Time to freezing of gait analysis : A time to event analysis

Contents

Executive summary

In this project, we will be using data from kaggle.We aim at analyzing the expected duration of time until freezing of gait (FOG) occurs (time to FOG). FOG is a pattern occurring in patient with Parkinson disease. It is a debilitating symptom that afflicts many people with Parkinson's disease. Some events like start walking hesitation, turning body could be observed and help to detect FOG occurrence.

Introduction

We aim at analyzing the time to FOG in 44 Parkinson disease patients, with median age of 69 years old. Here FOG is the event of interest and it is expected within the execution of a given task in lab, or during daily life activity.We will analyse data from lab tasks. Our main goal is to analyse and explain overall FOG time to event model, the influence/importance of some features in the time to occurrence of FOG from the beginning to the end of the task. To achieve this, we model our data under kaplan-Meier model.

Data preparation, inspection and preprocessing.

let's join all metadata tables before diving into analysis.

- defog_metadata.csv Identifies each series in the tdcsfog dataset by a unique Subject, Visit, Test, Medication condition.

  - `Visit` Lab visits consist of a baseline assessment, two post-treatment assessments for different treatment stages, and one follow-up assessment.
  - `Test` Which of three test types was performed, with 3 the most challenging.
  - `Medication` Subjects may have been either off or on anti-parkinsonian medication during the recording.

- subjects.csv Metadata for each Subject in the study, including their `Age` and `Sex` as well as:

  - `Visit` Only available for subjects in the daily and defog datasets.
  - `YearsSinceDx` Years since Parkinson's diagnosis.
  - `UPDRSIIIOn/UPDRSIIIOff` Unified Parkinson's Disease Rating Scale score during on/off medication respectively. `NFOGQ` Self-report FoG questionnaire score. See: https://pubmed.ncbi.nlm.nih.gov/19660949/

- events.csv Metadata for each FoG event in all data series. The event times agree with the labels in the data series.

  - `Id` The data series the event occured in.
  - `Init` Time (s) the event began.
  - `Completion` Time (s) the event ended.
  - `Type` Whether StartHesitation, Turn, or Walking.
  - `Kinetic` Whether the event was kinetic (1) and involved movement, or akinetic (0) and static.

- tasks.csv Task metadata for series in the defog dataset. (Not relevant for the series in the tdcsfog or daily datasets.)

  - `Id` The data series where the task was measured.
  - `Begin` Time (s) the task began.
  - `End` Time (s) the task ended.
    Task One of seven tasks types in the DeFOG protocol, described on this page.

Remove **Visit** from **subjects** table , as we do not use it

Note: we are going to consider only defog condition subjects.

Data structure

Ad event and tasks duration columns features:

- `eventsDuration <- Completion - Init`
- `tasksDuration <- Begin - end`

Data analysis

Subjects

How many unique subjects are in this dataset ?

There are 44 distinct subjects in this study.

Kinetic / Events

How many trials has missing kinetic/status ?

```
## [1] 0.48
```

48% of trials has missing Kinetic(status).

filter for missing kinetic entries/cases.

```
##
##  new table dimension

## [1] 2232    19

## # A tibble: 3 x 19
##   Id           Init Compl~1 Type  Kinetic Begin   End Task  Subject Visit Medic~2
##   <chr>       <dbl>   <dbl> <chr>   <dbl> <dbl> <dbl> <chr> <chr>   <dbl> <chr>
## 1 02ea782681 1377.   1378. Turn        1 1371. 1393. Turn~ ae2d35      2 on
## 2 02ea782681 1377.   1378. Turn        1 1371. 1393. Turn~ ae2d35      2 on
## 3 02ea782681 1466.   1467. Turn        1 1461. 1472. Hots~ ae2d35      2 on
## # ... with 8 more variables: Age <dbl>, Sex <chr>, YearsSinceDx <dbl>,
## #   UPDRSIII_On <dbl>, UPDRSIII_Off <dbl>, NFOGQ <dbl>, eventsDuration <dbl>,
## #   tasksDuration <dbl>, and abbreviated variable names 1: Completion,
## #   2: Medication
```

What is the proportion of observed Kinetic events ?

69 % of events has been observed.

Note : Given that each event is indicative of FOG, we will just gather them and consider that they form one class (Kinetic / events).

Age

What is the median age ?
Age numerical summary.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   28.00   61.00   69.00   67.44   73.00   82.00
```

At least half of subjects are 69 years old or above.

Sex

Sex numerical summary

```
## Sex
##    F    M
## 0.45 0.55
```

There is almost 55.1498127% of men in this cohort.

Tasks

What kind of tasks has been performed ?

```
##  [1] "TUG-DT"     "TUG-C"       "Turning-DT" "Hotspot2"    "Hotspot2-C"
##  [6] "Hotspot1"   "Turning-ST" "4MW"         "4MW-C"       "TUG-ST"
## [11] "Turning-C"  "Hotspot1-C" "MB10"        "MB11"        "MB13"
## [16] "MB12"
```

how many tasks had been performed ?

There had been 16 task performed.

Visit

How many rounds of visit did the patient had ?

