

# Replication of the Multi-State Model

2024-04-25

```
knitr::opts_chunk$set(  
  echo = FALSE,  
  warning = FALSE,  
  message = FALSE,  
  fig.align = "left",  
  out.width = "100%",  
  tidy = TRUE  
)
```

This script takes the prepared data (from *replication\_data\_prep*) and fits the multistate survival model to it. This is an attempt to replicate the outcomes from ten Broeke et al. (2022), and therefore much of the code is replicated, and should be accredited to, the authors of this paper.

## Descriptives

### Sample size

```
N <- dat[, uniqueN(user_id)]  
cat("Total Sample Size:\n")
```

```
## Total Sample Size:
```

```
N
```

```
## [1] 24859
```

```
N_append <- data.frame(grade = "total",  
                        N = N)  
N_grade <- dat[, .(N = uniqueN(user_id)), by = grade][order(grade)]  
cat("Sample Size Per Grade:\n")
```

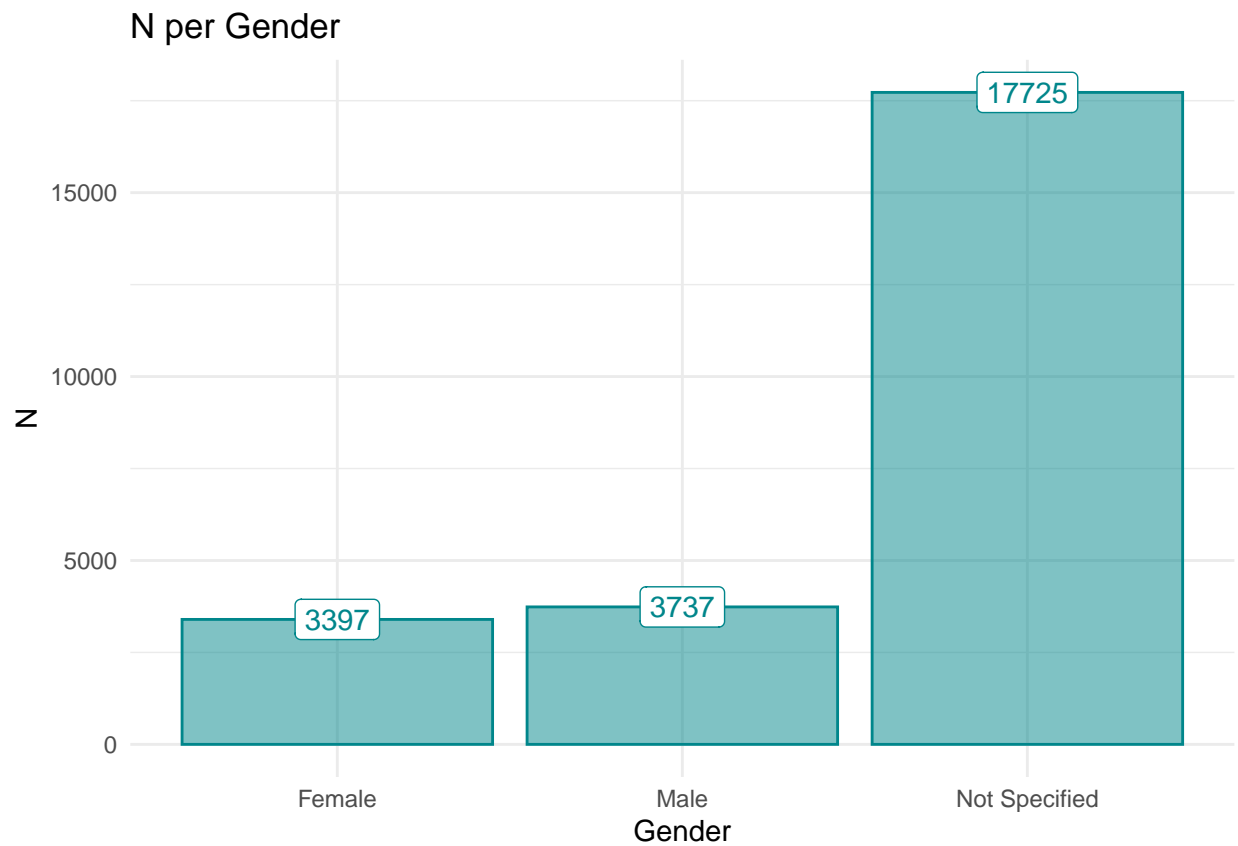
```
## Sample Size Per Grade:
```

```
N_grade
```

```
##   grade    N  
## 1:     3 4343  
## 2:     4 4816  
## 3:     5 4770  
## 4:     6 4304  
## 5:     7 3869  
## 6:     8 2787
```

