

# Freezing of gait survival analysis

## Contents

<b>Executive summary</b>	<b>1</b>
<b>Introduction</b>	<b>2</b>
<b>Data preparation, inspection and preprocessing.</b>	<b>2</b>
Data structure . . . . .	2
<b>Data analysis</b>	<b>3</b>
Subjects . . . . .	3
Kinetic / Events . . . . .	3
Age . . . . .	3
Sex . . . . .	3
Tasks . . . . .	4
Visit . . . . .	4
Medication . . . . .	4
<b>Modelling and Analysis : overall time to FOG.</b>	<b>5</b>
Univariate analysis : Survival difference in group . . . . .	5
Multivariate analysis : regression . . . . .	8
<b>Results</b>	<b>9</b>
<b>Conclusion</b>	<b>9</b>
<b>References</b>	<b>9</b>

## Executive summary

In this project, we used 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. a multivariate model, gives a better understanding of time to FOG in this cohort.

## 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, investigate the influence of some features in time to 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. We will consider only data from defog metadata.

- `defog_metadata.csv` Identifies each series in the defog 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.
  - **UPDRSIIIn/UPDRSIIOff** 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 occurred 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.

we removed **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:

- `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
```

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

What is the proportion by sex ?

```
## Sex
```

```
## Female  Male
```

```
##   0.45   0.55
```

There is almost 0.55 of men in this cohort.

## Tasks

What kind of tasks has been performed ?

```
## There is 7 different tasks with 16 levels
```

## 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
```

0.37 of total patients had one Visit, while the remaining had 2.

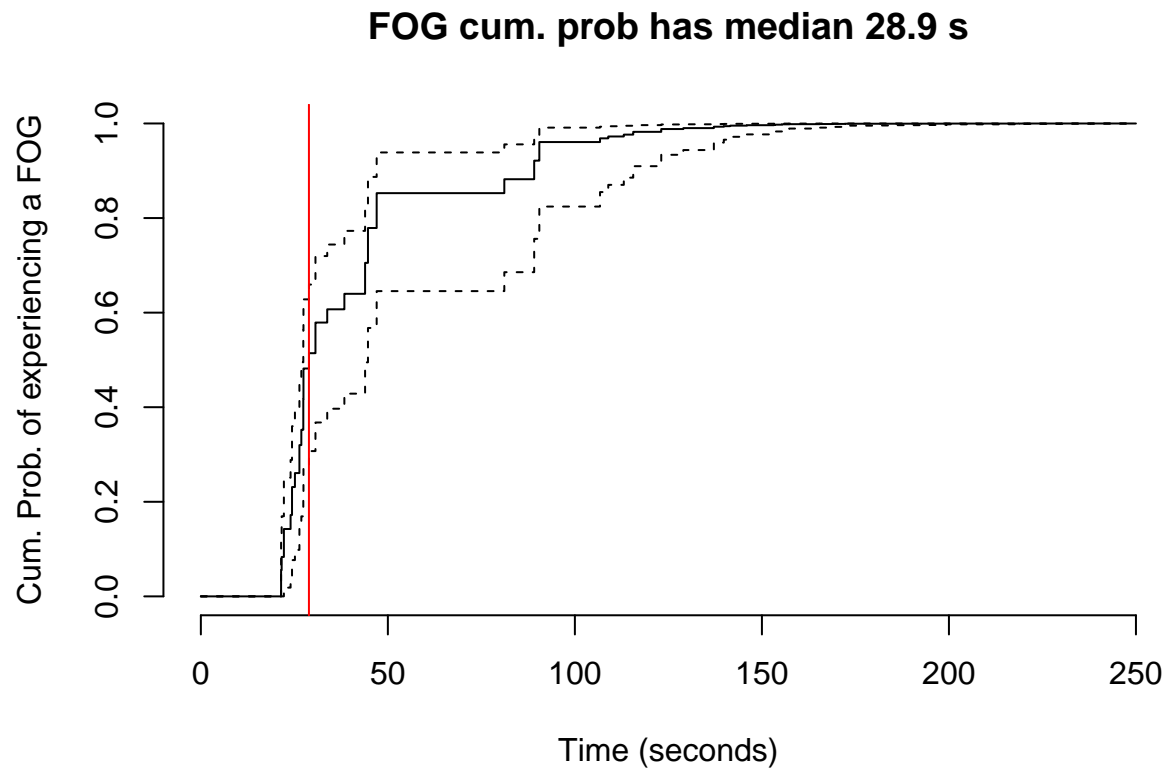
## Medication

What proportion of subjects are under Parkinson disease medication ?

```
## Medication
## off   on
## 0.69 0.31
```

0.31 of total patients are under medication.

