

Causal Inference for Cognitive Decline Using Mixed Effects Models: Detailed Analysis

Daniel Dema

2025-06-08

Introduction

This document provides the details for the construction of the models presented in the report, focusing primarily on the analysis of interaction terms and diagnostics for random effects and residuals. This document is written under the assumption that it is being read with the report readily available; details about the dataset and references are omitted to minimize repetition of the context established in the report.

Data Loading and Cleaning

Note that in the original dataset, SES is ranked backwards, in the sense that 1 is the highest and 5 is the lowest. When loading the data, we reverse the order for ease of interpretability.

```
df <- read.csv("D:/Daniel/Documents/MATH6642/final_project/oasis_longitudinal.csv")
dc <- df[!is.na(df$MMSE) & !is.na(df$SES), ]

dc$SES <- 5 - dc$SES

dc_orig <- dc

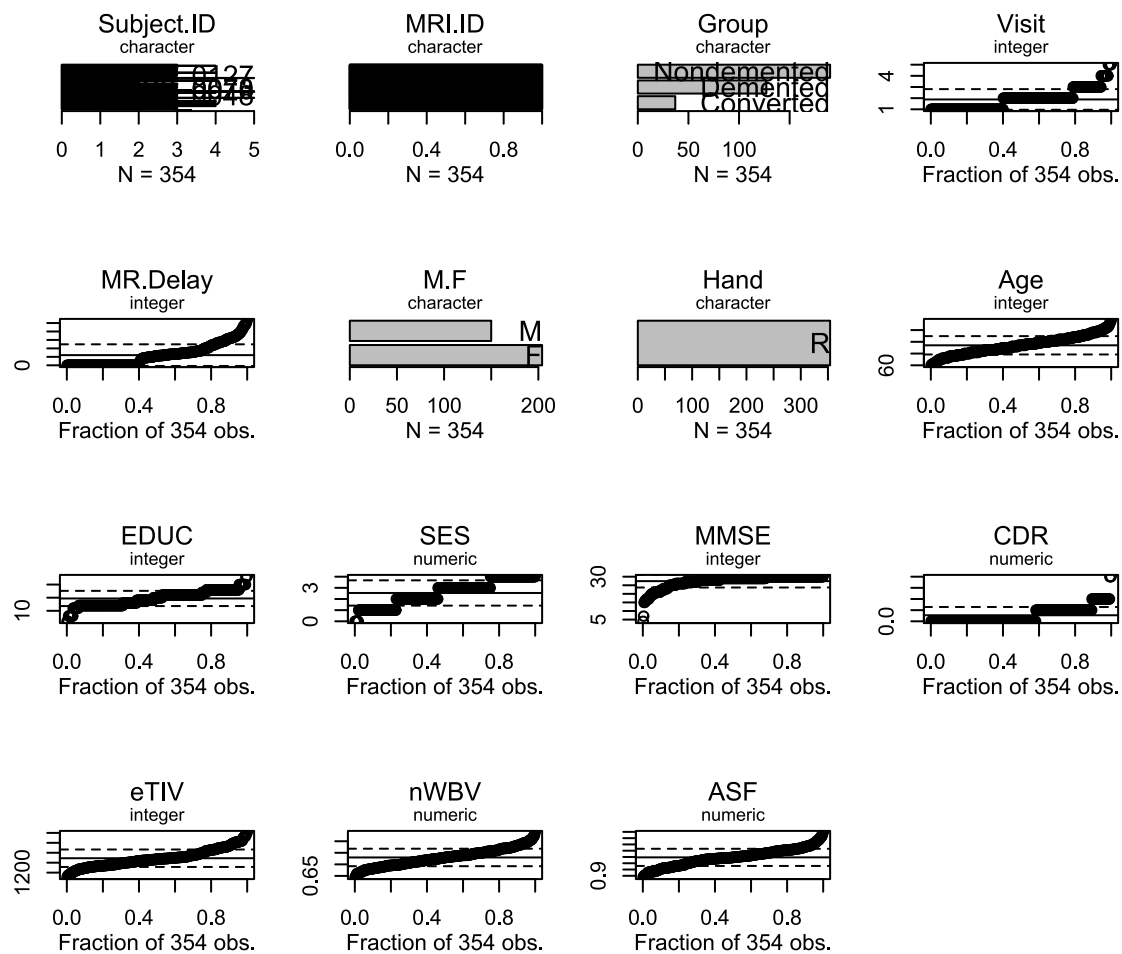
table_per_subject <- table(dc$Subject.ID)
table_per_subject[table_per_subject < 2]
```

```
## named integer(0)
```

```
dc <- dc[order(dc$Subject.ID, dc$Age), ]
```

