Concepts of multilevel, longitudinal, and mixed models: Group 3

Robrecht Van Der Bauwhede Renée Blanckaert Group
Charlotte Vercammen Group Raïsa Carmen

## Contents



# 1 Introduction

The data set results from a longitudinal observational study, the aim of which is to study the post-operative evolution of the cognitive status of elderly hipfracture patients and their pre-operative cognitive status, and to study the effects of housing situation and age on these evolutions. The physical ability is measured using the MMSE (Mini Mental State Examination) score, with values between 0 and 30, where low values correspond to a bad cognitive condition, while high scores correspond to high cognitive condition of the patient. The pre-operative cognitive status is measured through the so-called 'neuro-status' which is a binary indicator for being neuro-psychiatric.

- id: patient identification number
- age: age of the patient at entry
- neuro: neuro-psychiatric status of the patient (1: neuro-psychiatric, 0: not neuro-psychiatric)
- mmse: MMSE score
- time: day after operation at which the MMSE score has been measured (1, 3, 5, 8, or 12)
- housing: the housing situation prior to the hip fracture (1: alone, 2: with family or partner, 3: nursing home)

# 2 Data exploration

Considering the completeness of the data, it can be observed that 5 persons' housing situation is unknown. Additionally, there was some dropout over time as show in Table 1. The column "Nb" shows the number of respondent at each time instance. The column "Return" shows the number of respondents that participated at time $t$ while they did not participate at time $t-1$; there is, for instance, one respondent that did not participate at $time = 1$ while he did participate at $time = 3$. The column "Dropout" shows the number of

Table 1: Dropout and patient characteristics over time.

| Time | Nb | Return | Dropout | Mean age | % neuro-psychiatric | Housing %alone | %family/ partner | %nursing home | %NA |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 58 | 0 | 0 | 78.71 | 31.03 | 29.31 | 39.66 | 22.41 | 8.62 |
| 3 | 57 | 1 | 2 | 78.18 | 33.33 | 29.82 | 38.60 | 22.81 | 8.77 |
| 5 | 59 | 2 | 0 | 78.59 | 32.20 | 28.81 | 38.98 | 23.73 | 8.47 |
| 8 | 52 | 0 | 7 | 77.88 | 30.77 | 28.85 | 36.54 | 25.00 | 9.62 |
| 12 | 38 | 0 | 14 | 77.82 | 28.95 | 31.58 | 34.21 | 23.68 | 10.53 |

respondents that did not participate at time $t$ while they did participate at time $t-1$. It can be seen than many drop out at $time = 12$. The other columns show how the patient characteristics change over time as patients are added or lost from the study. Overall, there is not much variation which might indicate that the dropout of patients is not related to MMSE or patient characteristics. We will thus assume dropout is completely random.

Figure 1 shows the average evolution of MMSE over time (Loess curves) for groups of patients with different housing and/or neuro-psychiatric status. Patients that are not neuro-psychiatric seem to have higherMMSE that stays reasonably constant over time (except for nursing home patients). MMSE seems to go up over time for nursing home patients (but they are the smallest group), and down for neuro-spychiatric patients that live with their family or partner (or unknown housing). MMSE for neuro-psychiatric patients that live alone might have a quadratic evolution over time.

Figure 2 shows all patient profiles. There is quite a lot of variation between the patients' evolution. The non-psychiatric patients are also under-represented (except in the nursing home group).

The last variable that is explored, is the age. The loess curves in Figure 3 show that age might be negatively correlated with MMSE.

### 2.1 Variance analysis and correlation structure

Figure 4 shows that the variance is larger for the neuro-psychiatric patients (the smallest group).

Figure **??** shows that there seems to be an inverse relationship between the mean mmse and the variance for patients with either neuro-psychiatric status.

### 2.2 Conclusion exploratory analysis

The exploratory analysis has shows that the pattern of mmse over time is likely not constant or even linear. One might attempt quadratic, cubic, or logarithmic transformations of time to accommodate this. Both the level (intercept) of mmse differs quite a lot, depending on neuro-psychiatric status, housing and age.

There seems to be high intraclass correlation???