## Modelling and Analysis : overall time to FOG.



```
## 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
```

Half of patients experience FOG within 28.911 s.

## Univariate analysis : Survival difference in group

Sex

```
## Call: survfit(formula = Surv(Begin, End, Kinetic) ~ Sex, data = fog)
##
##      records n.max n.start events median 0.95LCL 0.95UCL
## Sex=Female      786    86     17   597   43.9    38.4     NA
## Sex=Male      1446   127     19   933   27.4    24.4     NA

## # A tibble: 1 x 5
##   term      estimate std.error statistic p.value
##   <chr>      <dbl>    <dbl>    <dbl>   <dbl>
## 1 SexMale    0.0877    0.0613     1.43    0.152
```

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

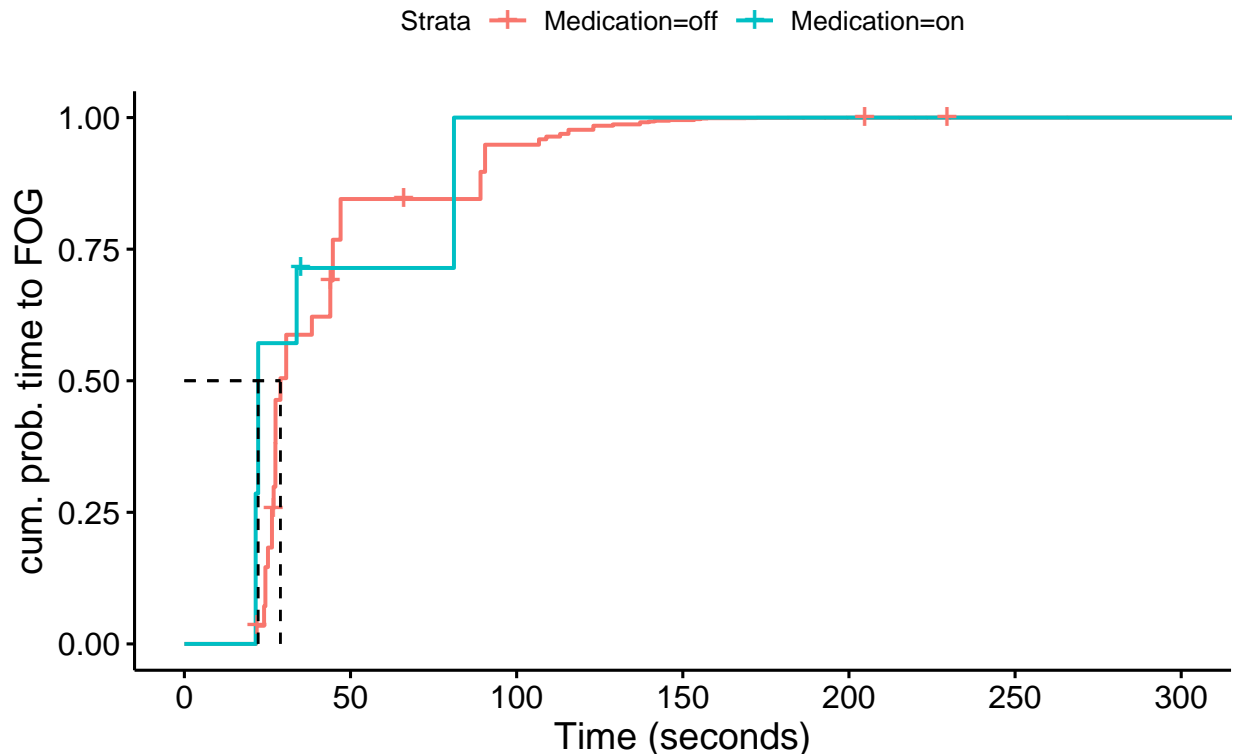
## Medication

```
## # A tibble: 1 x 5
##   term          estimate std.error statistic  p.value
##   <chr>         <dbl>    <dbl>    <dbl>    <dbl>
## 1 Medicationon    0.718    0.0677     10.6 2.75e-26
```

P-value = 0 and is less than 0.05. There is an effect of Medication on time to FOG , holding other variables fixed.

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

## Time to fog is short for subjects under medication



Half subjects under medication condition, experience FOG within 28.911, 22.203.