Data visualization

```
xqplot(dc)
```



```

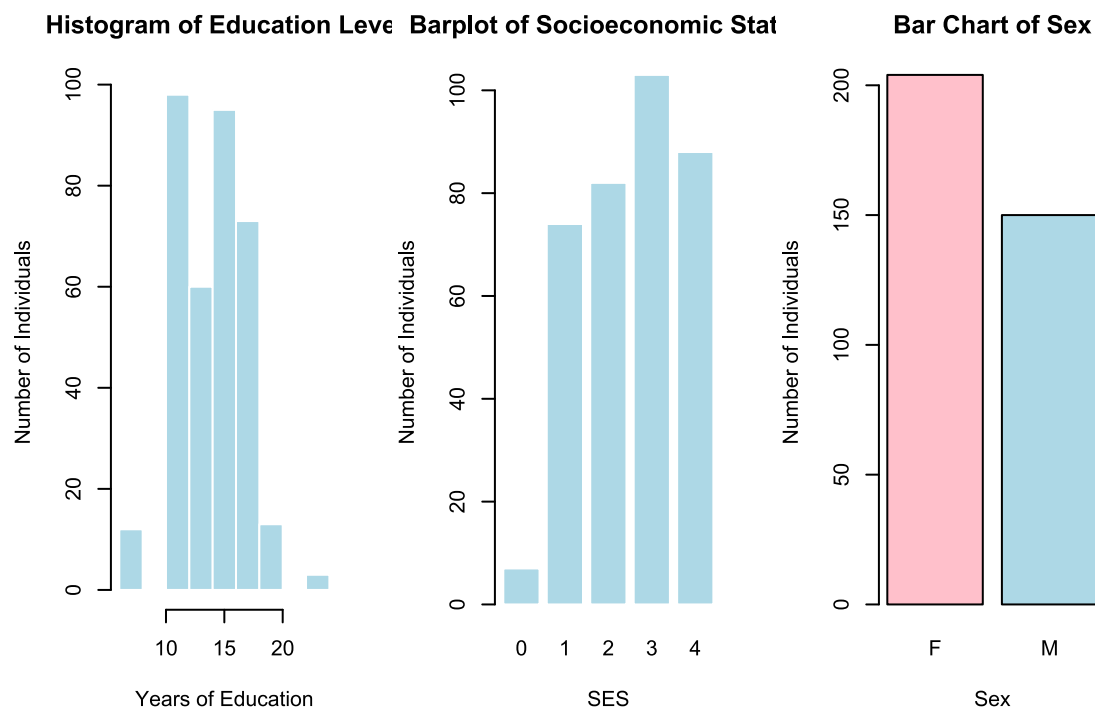
par(mfrow = c(1, 3))

hist(dc$EDUC,
     breaks = 10,
     main = "Histogram of Education Level",
     xlab = "Years of Education",
     ylab = "Number of Individuals",
     col = "lightblue",
     border = "white")

barplot(table(dc$SES),
       main = "Barplot of Socioeconomic Status",
       xlab = "SES",
       ylab = "Number of Individuals",
       col = "lightblue",
       border = "white")

barplot(table(dc$M.F),
       main = "Bar Chart of Sex",
       xlab = "Sex",
       ylab = "Number of Individuals",
       col = c("pink", "lightblue"))

```



```

par(mfrow = c(1, 1))

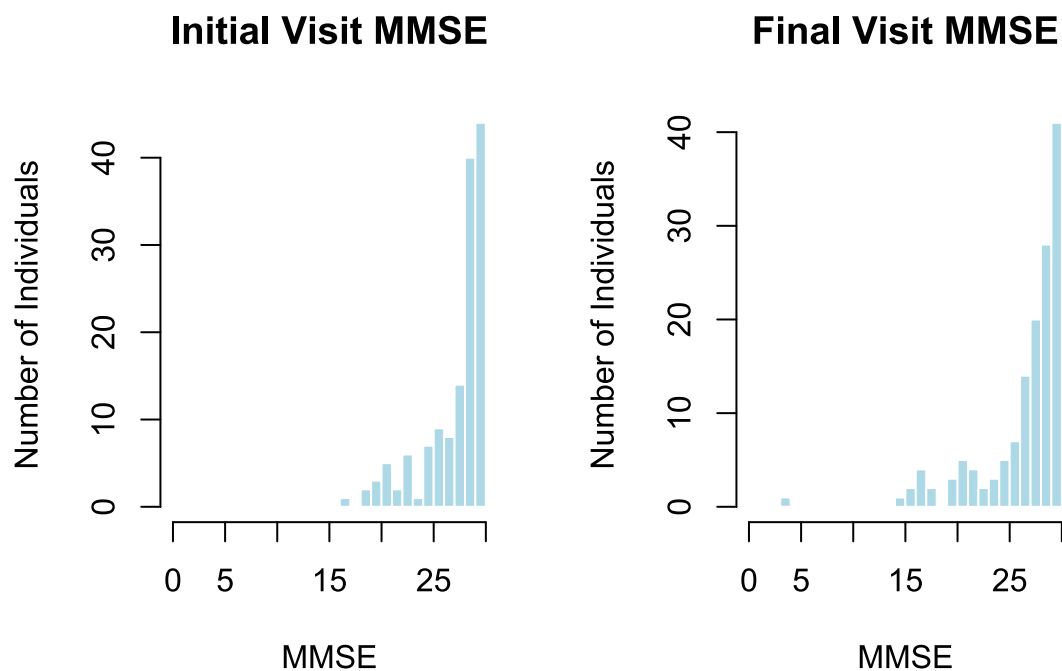
initial_visits <- dc[ave(dc$Visit, dc$Subject.ID, FUN = min) == dc$Visit, ]
final_visits <- dc[ave(dc$Visit, dc$Subject.ID, FUN = max) == dc$Visit, ]

#Set layout: 1 row, 2 columns
par(mfrow = c(1, 2))

#Histogram for initial visits
hist(initial_visits$MMSE,
      main = "Initial Visit MMSE",
      breaks = 0:30,
      xlab = "MMSE",
      ylab = "Number of Individuals",
      col = "lightblue",
      border = "white",
      xlim = c(0, 30))

#Histogram for final visits
hist(final_visits$MMSE,
      main = "Final Visit MMSE",
      breaks = 0:30,
      xlab = "MMSE",
      ylab = "Number of Individuals",
      col = "lightblue",
      border = "white",
      xlim = c(0, 30))