## 3 Methodology and results

This section, gradually develops a statistical model that seems to fit the data best. It starts with a simple model where only limited covariates are included in 3.1

### 3.1 A simple model

This first, simple model will assume a linear relationship between mmse and $log(time)$ for each patient. The model allows for subject-specific intercepts and slopes and the neuro-psychiatric status is the only additional

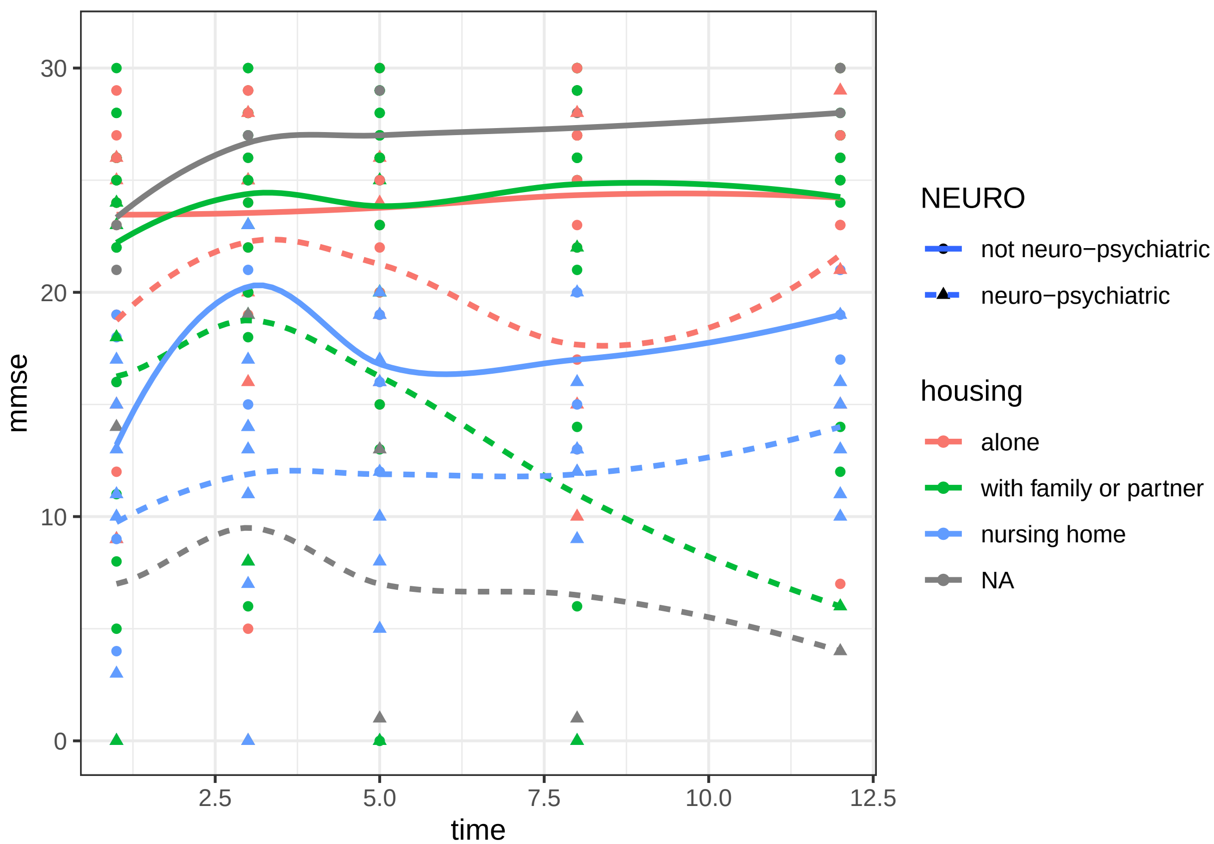


Figure 1: MMSE over time for all patient profiles, including smoothed curves.