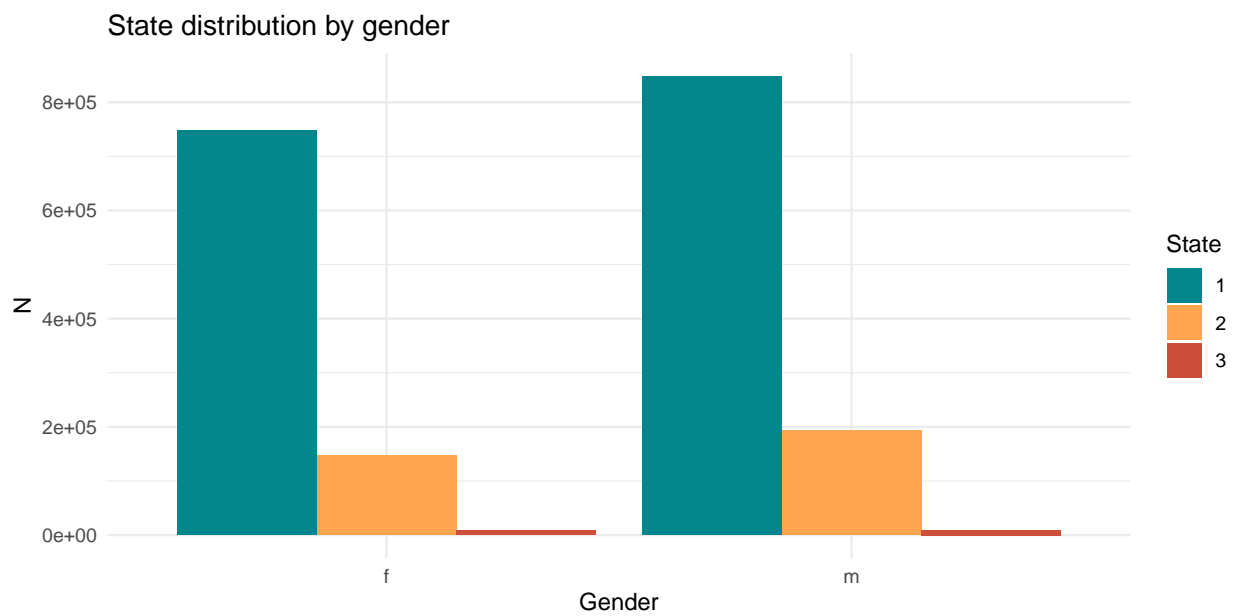
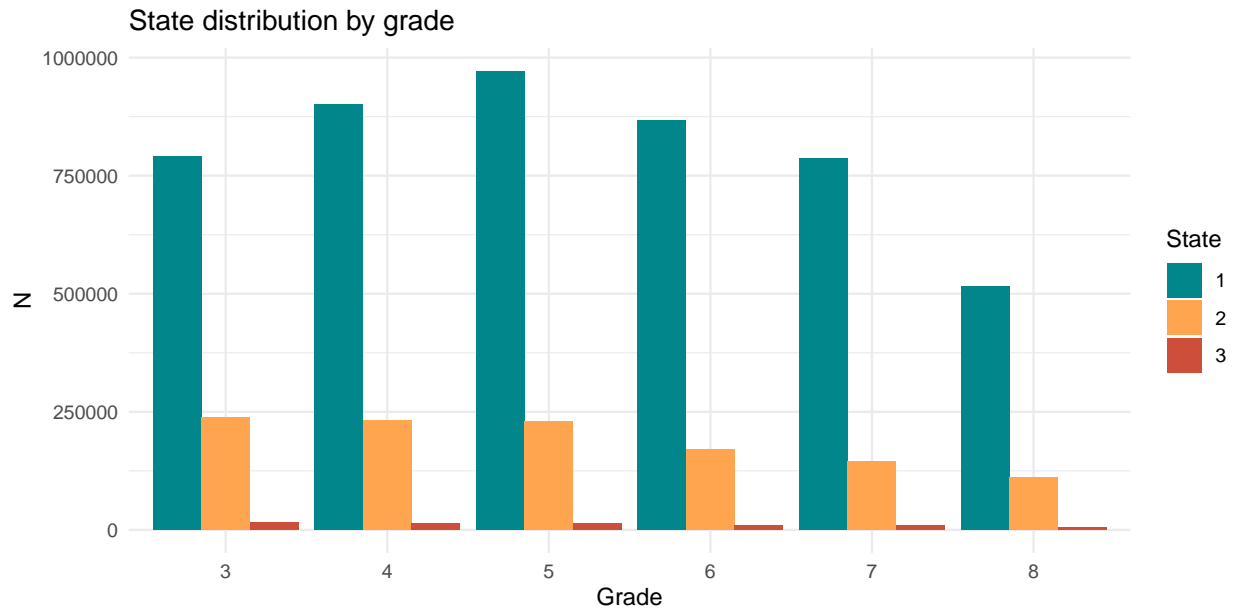
```
N_grade_with_total <- rbind(N_grade, N_append)
write.csv(N_grade_with_total,
  paste0(path_to_tables, "/N.csv"))
```