```



```
#Reset layout
par(mfrow = c(1, 1))

#Get initial visits per subject
initial_visits <- dc %>%
  group_by(Subject.ID) %>%
  filter(Visit == min(Visit)) %>%
  ungroup()

#Filter subjects who had CDR = 0 at initial visit
subjects_cdr0 <- initial_visits %>%
  filter(CDR == 0) %>%
  pull(Subject.ID)

#Initial visits for subjects with CDR = 0 at initial visit
initial_cdr0 <- dc %>%
  filter(Subject.ID %in% subjects_cdr0) %>%
  group_by(Subject.ID) %>%
  filter(Visit == min(Visit)) %>%
  ungroup()

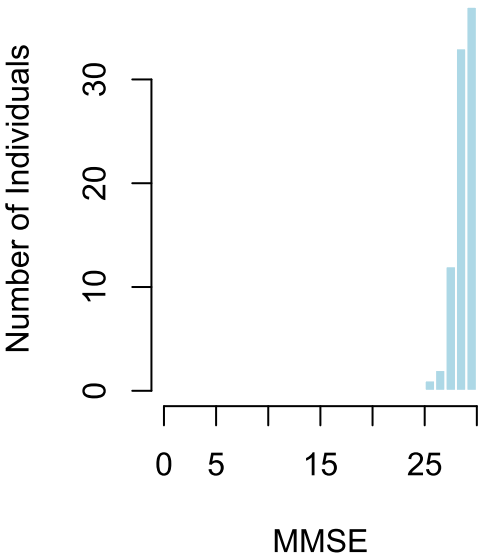
#Final visits for subjects with CDR = 0 at initial visit
final_cdr0 <- dc %>%
  filter(Subject.ID %in% subjects_cdr0) %>%
  group_by(Subject.ID) %>%
  filter(Visit == max(Visit)) %>%
  ungroup()

par(mfrow = c(1, 2))

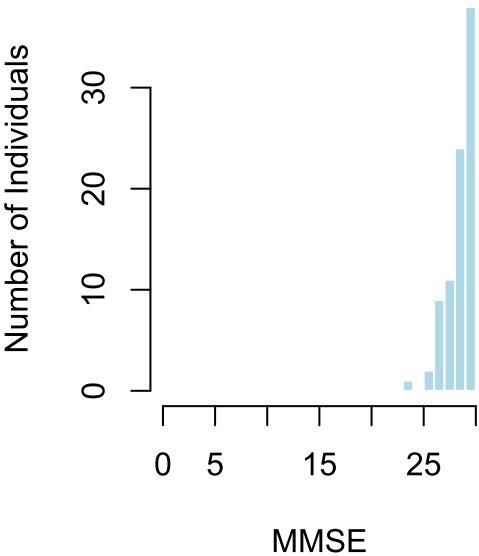
hist(initial_cdr0$MMSE,
      breaks = 0:30,
      main = "Initial Visit MMSE (CDR=0)",
      xlab = "MMSE",
      ylab = "Number of Individuals",
      col = "lightblue",
      border = "white",
      xlim = c(0, 30))

hist(final_cdr0$MMSE,
      breaks = 0:30,
      main = "Final Visit MMSE (CDR=0)",
      xlab = "MMSE",
      ylab = "Number of Individuals",
      col = "lightblue",
      border = "white",
      xlim = c(0, 30))
```

Initial Visit MMSE (CDR=0)



Final Visit MMSE (CDR=0)



```
par(mfrow = c(1, 1))

#Get initial visits per subject
initial_visits <- dc %>%
  group_by(Subject.ID) %>%
  filter(Visit == min(Visit)) %>%
  ungroup()

#Filter subjects who had CDR = 0.5 at initial visit
subjects_cdr05 <- initial_visits %>%
  filter(CDR == 0.5) %>%
  pull(Subject.ID)

#Initial visits for subjects with CDR = 0.5 at initial visit
initial_cdr05 <- dc %>%
  filter(Subject.ID %in% subjects_cdr05) %>%
  group_by(Subject.ID) %>%
  filter(Visit == min(Visit)) %>%
  ungroup()

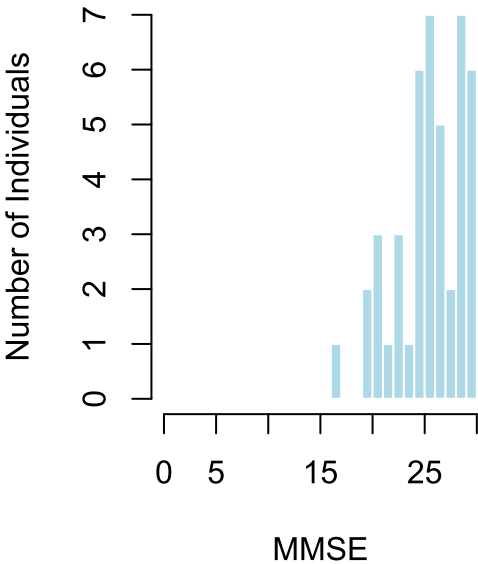
#Final visits for subjects with CDR = 0.5 at initial visit
final_cdr05 <- dc %>%
  filter(Subject.ID %in% subjects_cdr05) %>%
  group_by(Subject.ID) %>%
  filter(Visit == max(Visit)) %>%
  ungroup()

par(mfrow = c(1, 2))

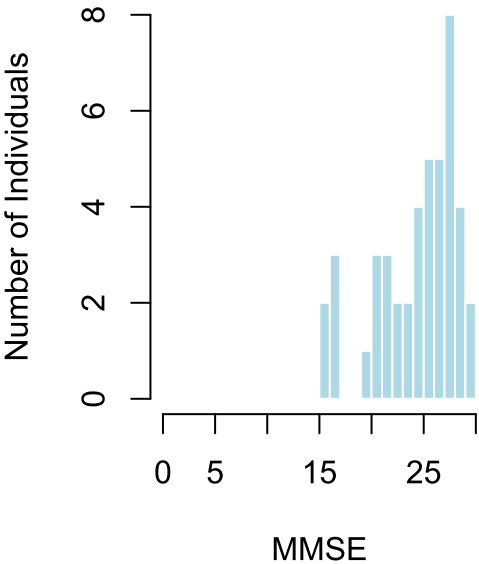
#Histogram for initial visits
hist(initial_cdr05$MMSE,
      breaks = 0:30,
      main = "Initial Visit MMSE (CDR=0.5)",
      xlab = "MMSE",
      ylab = "Number of Individuals",
      col = "lightblue",
      border = "white",
      xlim = c(0, 30))

#Histogram for final visits
hist(final_cdr05$MMSE,
      breaks = 0:30,
      main = "Final Visit MMSE (CDR=0.5)",
      xlab = "MMSE",
      ylab = "Number of Individuals",
      col = "lightblue",
      border = "white",
      xlim = c(0, 30))
```