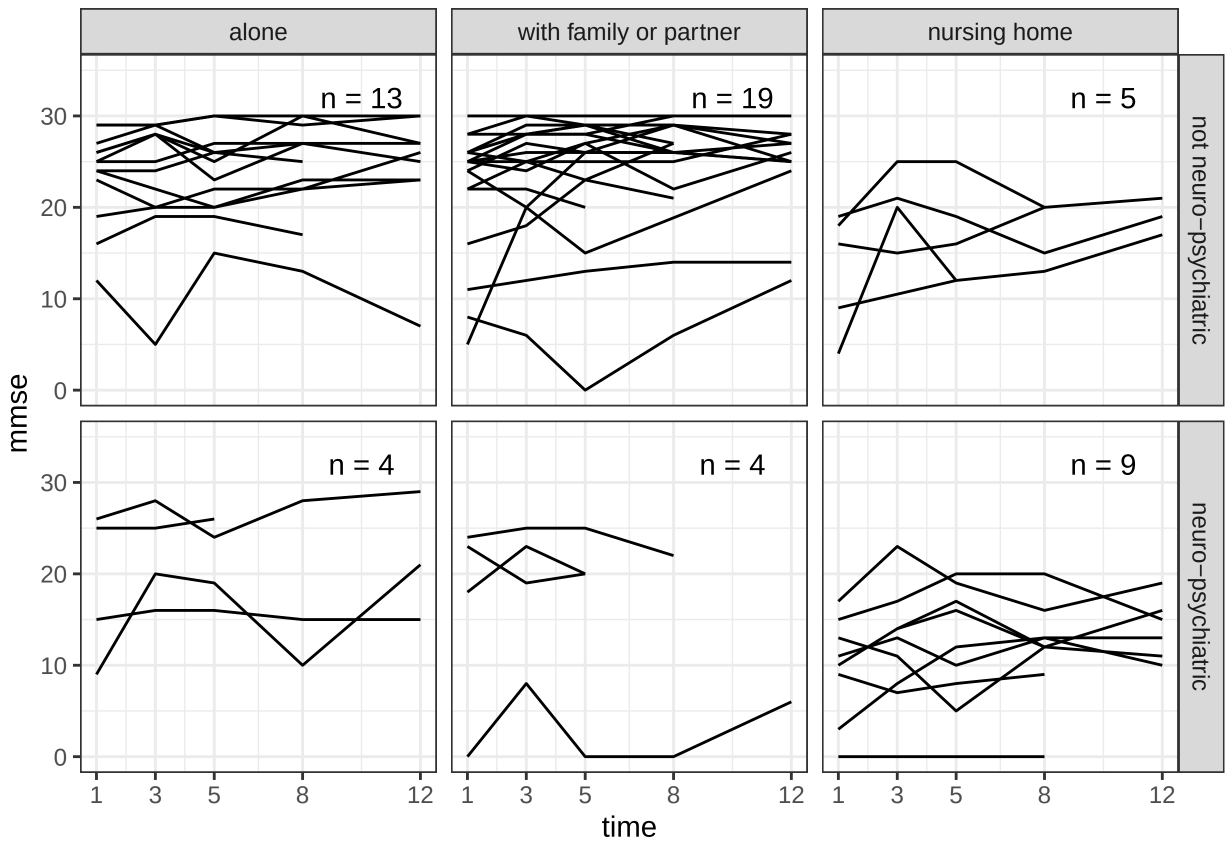


Figure 2: Patient profiles of MMSE over time.