```
## [1] 2
```

What is the proportion of patient within each number of visit round ?

```
## Visit
##    1    2
## 0.37 0.63
```

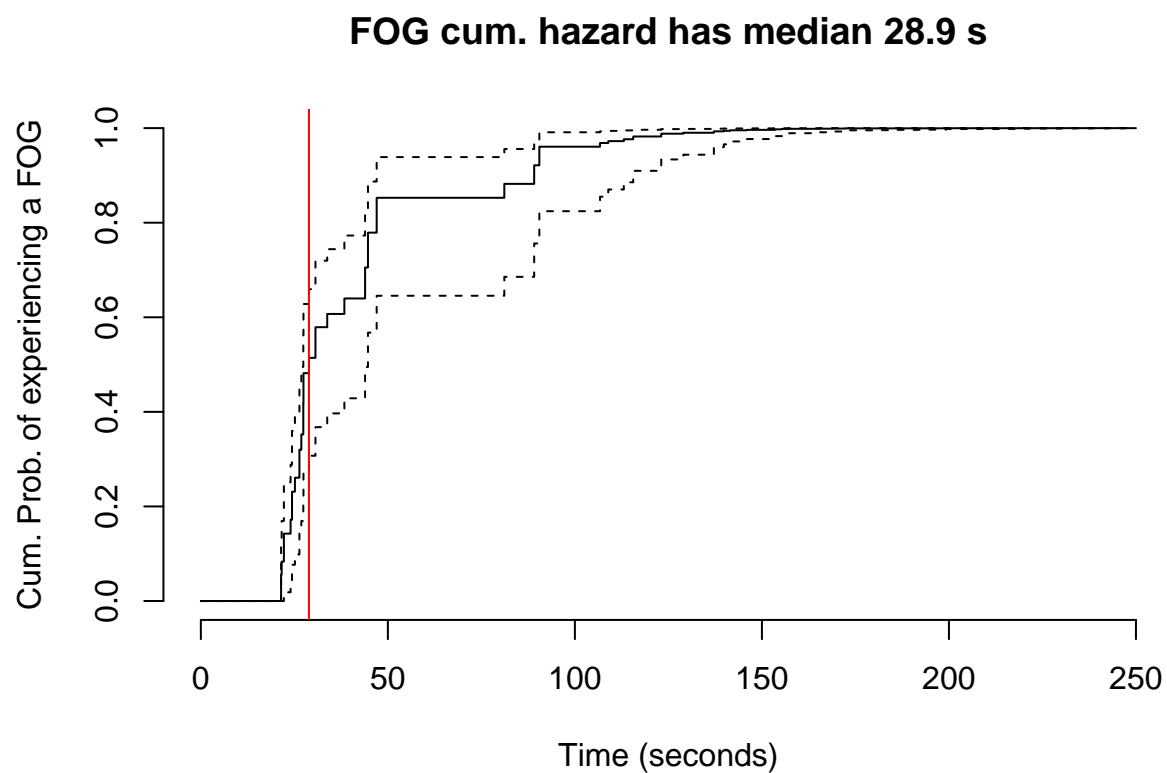37% of patients had one Visit, while the remaining had 2.

Medication

How many subjects are under parkinsonian-medication ?

```
## Medication
##   off    on
## 68.89 31.11
```

31.11% are under medication.

Modelling and Analysis : overall time to FOG.

FOG cum. hazard has median 28.9 s


```
## Call: survfit(formula = Surv(Begin, End, Kinetic) ~ 1, data = fog)
##
##        records n.max n.start events median 0.95LCL 0.95UCL
## [1,]     2232   190      36   1530   28.9    27.4    44.7
```

When considering time to the end of the task, half of patients experience FOG within 28.911 s.

Univariate analysis : Survival difference in group

Sex

```
## Call: survfit(formula = Surv(Begin, End, Kinetic) ~ factor(Sex), data = fog)
```

```
## 
##              records n.max n.start events median 0.95LCL 0.95UCL
## factor(Sex)=F     786    86      17     597   43.9    38.4      NA
## factor(Sex)=M    1446   127      19     933   27.4    24.4      NA
```

p-value is 0.15 .There is no effect of sex on time to FOG.

Medication

```
## Call: survfit(formula = Surv(Begin, End, Kinetic) ~ factor(Medication),
##     data = fog)
## 
##                          records n.max n.start events median 0.95LCL 0.95UCL
## factor(Medication)=off      1399   180      29     960   28.9    27.4    44.7
## factor(Medication)=on        833   117       7     570   22.2    21.5      NA
```

P-value is 0 which is less than 0.05. There is huge effect of Medication in experiencing FOG holding other variables fixed.

Task type