Initial Visit MMSE (CDR=0.5)



Final Visit MMSE (CDR=0.5)




```
par(mfrow = c(1, 1))

#Get initial visits per subject
initial_visits <- dc %>%
  group_by(Subject.ID) %>%
  filter(Visit == min(Visit)) %>%
  ungroup()

#Filter subjects who had CDR = 1 at initial visit
subjects_cdr1 <- initial_visits %>%
  filter(CDR == 1) %>%
  pull(Subject.ID)

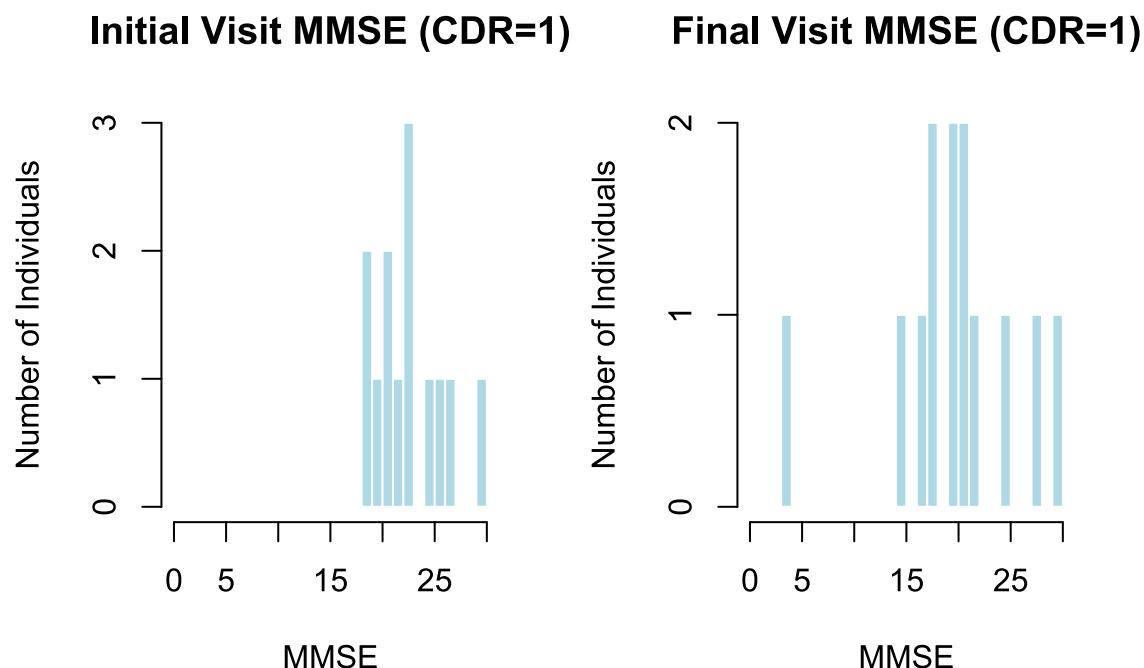
#Initial visits for subjects with CDR = 1 at initial visit
initial_cdr1 <- dc %>%
  filter(Subject.ID %in% subjects_cdr1) %>%
  group_by(Subject.ID) %>%
  filter(Visit == min(Visit)) %>%
  ungroup()

#Final visits for subjects with CDR = 1 at initial visit
final_cdr1 <- dc %>%
  filter(Subject.ID %in% subjects_cdr1) %>%
  group_by(Subject.ID) %>%
  filter(Visit == max(Visit)) %>%
  ungroup()

par(mfrow = c(1, 2))

#Histogram for initial visits
hist(initial_cdr1$MMSE,
      breaks = 0:30,
      main = "Initial Visit MMSE (CDR=1)",
      xlab = "MMSE",
      ylab = "Number of Individuals",
      col = "lightblue",
      border = "white",
      xlim = c(0, 30))

#Histogram for final visits
hist(final_cdr1$MMSE,
      breaks = 0:30,
      main = "Final Visit MMSE (CDR=1)",
      xlab = "MMSE",
      ylab = "Number of Individuals",
      col = "lightblue",
      border = "white",
      xlim = c(0, 30))
```



```
par(mfrow = c(1, 1))
```

We see immediately from the xqplot that MMSE is heavily left-skewed, the data has slightly more females than males, and socioeconomic status tends to the higher end. Histograms and bar plots are attached for ease of visualization, and we can hypothesize from inspecting the MMSE histograms stratified by CDR that the overall decline in MMSE is in majority due to the decline in the groups CDR = 0.5 and CDR = 1.