## Task type

```
## 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
```

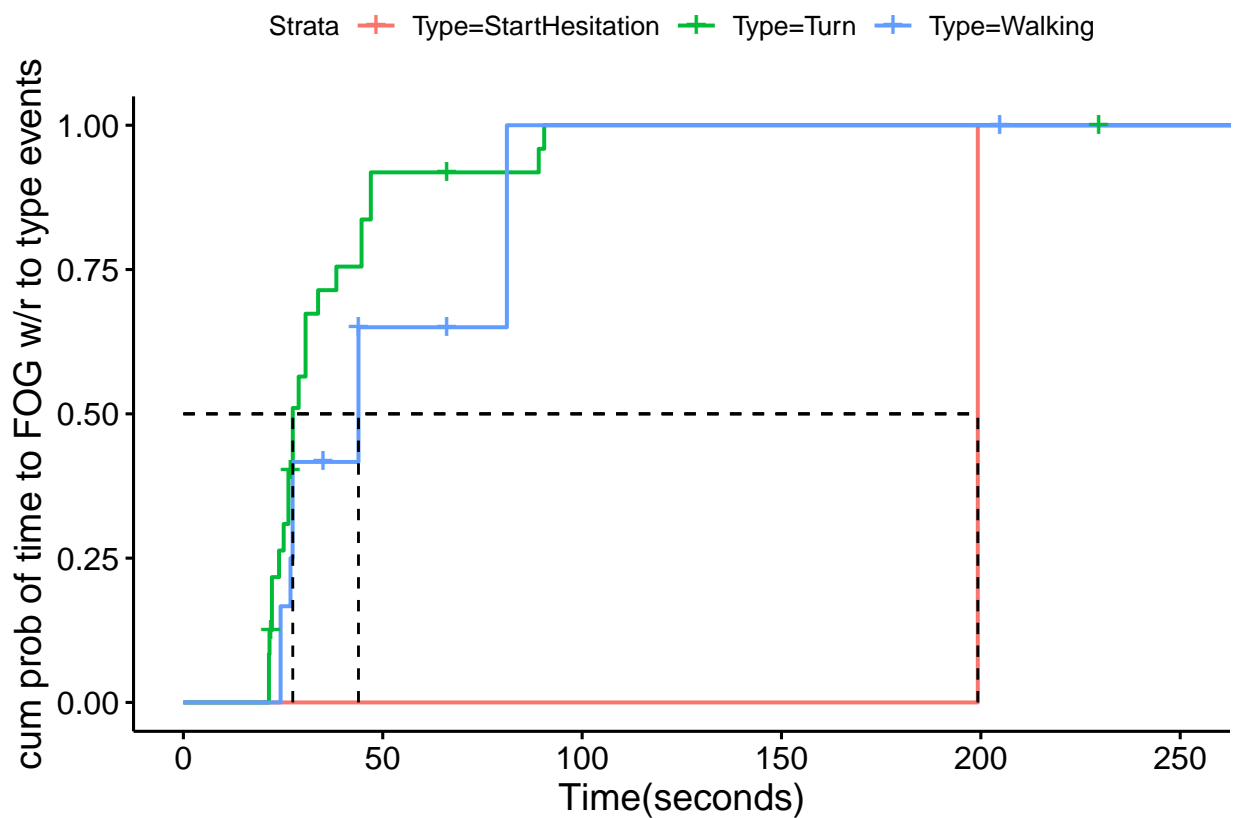
```
##      <chr>          <dbl>  <dbl>
## 1 TaskHotspot1      1.03 0.0127
## 2 TaskTUG-C         1.08 0.0186
## 3 TaskTurning-DT   -1.24 0.00165
```

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

## events Type

```
## # A tibble: 0 x 3
## # ... with 3 variables: term <chr>, estimate <dbl>, p.value <dbl>
```

There is no effect of type of events on time to FOG.