```
## Call: survfit(formula = Surv(Begin, End, Kinetic) ~ factor(Task), data = fog)
## 
##                        records n.max n.start events  median 0.95LCL 0.95UCL
## factor(Task)=4MW            42    34      34     29    30.6    27.4    44.7
## factor(Task)=4MW-C          12     6       4     12    89.8    89.1      NA
## factor(Task)=Hotspot1       96    24       2     83   592.9      NA      NA
## factor(Task)=Hotspot1-C     22     8       2     19   919.0      NA      NA
## factor(Task)=Hotspot2      329    85       8    253   695.8   695.8      NA
## factor(Task)=Hotspot2-C     54    20      20     43  1276.3  1229.7  1554.3
## factor(Task)=MB10            4     4       4      4   530.6      NA      NA
## factor(Task)=MB12            7     2       1      5    81.1      NA      NA
## factor(Task)=TUG-C          20     4       2     17   223.3      NA      NA
## factor(Task)=TUG-DT        188    23       2    145   106.7      NA      NA
## factor(Task)=TUG-ST        112    14       2     99    24.4      NA      NA
## factor(Task)=Turning-DT    715    96       1    380   237.7      NA      NA
## factor(Task)=Turning-ST    631   105      21    441   357.6   357.6   371.4
```

```
## Warning in agreg.fit(X, Y, istrat, offset, init, control, weights = weights, :
## Loglik converged before variable 6 ; beta may be infinite.
```

```
## # A tibble: 3 x 3
##   term                 estimate p.value
##   <chr>                   <dbl>   <dbl>
## 1 factor(Task)Hotspot1     1.03  0.0127
## 2 factor(Task)TUG-C        1.08  0.0186
## 3 factor(Task)Turning-DT  -1.24  0.00165
```

Three levels have significant effect on time to FOG,holding other variables fixed.

events Type

```
## Call: survfit(formula = Surv(Begin, End, Kinetic) ~ factor(Type), data = fog)
##
##                               records n.max n.start events median 0.95LCL
## factor(Type)=StartHesitation      11     4       2      9  199.2      NA
## factor(Type)=Turn               1787   178      24   1195   27.4    26.4
## factor(Type)=Walking             434    59      12    326   43.9    27.4
##                               0.95UCL
## factor(Type)=StartHesitation      NA
## factor(Type)=Turn               44.7
## factor(Type)=Walking              NA
```

```
## # A tibble: 2 x 3
##   term                estimate p.value
##   <chr>                  <dbl>   <dbl>
## 1 factor(Type)Turn      -0.675  0.0613
## 2 factor(Type)Walking   -0.285  0.435
```

All p.values are greater than 0.05. There is no effect of Task type on FOG occurrence.

visit

```
## Call:
## coxph(formula = Surv(Begin, End, Kinetic) ~ factor(Visit), data = fog)
##
##                   coef exp(coef) se(coef)     z        p
## factor(Visit)2 0.25417   1.28939  0.06223 4.085 4.42e-05
##
## Likelihood ratio test=16.35  on 1 df, p=5.273e-05
## n= 2232, number of events= 1530
```

```
## # A tibble: 1 x 3
##   term            estimate   p.value
##   <chr>              <dbl>     <dbl>
## 1 factor(Visit)2     0.254 0.0000442
```

p-value less than 0.05. Having more than one (1) round visit to hospital, has an influence on FOG occurrence.

yearSinceDx : year since diagnosis.

```
## Call:
## coxph(formula = Surv(Begin, End, Kinetic) ~ YearsSinceDx, data = fog)
##
##                  coef exp(coef) se(coef)     z    p
## YearsSinceDx 0.002764  1.002767 0.004199 0.658 0.51
##
## Likelihood ratio test=0.43  on 1 df, p=0.5113
## n= 2232, number of events= 1530
```

p-value is 0.5104841 which is > 0.05. There is no effect of Year since diagnosis on FOG occurrence.

Multivariate analysis

```
## # A tibble: 2 x 5
##   term          estimate std.error statistic  p.value
##   <chr>            <dbl>     <dbl>     <dbl>    <dbl>
## 1 Medicationon     0.740    0.0702     10.5  5.84e-26
## 2 Visit            0.240    0.0630      3.81 1.36e- 4
```

Medication and Visit present a p value < 0.05. Medication increases the risk of experiencing FOG by 0.7 holding visit fixed, while , visit does the same with a by a rate of 0.2 while holding Medication fixed.

Results

This cohort has 44 subjects, half of the subjects are at least 69 years old. 31.11 % are under medication, 63% present more than 1 round visit to hospital. Our analysis shows that at least half of the subjects time to FOG is 28.911 seconds. Subjects under medication and having frequent visit to hospital are more likely to experience a FOG episode during a motor/kinetic task like gait.

Conclusion

We aim at analyzing time to FOG in this study. Our analysis shows that FOG occurrence is related to Medication and Visit to hospital. Overall, half of subjects experience FOG within 28.9 seconds.It could be interesting to investigate also survival difference given kinetic outcomes and tasks.

References

Cox, David R. 1972. "Regression Models and Life-Tables." Journal of the Royal Statistical Society: Series B (Methodological) 34 (2): 187–202.

Kaplan, Edward L, and Paul Meier. 1958. "Nonparametric Estimation from Incomplete Observations." Journal of the American Statistical Association 53 (282): 457–81.

To be continued !!!!