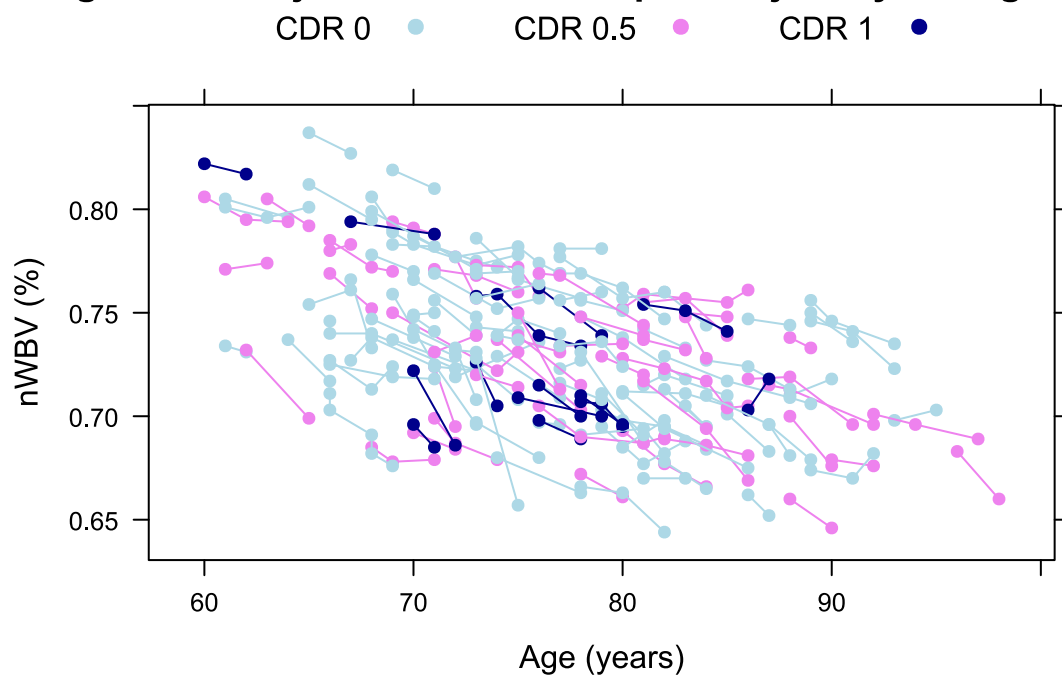
Longitudinal Plots

We recreate the by subject nWBV trajectories originally presented in (Marcus, Fotenos, et al., 2010), as well as the analogous MMSE trajectories. We compute the mean of the slopes of the MMSE trajectories.

```
cdr_colors <- c("0" = "lightblue", "0.5" = "violet", "1" = "darkblue")

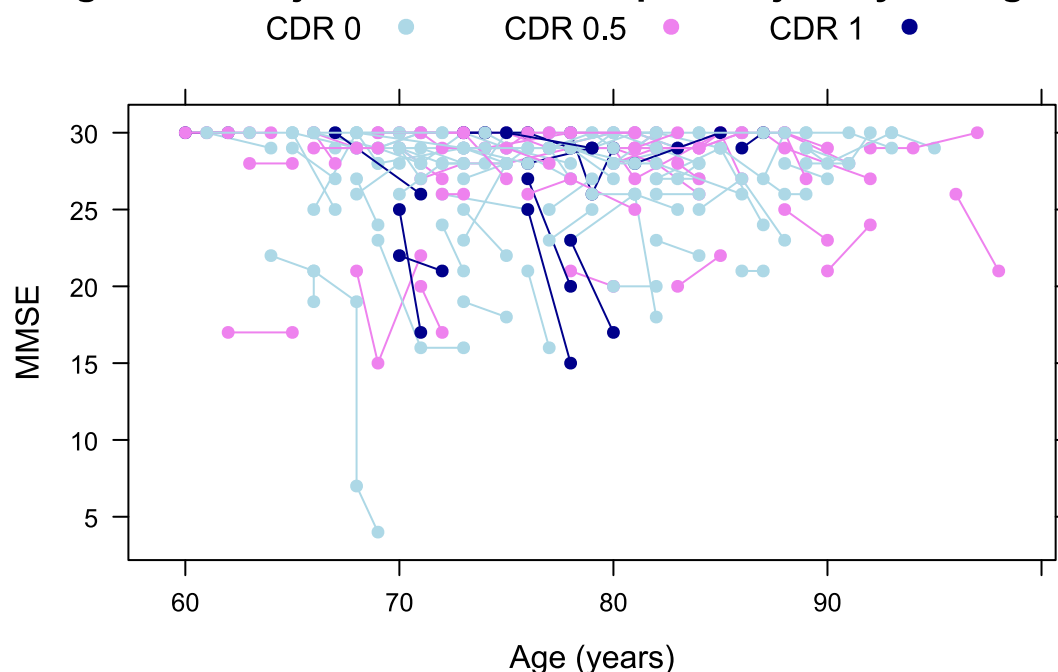
# nWBV trajectories
xyplot(nWBV ~ Age, data = dc,
       groups = Subject.ID,
       type = "b",
       lwd = 1,
       pch = 16,
       col = cdr_colors[as.character(dc$CDR)],
       xlab = "Age (years)",
       ylab = "nWBV (%)",
       main = "Longitudinal trajectories of nWBV per subject by CDR group",
       key = list(text = list(c("CDR 0", "CDR 0.5", "CDR 1")),
                  points = list(pch = 16, col = c("lightblue", "violet", "darkblue")),
                  columns = 3))
```