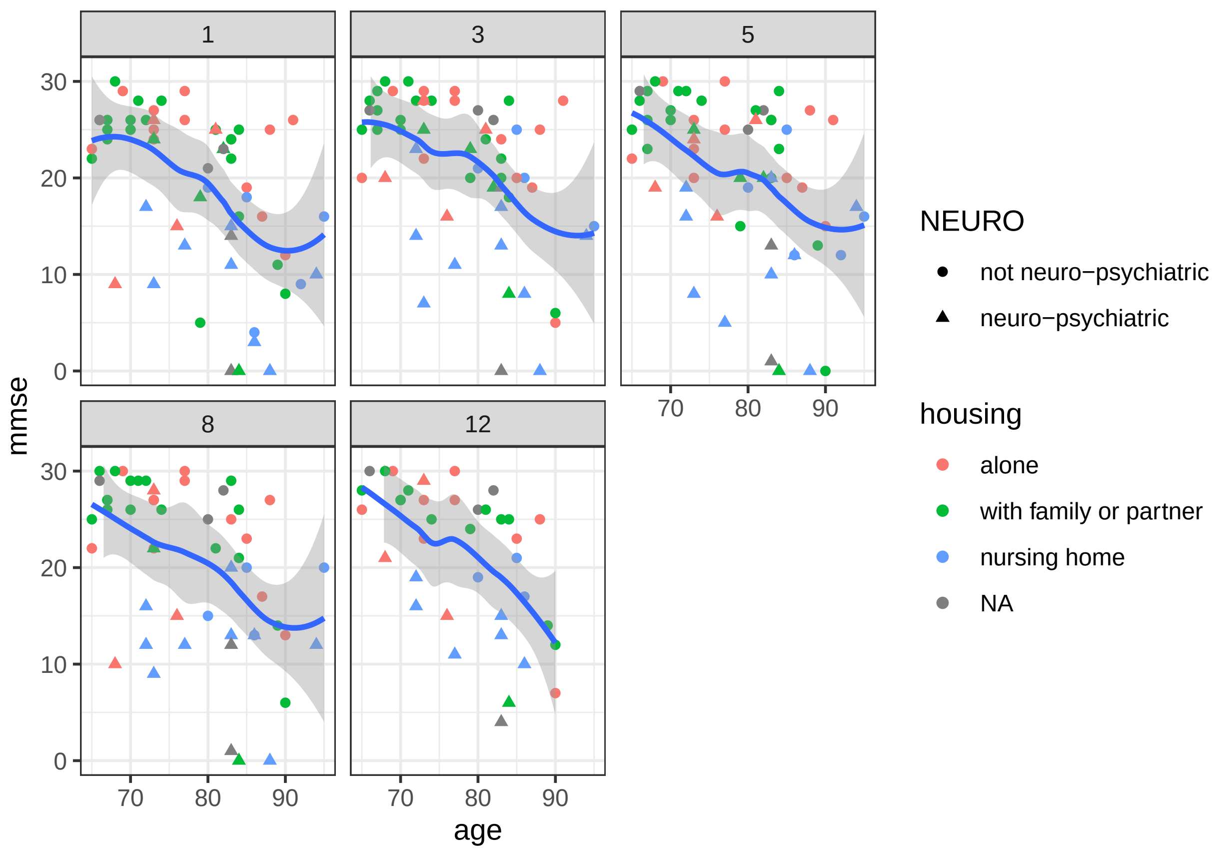


Figure 3: MMSE versus age. Facets show the time instances.

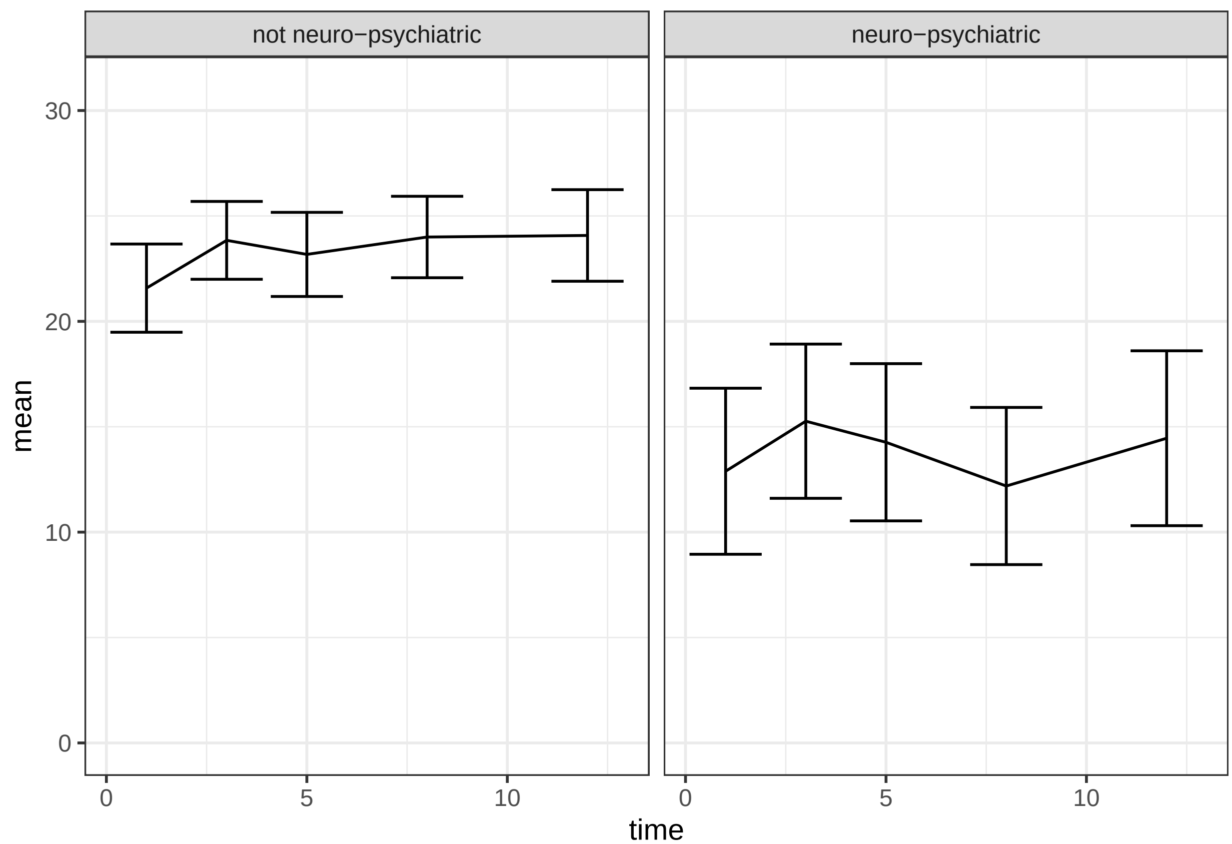


Figure 4: Error bars with 95 percent confidence intervals.