There are slightly more males compared to females in the data.



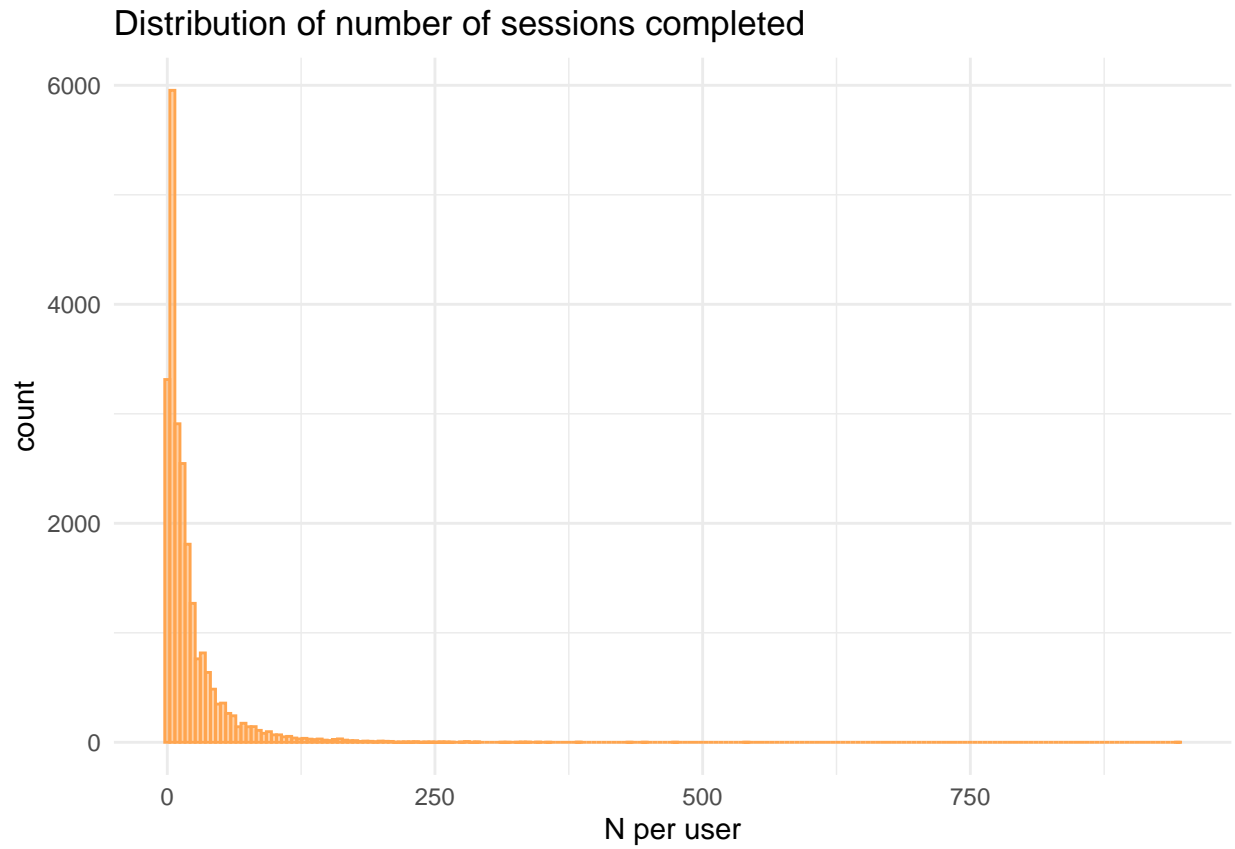
## State distribution

Here, I examine how the frequency of being in a certain state (1, 2, or 3) varies across grade and gender in the data.