Longitudinal trajectories of nWBV per subject by CDR group



```
# MMSE trajectories
xyplot(MMSE ~ Age, data = dc,
       groups = Subject.ID,
       type = "b",
       lwd = 1,
       pch = 16,
       col = cdr_colors[as.character(dc$CDR)],
       xlab = "Age (years)",
       ylab = "MMSE",
       main = "Longitudinal trajectories of MMSE per subject by CDR group",
       key = list(text = list(c("CDR 0", "CDR 0.5", "CDR 1")),
                  points = list(pch = 16, col = c("lightblue", "violet", "darkblue")),
                  columns = 3))
```

Longitudinal trajectories of MMSE per subject by CDR group



```
# Copy dc to avoid modifying original
dc_copy <- dc

# Calculate slopes per subject manually
subject_slopes <- dc_copy %>%
  group_by(Subject.ID, CDR) %>%
  filter(n() > 1) %>% # need at least two points per subject
  summarise(
    slope = {
      fit <- lm(MMSE ~ Age, data = cur_data())
      coef(fit)["Age"]
    },
    .groups = "drop"
  )
```

```
## Warning: There was 1 warning in `summarise()`.
## i In argument: `slope = { ... }`.
## i In group 1: `Subject.ID = "OAS2_0001"` `CDR = 0`.
## Caused by warning:
## ! `cur_data()` was deprecated in dplyr 1.1.0.
## i Please use `pick()` instead.
```

```
# Average slopes by CDR group
avg_slopes_by_CDR_old <- subject_slopes %>%
  group_by(CDR) %>%
  summarise(
    avg_slope = mean(slope, na.rm = TRUE),
    n = n()
  )

print(avg_slopes_by_CDR_old)
```

```
## # A tibble: 3 × 3
##   CDR avg_slope    n
##   <dbl>   <dbl> <int>
## 1     0   -0.0306    76
## 2   0.5   -0.464    38
## 3     1   -0.826     9
```

Group CDR = 0 has mean MMSE slope -0.0306, group CDR = 0.5 has mean MMSE slope -0.464, and group CDR = 1 has mean MMSE slope -0.826.

Linear Mixed Effects Modelling

```
# We begin with considering random effects. As a preliminary model, we include each predictor in the fixed effects and
# consider Age as a random effect since it's time-variant.

fit1 <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC,
            data = dc,
            random = ~ 1 + Age | Subject.ID,
            correlation = corAR1(form = ~ 1 | Subject.ID))

summary(fit1)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC logLik
## 1676.46 1718.834 -827.23
##
## Random effects:
## Formula: ~1 + Age | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev      Corr
## (Intercept) 2.142898e+00 (Intr)
## Age         1.322308e-05 0
## Residual    2.474990e+00
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.5880266
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC
##           Value Std.Error DF   t-value p-value
## (Intercept) -17.77207  7.414292 210 -2.397002  0.0174
## nWBV         45.38363  7.185439 210  6.316055  0.0000
## Age          0.11360  0.036311 210  3.128682  0.0020
## M.FM         -0.37804  0.534158 138 -0.707734  0.4803
## SES          0.28756  0.327138 138  0.879032  0.3809
## EDUC         0.17950  0.127845 138  1.404058  0.1625
## Correlation:
##      (Intr) nWBV   Age    M.FM   SES
## nWBV -0.929
## Age  -0.802  0.585
## M.FM -0.271  0.266  0.182
## SES  0.042  0.058 -0.030  0.040
## EDUC -0.145 -0.053  0.035 -0.081 -0.727
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -5.3352764 -0.2424682  0.1432472  0.5320371  1.5676575
##
## Number of Observations: 354
## Number of Groups: 142
```

We will compare it to the same fixed effects but with just a random intercept.

```
test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC,
            data = dc,
            random = ~ 1 | Subject.ID,
            correlation = corAR1(form = ~ 1 | Subject.ID))
summary(test)
```

```

## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC logLik
## 1672.46 1707.13 -827.23
##
## Random effects:
## Formula: ~1 | Subject.ID
##      (Intercept) Residual
## StdDev:      2.142898  2.47499
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.5880268
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC
##              Value Std.Error DF   t-value p-value
## (Intercept) -17.77207  7.414292 210 -2.397002  0.0174
## nWBV          45.38363  7.185439 210  6.316055  0.0000
## Age           0.11360  0.036311 210  3.128682  0.0020
## M.FM          -0.37804  0.534158 138 -0.707734  0.4803
## SES           0.28756  0.327138 138  0.879032  0.3809
## EDUC          0.17950  0.127845 138  1.404058  0.1625
## Correlation:
##      (Intr) nWBV   Age    M.FM   SES
## nWBV -0.929
## Age  -0.802  0.585
## M.FM -0.271  0.266  0.182
## SES  0.042  0.058 -0.030  0.040
## EDUC -0.145 -0.053  0.035 -0.081 -0.727
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -5.3352766 -0.2424683  0.1432472  0.5320371  1.5676574
##
## Number of Observations: 354
## Number of Groups: 142

```

```
anova(fit1, test)
```

```

##      Model df      AIC      BIC logLik Test      L.Ratio p-value
## fit1      1 11 1676.46 1718.834 -827.23
## test      2  9 1672.46 1707.130 -827.23 1 vs 2 3.833557e-07      1

```

```
# Since the AIC is lower and p-value is high, we proceed without the random effect of Age.

fit1 <- update(fit1, random = ~ 1 | Subject.ID)

# nWBV is also time-variant, so we will test it as a random effect.

test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC,
            data = dc,
            random = ~ 1 + nWBV | Subject.ID,
            correlation = corAR1(form = ~ 1 | Subject.ID),
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```



