ratio (exp(coef) = 1.05052). As the ratio is greater than 1, the relation between the risk of death and the age is positive, meaning that suffering from stroke at a later age increases the risk of death. Each year increases the risk of death by 5.1%.

Interestingly, the sex is not significantly related to the risk of death (p-value = 0.82), contrary to what has been observe by means of the log-rank test. Knowing that females tend to suffer from stroke at a later age compared to males (Figure 1, right panel), what was observed on Figure 3 and with the log-rank test was in fact the effect of the age. As a matter of fact, the higher risk of death (i.e. the lower survival) in females is related to their age.

4.2. Survival according to the other factors

As the study aims to assess the impact of covariables on the survival of patients, first, the differents factors have been tested separately using a Cox Proportional-Hazards model. It appeared that all factors except diabetes and history of hypertension are somehow related to survival (positively or negatively, depending whether it increases or decreases the survival).

Depending on the diagnosis, the risk of death is significantly reduced. Indeed, the risk is significantly decreased by 52.01% (p-value = 9.6e-07) and 53.64% (p-value = 0.0028) for `INF` and `SAH` diagnosis respectively. However, experiencing coma upon stroke and having history of myocardial infraction, increases the risk of death by 1184.6% and 55.2% respectively.

5. Analysis

Once all the factors have been tested one by one, only those that displayed significant results are implemented in the final model. It should be noted that although the history of hypertension have not shown a significant effect (p-value = 0.063), the factor is still taken into account as the p-value is relatively close to 0.05. The final analysis is performed using a Cox Proportional-Hazards model.

```
## Call:
## coxph(formula = stroke_surv ~ stroke$sex + stroke$age + stroke$dgn +
##     stroke$coma + stroke$minf + stroke$han)
##
##                     coef exp(coef) se(coef)     z       p
## stroke$sexMale  0.00917   1.00921  0.10383  0.09 0.92965
## stroke$age      0.04872   1.04993  0.00475 10.25 < 2e-16
## stroke$dgnID   -0.26393   0.76803  0.16969 -1.56 0.11987
## stroke$dgnINF  -0.62514   0.53518  0.15840 -3.95 7.9e-05
## stroke$dgnSAH  -0.43181   0.64933  0.25872 -1.67 0.09512
## stroke$comaYes  2.49644  12.13917  0.14973 16.67 < 2e-16
## stroke$minfYes  0.47692   1.61110  0.12953  3.68 0.00023
## stroke$hanYes  -0.03555   0.96508  0.09475 -0.38 0.70751
##
## Likelihood ratio test=413  on 8 df, p=0
## n= 816, number of events= 473
##    (13 observations deleted due to missingness)
```

The final model exposes consistent results with what has been previously observed. Age at stroke, experiencing coma and history of myocardial infraction significantly increase the risk of death. The fact that stroke is prevalent at a later age is quite intuitive as the overall health of an individual tend to decline with time. The variable that has the larger impact is coma (exp(coef) = 12.13917), which indicates that falling into coma after stroke, strongly decreases the survival (p-value = < 2e-16). Patients with history of myocardial infraction have a significant higher risk of death as well (p-value = 0.00023). In contrary, the risk of death is significantly decreased by 46.5% in patients with `INF` diagnosis (p-value = 7.9e-05).

6. Conclusion

The results from the final model highlighted the fact that some factors tend to decrease the survival while, surprisingly, others can increase it. The condition `INF` (infarction, ischaemic) was the only one that increases the survival in the current study. In contrary, the age, falling into coma and having history of myocardial infraction decrease the chance of survival after having suffered from stroke.