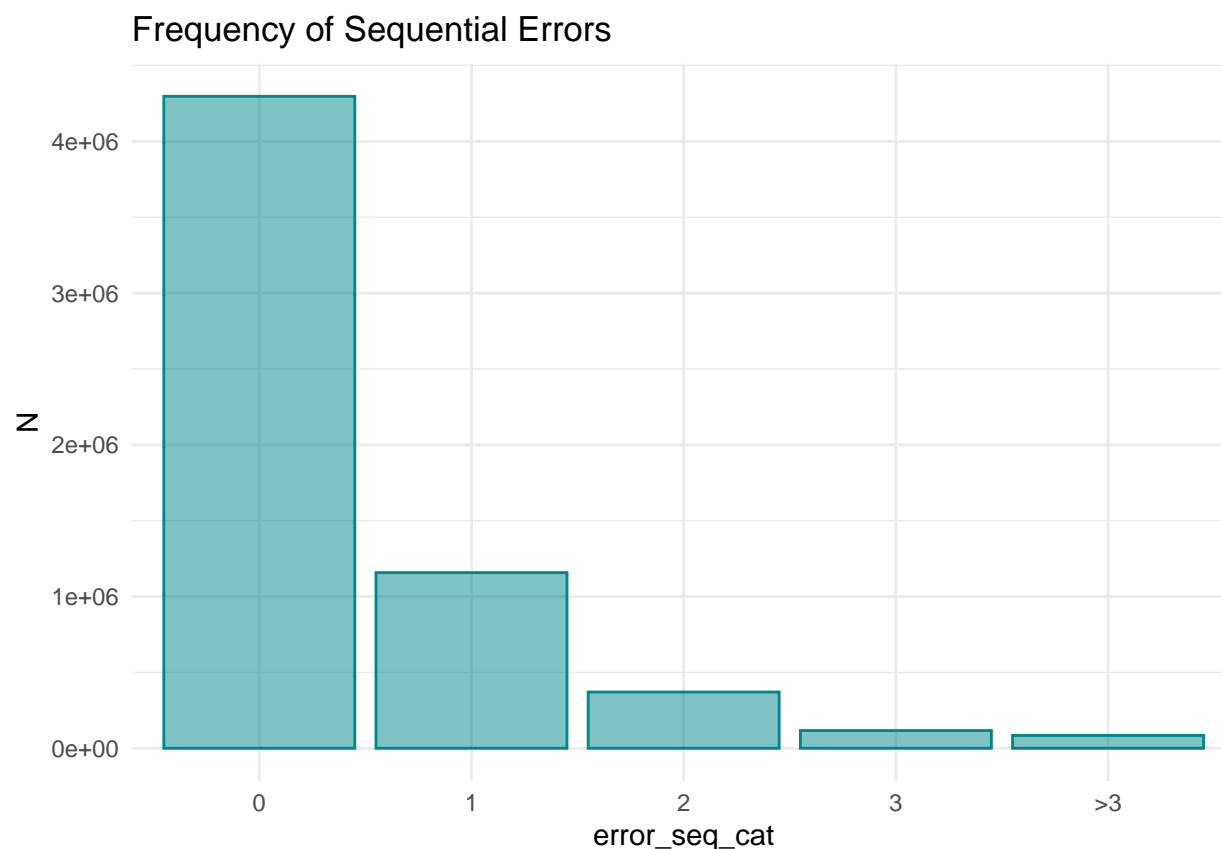
## Number of sessions completed

The majority of users have completed 1-10 full sessions. This ranges from 1 to 945 completed sessions.



### Sequential errors

There is a large amount of errors (sequential errors = 1), but the frequency of 2, 3, and more than 3 sequential errors gradually decreases.



## Multistate Survival Model

### Transition matrix

For a model with three states, a transition intensity matrix ( $Q$ ) is defined:

##	Persisting	Soft-Quit	Hard-Quit
## Persisting	1	1	1
## Soft-Quit	1	1	1
## Hard-Quit	0	0	0

### Model fitting

The model fit procedures are specified in a separate script, “*modelfit\_msm.R*”. The outputs from that model fitting have been saved and can be loaded into the environment below.

*msm\_constrained\_all*, *msm\_constrained\_rt* and *msm\_constrained\_tz* are the constrained models. That is, the multistate models without any covariates added.

*msm\_covariate\_all*, *msm\_covariate\_rt* and *msm\_covariate\_tz* are the covariate models. The following covariates were added to these models:

- Gender

- Grade (higher grades vs. lower grades)
- Difficulty (high vs. low, and high vs. medium)
- Speed of errors (fast error vs. no error and slow error vs. no error)
- Playing during vs. outside school hours
- Sequential errors (vs. no error: 1, 2, 3, >3)

*hr\_msm\_covariate\_rt* and *hr\_msm\_covariate\_tz* specify the transition probabilities for each covariate (hazard ratios) in the rekentuin and taalzee models, respectively.