Table 2: Fixed effects for the simple model with random intercept and slope

| | Estimate | Std. Error | df | t value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| (Intercept) | 21.762 | 1.142 | 56.891 | 19.065 | 0.000 |
| logtime | 0.953 | 0.227 | 53.841 | 4.194 | 0.000 |
| NEUROneuro-psychiatric | -8.381 | 2.016 | 57.314 | -4.158 | 0.000 |
| logtime:NEUROneuro-psychiatric | -0.327 | 0.415 | 57.976 | -0.788 | 0.434 |

Table 3: Random effects

| grp | var1 | var2 | vcov | sdcor |
|---|---|---|---|---|
| id | (Intercept) | NA | 47.383 | 6.884 |
| id | logtime | NA | 0.203 | 0.451 |
| id | (Intercept) | logtime | -2.370 | -0.764 |
| Residual | NA | NA | 5.957 | 2.441 |

explanatory variable that is taken into account.

Table 2 shows the fixed effects for the simple model where time is on a log scale, with a random intercept and slope. For the average not-neuro-psychiatric patient, the slope is positive and the intercept is at 21.7624198. For the average neuro-psychiatric patients, the intercept is at 13.3817701 and the slope is less steep (as shown in Figure 6) but the difference in slope is not significant.

Figure 5 shows that the model performs reasonably well; there are no real remaining trends (all curves going up or down) and most residuals seem to randomly fluctuate around zero (no profiles are consistently above or below zero). There are, however, some individuals whose residuals are far larger (both positive and negative) than others.

Figure 7 shows the inverse relationship between the slope and the intercept: the higher the intercept, the lower the slope for each of the neuro-psychiatric statuses. This makes sense because MMSE is not a real continuous variable. It has a maximum value of 30 which means that there is less room for an increase in MMSE (meaning a lower slope) if the starting value (intercept) for MMSE is already high.

## 3.2 Model with housing and age

Before adding age to the model, it is centered around the mean (mean age is 78.28). Table 4 shows the results for the fixed effects. Non neuro-phsychiatric patients that are 78.28 years old and live alone are the baseline and have the highest intercept (23.12). For each increase in age of one year, the intercept is estimated to be 16.8659782 lower and neuro-psychiatric patients still have lower intercept and slope. Patients that live with family or partner have slightly lower intercept (not significant) than those who live alone. Patients that live in a nursing home have a significantly lower intercept than those who live alone. There is no significant difference in slope between neuro-psychiatric and non neuro-psychiatric patients.

## 3.3 A model with a dichotomized dependent variable

Lastly, a model is tested where mmse is dichotomized using a median split; all mmse scores equal to or below the median (23) are cetagorized as “Low” and all mmse scores over 23 are categorized as “High”. It should be noted that the literature warns against this technique as it is rarely justified from either a conceptual or statistical perspective [@maccallum2002practice].