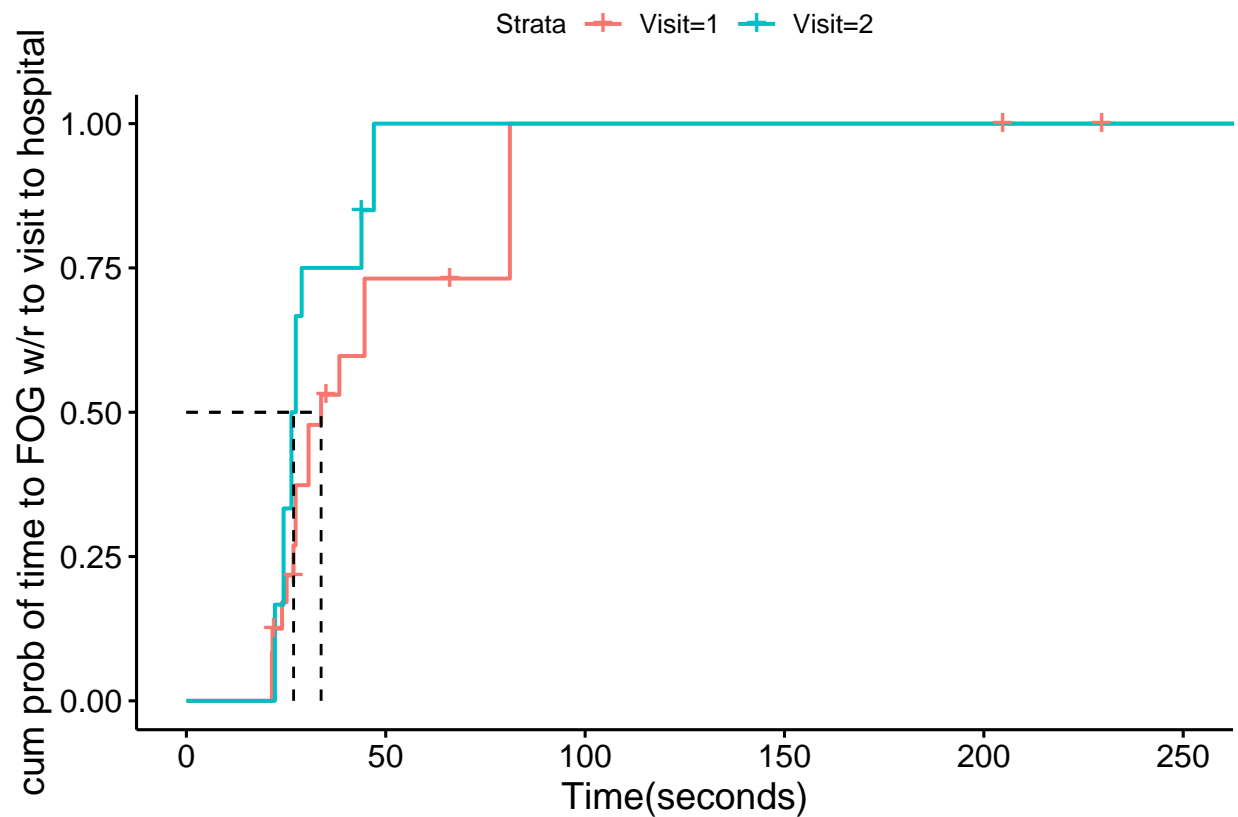


## visit

```
## Call: survfit(formula = Surv(Begin, End, Kinetic) ~ Visit, data = fog)
##
##      records n.max n.start events median 0.95LCL 0.95UCL
## Visit=1    1561   149    24   1017   33.8    27.4    NA
## Visit=2     671    76    12    513   26.9    24.4    NA
```

```
## # A tibble: 1 x 3
##   term      estimate p.value
```

```
##      <chr>      <dbl>      <dbl>
## 1 Visit2      0.254 0.0000442
```



p-value is less than 0.05. There is a difference in time to FOG with respect to visit to hospital, holding other variables fixed.

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 = 0.51 which is  $> 0.05$ . There is no effect of Year since diagnosis on FOG occurrence.

Multivariate analysis : regression

```
## # A tibble: 2 x 8
##   term          estimate std.error statistic  p.value conf.low conf.high Hazard-1
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 Medicationon  0.740    0.0702    10.5 5.84e-26  0.602    0.878    2.10
```



```
## 2 Visit2          0.240    0.0630      3.81 1.36e- 4    0.117    0.364    1.27
## # ... with abbreviated variable name 1: Hazard_Rate
```

Medication and Visit present a p value  $< 0.05$ . Medication increases the risk of experiencing FOG by 2.09s holding other variables fixed, while , visit does the same with an increase of time to FOG by a rate of 1.27s while holding other features 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 early occurrence when subjects are under Medication and having Visit to hospital. Overall, half of subjects experience FOG within 28.9 seconds. It could be interesting to investigate why subjects under medication and having visit to hospital experience earlier FOG. Also, survival difference given kinetic outcomes and tasks could be investigate too.

## 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.

<https://www.kaggle.com/competitions/tlvmc-parkinsons-freezing-gait-prediction/data>