## State transition rates

### Transition intensities

Here, the estimated Q-matrices for the constrained and covariate models are extracted.

```
model <- "covariate" # "constrained" or "covariate"
```

```
##               estimate      SE          95% CI
## Persisting > Persisting -0.057 0.000 [ -0.058  -0.056 ]
## Persisting > Soft-Quit  0.241 0.002 [  0.238   0.245 ]
## Soft-Quit > Persisting  0.034 0.000 [  0.034   0.035 ]
## Soft-Quit > Soft-Quit  -0.280 0.002 [ -0.283  -0.276 ]
## Hard-Quit > Persisting  0.023 0.000 [  0.022   0.023 ]
## Hard-Quit > Soft-Quit   0.038 0.001 [  0.037   0.04 ]
```

### Transition Probabilities

Also extracting state transition probability matrices

```
## Transition probabilities for covariate model
##           Persisting Soft-Quit Hard-Quit
## Persisting 0.9480690 0.02923061 0.02270036
## Soft-Quit  0.2047395 0.75930115 0.03595936
## Hard-Quit  0.0000000 0.00000000 1.00000000
```

### Likelihood Ratio Test

```
## [1] NA 44
```

```
##           model      AIC minus_2_loglik lr_test_loglik lr_test_df lr_test_df_diff
## 1 constrained 1668098      1668090      -834045         4          NA
## 2  covariate  429940      429844      -214922        48          44
## lr_test_chisq lr_test_p
## 1             NA       NA
## 2      1238246         0
```

The covariate model significantly improves model fit, ( $\chi^2(44) = 1238246$ ,  $p < 0.001$ ), supporting the further use of this model.