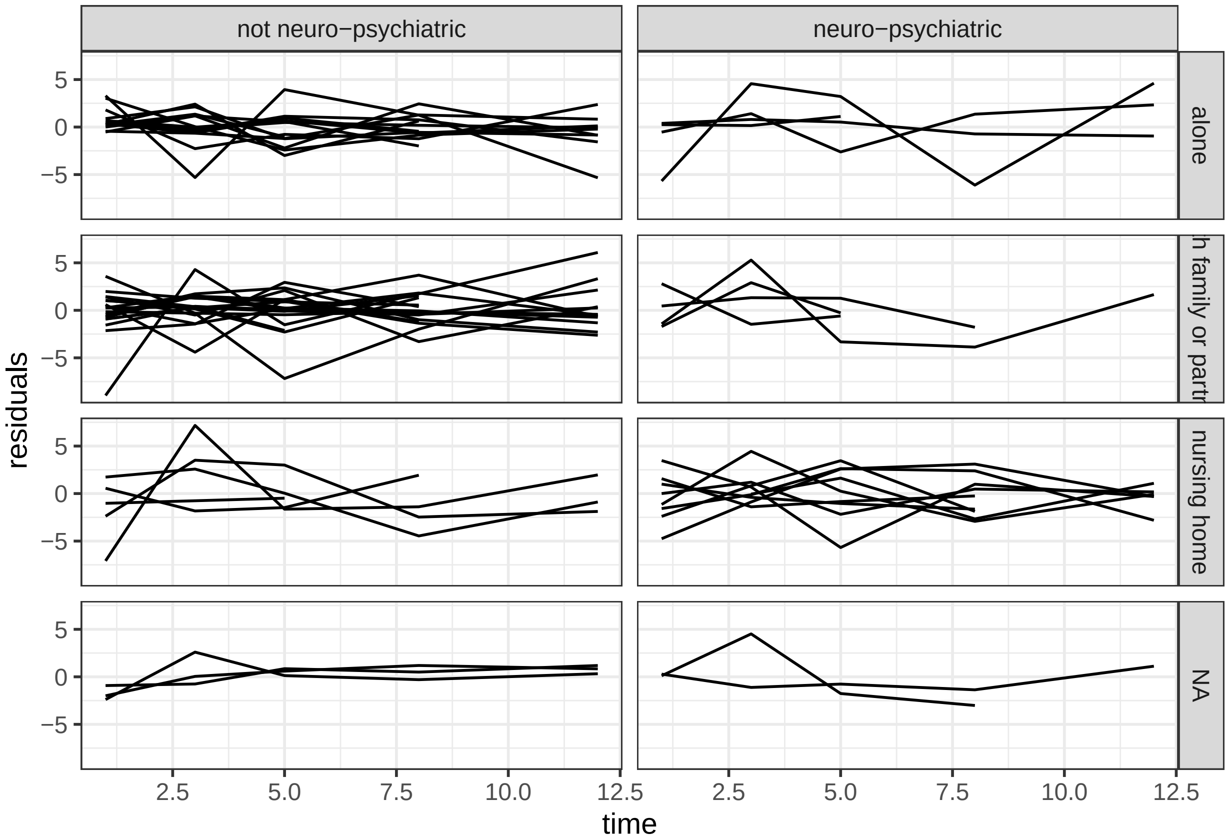


Figure 5: Residuals and time.

Table 4: Fixed effects for a model with random intercept and slope, including age and housing

| | Estimate | Std. Error | df | t value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| (Intercept) | 23.123 | 1.371 | 56.189 | 16.866 | 0.000 |
| logtime | 0.851 | 0.243 | 49.821 | 3.508 | 0.001 |
| NEUROneuro-psychiatric | -5.597 | 1.859 | 53.010 | -3.010 | 0.004 |
| agecenter | -0.368 | 0.090 | 49.306 | -4.084 | 0.000 |
| housingwith family or partner | -1.503 | 1.632 | 48.942 | -0.921 | 0.362 |
| housingnursing home | -5.282 | 2.020 | 49.305 | -2.615 | 0.012 |
| logtime:NEUROneuro-psychiatric | -0.194 | 0.447 | 53.060 | -0.434 | 0.666 |

Table 5: Random effects

| grp | var1 | var2 | vcov | sdcor |
|---|---|---|---|---|
| id | (Intercept) | NA | 29.056 | 5.390 |
| id | logtime | NA | 0.195 | 0.442 |
| id | (Intercept) | logtime | -1.665 | -0.699 |
| Residual | NA | NA | 6.256 | 2.501 |