```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1632.026 1674.4 -805.0128
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 30.125603 (Intr)
## nWBV        38.256604 -1
## Residual    1.882025
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3503827
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC
##           Value Std.Error DF   t-value p-value
## (Intercept) 0.294103 7.157217 210 0.041092 0.9673
## nWBV        29.823705 7.116504 210 4.190781 0.0000
## Age         0.043803 0.031988 210 1.369333 0.1724
## M.FM        -0.666225 0.442148 138 -1.506791 0.1342
## SES         0.093875 0.262842 138 0.357152 0.7215
## EDUC        0.138319 0.104192 138 1.327536 0.1865
## Correlation:
##      (Intr) nWBV  Age    M.FM  SES
## nWBV -0.947
## Age  -0.786 0.597
## M.FM -0.223 0.232 0.148
## SES  -0.027 0.103 0.017 0.093
## EDUC -0.097 -0.072 0.010 -0.138 -0.696
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.9190676 -0.2946575 0.1071537 0.4679001 2.9629229
##
## Number of Observations: 354
## Number of Groups: 142
```

We see that the AIC is much lower and p-value is very low, so we proceed by including the random effect of nWBV.

```
fit2 <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(fit2)
```

```

## Linear mixed-effects model fit by REML
##   Data: dc
##           AIC      BIC    logLik
##  1632.026 1674.4 -805.0128
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev    Corr
## (Intercept) 30.125603 (Intr)
## nWBV        38.256604 -1
## Residual    1.882025
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##           Phi
## 0.3503827
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC
##           Value Std.Error  DF   t-value p-value
## (Intercept) 0.294103  7.157217 210  0.041092  0.9673
## nWBV        29.823705  7.116504 210  4.190781  0.0000
## Age         0.043803  0.031988 210  1.369333  0.1724
## M.FM        -0.666225  0.442148 138 -1.506791  0.1342
## SES         0.093875  0.262842 138  0.357152  0.7215
## EDUC        0.138319  0.104192 138  1.327536  0.1865
## Correlation:
##           (Intr) nWBV  Age    M.FM  SES
## nWBV -0.947
## Age  -0.786  0.597
## M.FM -0.223  0.232  0.148
## SES  -0.027  0.103  0.017  0.093
## EDUC -0.097 -0.072  0.010 -0.138 -0.696
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.9190676 -0.2946575  0.1071537  0.4679001  2.9629229
##
## Number of Observations: 354
## Number of Groups: 142

```

Now we consider interaction terms. We consider three possible 3-way Interaction terms:

*# 1) nWBV * Age * SES: SES influences access to healthcare, slowing decay in nWBV as the subjects ages.*

*# 2) nWBV * Age * M.F: Possible differences in nWBV decay by sex as the subject ages.*

*# 3) nWBV * Age * EDUC: Higher education slows the loss of brain volume as the subject ages.*

Case 1)

```
test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age * SES
           + nWBV * Age + nWBV * SES + Age * SES,
           data = dc,
           random = ~ 1 + nWBV | Subject.ID,
           correlation = corAR1(form = ~ 1 | Subject.ID),
           control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```

## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1679.897 1737.507 -824.9487
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 0.1954965 (Intr)
## nWBV         0.2027344 -0.371
## Residual     3.2745790
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.7649814
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age * SES + nWBV *      Age + n
WBV * SES + Age * SES
##           Value Std.Error  DF    t-value p-value
## (Intercept) -109.64519  93.31457 206 -1.1750061  0.2413
## nWBV         159.75950 127.16399 206  1.2563266  0.2104
## Age          1.02400   1.22193 206  0.8380188  0.4030
## M.FM         -0.50329   0.53906 138 -0.9336385  0.3521
## SES          29.71438  32.38970 138  0.9174020  0.3605
## EDUC          0.20386   0.12943 138  1.5750715  0.1175
## nWBV:Age     -1.10089   1.68083 206 -0.6549678  0.5132
## nWBV:SES     -36.11864  44.43754 206 -0.8127956  0.4173
## Age:SES      -0.26514   0.42249 206 -0.6275692  0.5310
## nWBV:Age:SES  0.30722   0.58460 206  0.5255143  0.5998
## Correlation:
##           (Intr) nWBV   Age    M.FM   SES    EDUC   nWBV:Ag nWBV:S Ag:SES
## nWBV         -0.997
## Age          -0.991  0.994
## M.FM          0.115 -0.120 -0.126
## SES          -0.919  0.920  0.910 -0.153
## EDUC          -0.090  0.065  0.077 -0.091  0.060
## nWBV:Age      0.982 -0.991 -0.997  0.129 -0.905 -0.066
## nWBV:SES      0.913 -0.920 -0.910  0.151 -0.997 -0.057  0.911
## Age:SES       0.912 -0.918 -0.920  0.148 -0.990 -0.066  0.921  0.993
## nWBV:Age:SES -0.900  0.912  0.914 -0.145  0.981  0.055 -0.921 -0.990 -0.997
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -6.6203240 -0.2772330  0.2388486  0.6347886  1.7999549
##
## Number of Observations: 354
## Number of Groups: 142

```

Switch to ML so we can use ANOVA:

```
test_ml <- update(test, method = "ML")
fit2_ml <- update(fit2, method = "ML")

anova(test_ml, fit2_ml)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	test_ml	1 15	1621.839	1679.879	-795.9196			
##	fit2_ml	2 11	1625.603	1668.166	-801.8017	1 vs 2	11.76423	0.0192