## Hazard ratios

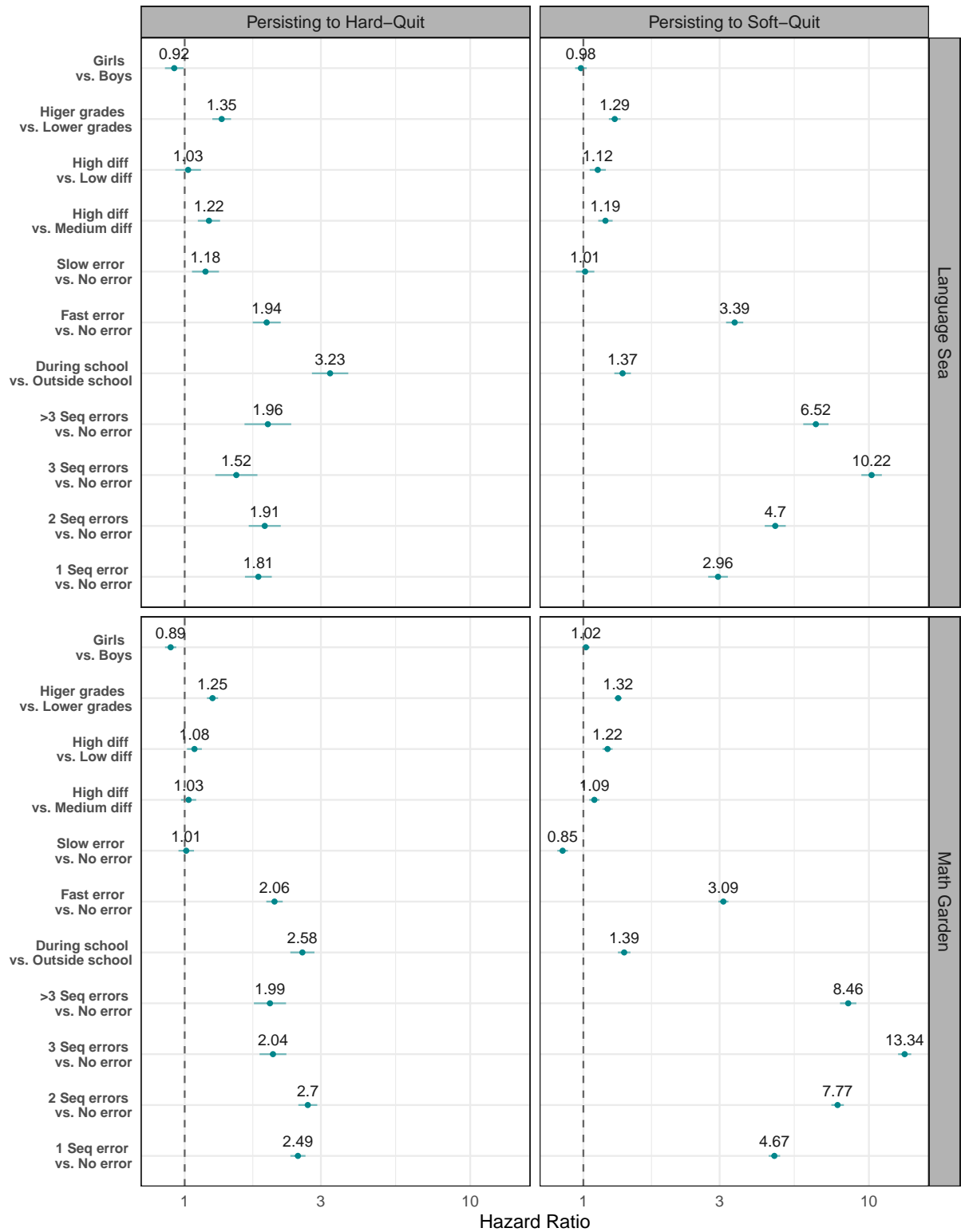
The effect of each covariate on the model is expressed in the form of Hazard Ratios. Specifically, they are a maximum likelihood computation of the relative effect of the covariate on the probability of a state transition.