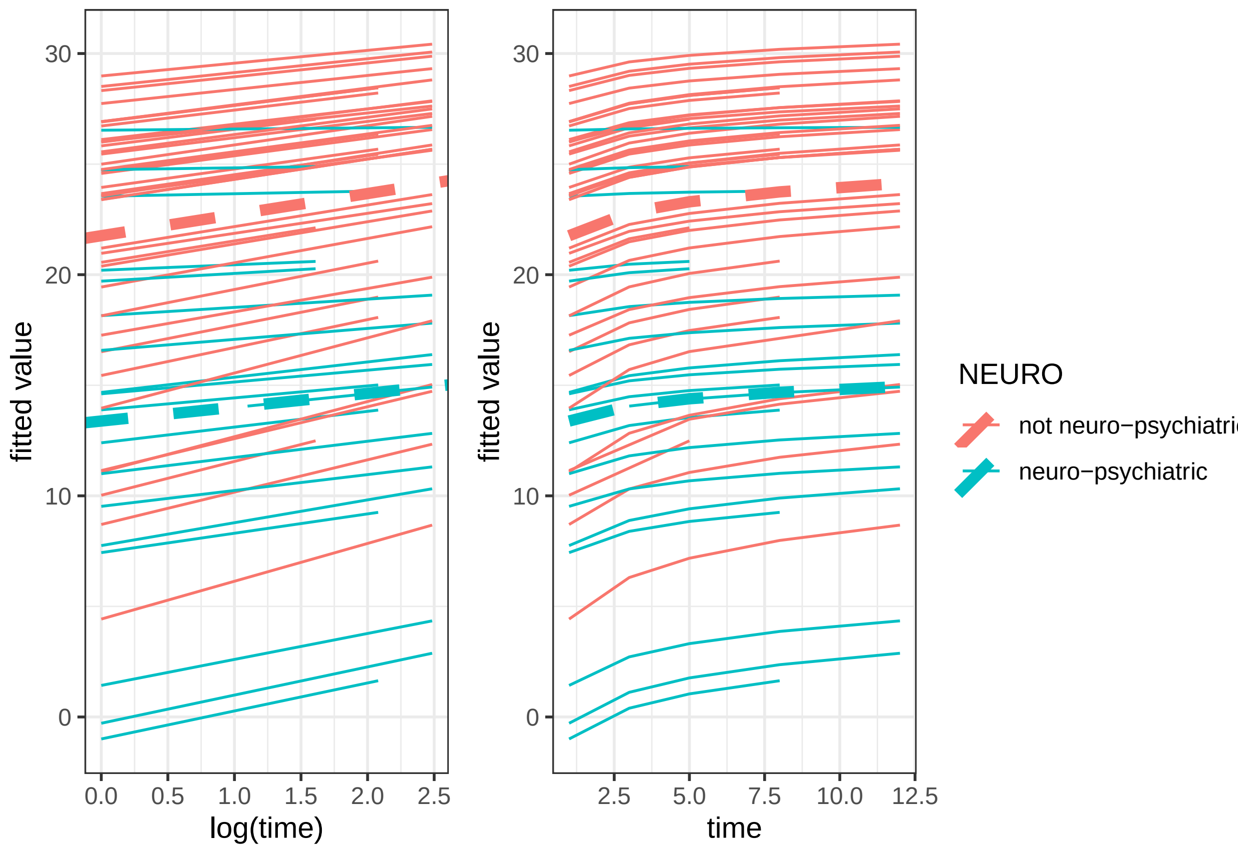


Figure 6: Fitted values plotted against log(time) and time. Dashed, thicker lines are the predicted trends based on the fixed effects only, for each group.

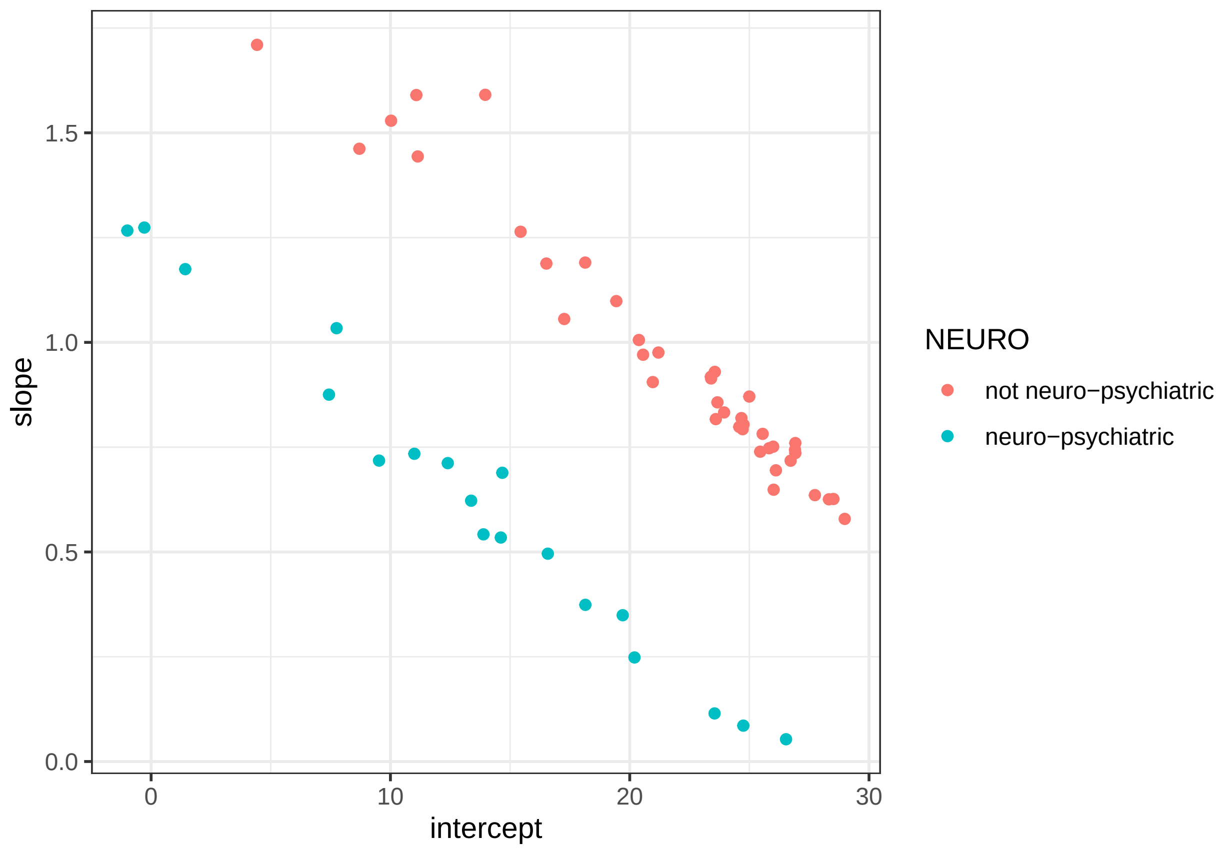


Figure 7: Scatterplot of the fitted intercept and slopes (log(time)).

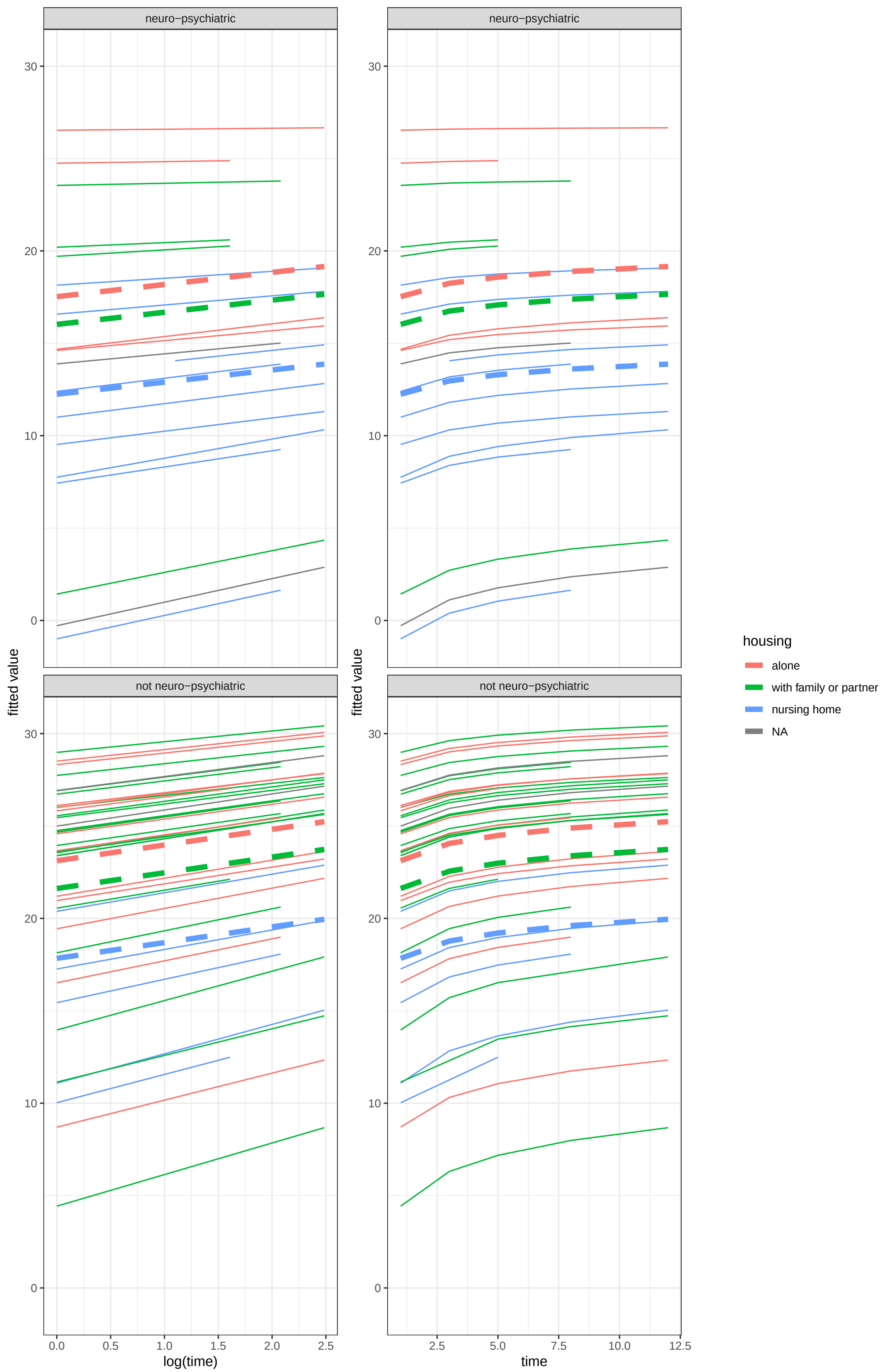


Figure 8: Fitted values plotted against log(time) and time. Dashed, thicker lines are the predicted trends based on the fixed effects only, for an average-aged patient in each group.

# 4 Conclusion

# 5 Bibliography