The below code extracts the hazard ratios from the covariate model, and adds them to a table (*hazardratios*) that can be used to plot the data.

## Plot

I plot the hazard ratios and their confidence intervals for the persisting > soft-quit and persisting > hard-quit transitions for Math Garden and Language Sea, respectively. The plot gets saved to the directory under the folder “plots”.

```
## pdf
## 2
```



## pdf  
## 2



## Hazard ratio tables

In order to easily compare the values to the previous paper, here I extract the hazard ratios, with their confidence intervals, in table format and save them to the directory.

##	Transition	Covariate	Hazard Ratio
## 1:	Persisting to Soft-Quit	1 Seq error \n vs. No error	4.666
## 2:	Persisting to Soft-Quit	2 Seq errors \n vs. No error	7.768
## 3:	Persisting to Soft-Quit	3 Seq errors \n vs. No error	13.341
## 4:	Persisting to Soft-Quit	>3 Seq errors \n vs. No error	8.462
## 5:	Persisting to Soft-Quit	Fast error \n vs. No error	3.092
## 6:	Persisting to Soft-Quit	Slow error \n vs. No error	0.846
## 7:	Persisting to Soft-Quit	During school \n vs. Outside school	1.390
## 8:	Persisting to Soft-Quit	High diff \n vs. Medium diff	1.093
## 9:	Persisting to Soft-Quit	High diff \n vs. Low diff	1.215
## 10:	Persisting to Soft-Quit	Higer grades \n vs. Lower grades	1.324
## 11:	Persisting to Soft-Quit	Girls \n vs. Boys	1.021
## 12:	Persisting to Hard-Quit	1 Seq error \n vs. No error	2.493
## 13:	Persisting to Hard-Quit	2 Seq errors \n vs. No error	2.699
## 14:	Persisting to Hard-Quit	3 Seq errors \n vs. No error	2.038
## 15:	Persisting to Hard-Quit	>3 Seq errors \n vs. No error	1.989
## 16:	Persisting to Hard-Quit	Fast error \n vs. No error	2.064
## 17:	Persisting to Hard-Quit	Slow error \n vs. No error	1.013
## 18:	Persisting to Hard-Quit	During school \n vs. Outside school	2.584
## 19:	Persisting to Hard-Quit	High diff \n vs. Medium diff	1.032
## 20:	Persisting to Hard-Quit	High diff \n vs. Low diff	1.082
## 21:	Persisting to Hard-Quit	Higer grades \n vs. Lower grades	1.252
## 22:	Persisting to Hard-Quit	Girls \n vs. Boys	0.893
##	Transition	Covariate	Hazard Ratio
##	95% CI		
## 1:	[ 4.455 , 4.887 ]		
## 2:	[ 7.386 , 8.171 ]		
## 3:	[ 12.652 , 14.068 ]		
## 4:	[ 7.923 , 9.039 ]		
## 5:	[ 2.969 , 3.22 ]		
## 6:	[ 0.811 , 0.882 ]		
## 7:	[ 1.321 , 1.462 ]		
## 8:	[ 1.049 , 1.138 ]		
## 9:	[ 1.167 , 1.265 ]		
## 10:	[ 1.285 , 1.364 ]		
## 11:	[ 0.99 , 1.052 ]		
## 12:	[ 2.345 , 2.649 ]		
## 13:	[ 2.501 , 2.913 ]		
## 14:	[ 1.829 , 2.271 ]		
## 15:	[ 1.748 , 2.265 ]		
## 16:	[ 1.932 , 2.205 ]		
## 17:	[ 0.951 , 1.078 ]		
## 18:	[ 2.344 , 2.848 ]		
## 19:	[ 0.971 , 1.097 ]		
## 20:	[ 1.019 , 1.149 ]		
## 21:	[ 1.197 , 1.309 ]		
## 22:	[ 0.854 , 0.934 ]		
##	95% CI		

##	Transition	Covariate	Hazard Ratio
## 1:	Persisting to Soft-Quit	1 Seq error \n vs. No error	2.962
## 2:	Persisting to Soft-Quit	2 Seq errors \n vs. No error	4.699
## 3:	Persisting to Soft-Quit	3 Seq errors \n vs. No error	10.221
## 4:	Persisting to Soft-Quit	>3 Seq errors \n vs. No error	6.520
## 5:	Persisting to Soft-Quit	Fast error \n vs. No error	3.387
## 6:	Persisting to Soft-Quit	Slow error \n vs. No error	1.014
## 7:	Persisting to Soft-Quit	During school \n vs. Outside school	1.372
## 8:	Persisting to Soft-Quit	High diff \n vs. Medium diff	1.195
## 9:	Persisting to Soft-Quit	High diff \n vs. Low diff	1.123
## 10:	Persisting to Soft-Quit	Higer grades \n vs. Lower grades	1.288
## 11:	Persisting to Soft-Quit	Girls \n vs. Boys	0.980
## 12:	Persisting to Hard-Quit	1 Seq error \n vs. No error	1.812
## 13:	Persisting to Hard-Quit	2 Seq errors \n vs. No error	1.907
## 14:	Persisting to Hard-Quit	3 Seq errors \n vs. No error	1.517
## 15:	Persisting to Hard-Quit	>3 Seq errors \n vs. No error	1.956
## 16:	Persisting to Hard-Quit	Fast error \n vs. No error	1.937
## 17:	Persisting to Hard-Quit	Slow error \n vs. No error	1.183
## 18:	Persisting to Hard-Quit	During school \n vs. Outside school	3.231
## 19:	Persisting to Hard-Quit	High diff \n vs. Medium diff	1.216
## 20:	Persisting to Hard-Quit	High diff \n vs. Low diff	1.028
## 21:	Persisting to Hard-Quit	Higer grades \n vs. Lower grades	1.348
## 22:	Persisting to Hard-Quit	Girls \n vs. Boys	0.919
##	Transition	Covariate	Hazard Ratio
##	95% CI		
## 1:	[ 2.736 , 3.207 ]		
## 2:	[ 4.318 , 5.115 ]		
## 3:	[ 9.41 , 11.101 ]		
## 4:	[ 5.888 , 7.22 ]		
## 5:	[ 3.16 , 3.631 ]		
## 6:	[ 0.941 , 1.092 ]		
## 7:	[ 1.283 , 1.467 ]		
## 8:	[ 1.127 , 1.266 ]		
## 9:	[ 1.052 , 1.199 ]		
## 10:	[ 1.228 , 1.35 ]		
## 11:	[ 0.935 , 1.027 ]		
## 12:	[ 1.626 , 2.019 ]		
## 13:	[ 1.676 , 2.17 ]		
## 14:	[ 1.279 , 1.798 ]		
## 15:	[ 1.62 , 2.361 ]		
## 16:	[ 1.73 , 2.169 ]		
## 17:	[ 1.061 , 1.319 ]		
## 18:	[ 2.79 , 3.741 ]		
## 19:	[ 1.112 , 1.33 ]		
## 20:	[ 0.927 , 1.14 ]		
## 21:	[ 1.25 , 1.453 ]		
## 22:	[ 0.853 , 0.991 ]		
##	95% CI		