The test model as a higher AIC with low p-value, so it appears to perform better. But it's worth noting that all

coefficients have p-values > 0.05. We will give this model a proper name and return to it later.

```
fit3 <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age * SES
           + nWBV * Age + nWBV * SES + Age * SES,
           data = dc,
           random = ~ 1 + nWBV | Subject.ID,
           correlation = corAR1(form = ~ 1 | Subject.ID),
           control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(fit3)
```

```

## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1679.897 1737.507 -824.9487
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 0.1954965 (Intr)
## nWBV         0.2027344 -0.371
## Residual     3.2745790
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.7649814
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age * SES + nWBV *      Age + n
WBV * SES + Age * SES
##           Value Std.Error DF    t-value p-value
## (Intercept) -109.64519  93.31457 206 -1.1750061  0.2413
## nWBV         159.75950 127.16399 206  1.2563266  0.2104
## Age          1.02400   1.22193 206  0.8380188  0.4030
## M.FM         -0.50329   0.53906 138 -0.9336385  0.3521
## SES          29.71438  32.38970 138  0.9174020  0.3605
## EDUC          0.20386   0.12943 138  1.5750715  0.1175
## nWBV:Age     -1.10089   1.68083 206 -0.6549678  0.5132
## nWBV:SES     -36.11864  44.43754 206 -0.8127956  0.4173
## Age:SES      -0.26514   0.42249 206 -0.6275692  0.5310
## nWBV:Age:SES  0.30722   0.58460 206  0.5255143  0.5998
## Correlation:
##           (Intr) nWBV   Age    M.FM   SES    EDUC   nWBV:Ag nWBV:S Ag:SES
## nWBV       -0.997
## Age        -0.991  0.994
## M.FM        0.115 -0.120 -0.126
## SES        -0.919  0.920  0.910 -0.153
## EDUC       -0.090  0.065  0.077 -0.091  0.060
## nWBV:Age    0.982 -0.991 -0.997  0.129 -0.905 -0.066
## nWBV:SES    0.913 -0.920 -0.910  0.151 -0.997 -0.057  0.911
## Age:SES     0.912 -0.918 -0.920  0.148 -0.990 -0.066  0.921  0.993
## nWBV:Age:SES -0.900  0.912  0.914 -0.145  0.981  0.055 -0.921 -0.990 -0.997
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -6.6203240 -0.2772330  0.2388486  0.6347886  1.7999549
##
## Number of Observations: 354
## Number of Groups: 142

```

Case 2)

```
test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age * SES
            + nWBV * Age + nWBV * SES + Age * SES + nWBV * Age * M.F
            + nWBV * M.F + Age * M.F,
            data = dc,
            random = ~ 1 + nWBV | Subject.ID,
            correlation = corAR1(form = ~ 1 | Subject.ID),
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```

## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1625.132 1694.106 -794.5662
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 31.814181 (Intr)
## nWBV         40.715213 -1
## Residual     1.875835
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3417534
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age * SES + nWBV * Age + n
WBV * SES + Age * SES + nWBV * Age * M.F + nWBV * M.F + Age * M.F
##           Value Std.Error  DF    t-value p-value
## (Intercept) -44.17196 107.61065 203 -0.4104795 0.6819
## nWBV         71.82271 144.05676 203 0.4985723 0.6186
## Age          0.16333 1.41441 203 0.1154779 0.9082
## M.FM        -85.11333 78.95232 138 -1.0780346 0.2829
## SES          30.88150 33.03642 138 0.9347712 0.3515
## EDUC         0.16656 0.10373 138 1.6057795 0.1106
## nWBV:Age      0.06826 1.90198 203 0.0358899 0.9714
## nWBV:SES     -33.51906 44.30109 203 -0.7566192 0.4502
## Age:SES      -0.20361 0.43204 203 -0.4712662 0.6380
## nWBV:M.FM    106.33181 105.87054 203 1.0043570 0.3164
## Age:M.FM      1.02124 1.03661 203 0.9851756 0.3257
## nWBV:Age:SES 0.16965 0.58367 203 0.2906663 0.7716
## nWBV:Age:M.FM -1.27994 1.39956 203 -0.9145332 0.3615
## Correlation:
##           (Intr) nWBV   Age    M.FM   SES    EDUC   nWBV:Ag nWBV:S Ag:SES
## nWBV        -0.998
## Age         -0.992 0.994
## M.FM        -0.525 0.530 0.518
## SES         -0.898 0.898 0.890 0.281
## EDUC         0.101 -0.116 -0.101 -0.223 -0.060
## nWBV:Age     0.986 -0.992 -0.998 -0.519 -0.888 0.106
## nWBV:SES     0.896 -0.901 -0.893 -0.288 -0.998 0.063 0.894
## Age:SES      0.891 -0.895 -0.898 -0.273 -0.989 0.050 0.900 0.992
## nWBV:M.FM    0.527 -0.533 -0.520 -0.998 -0.288 0.220 0.523 0.296 0.280
## Age:M.FM     0.519 -0.525 -0.516 -0.991 -0.279 0.213 0.518 0.286 0.269
## nWBV:Age:SES -0.885 0.893 0.896 0.280 0.982 -0.058 -0.901 -0.989 -0.998
## nWBV:Age:M.FM -0.518 0.525 0.515 0.984 0.285 -0.211 -0.518 -0.291 -0.273
##           nWBV:M A:M.FM nWBV:A:S
## nWBV
## Age
## M.FM

```



```
## SES
## EDUC
## nWBV:Age
## nWBV:SES
## Age:SES
## nWBV:M.FM
## Age:M.FM      0.993
## nWBV:Age:SES  -0.286 -0.274
## nWBV:Age:M.FM -0.991 -0.998  0.279
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.9236543 -0.3069357  0.1510325  0.4549005  2.8506166
##
## Number of Observations: 354
## Number of Groups: 142
```

```
# Switch to ML so we can use ANOVA:
test_ml <- update(test, method = "ML")
fit3_ml <- update(fit3, method = "ML")

anova(test_ml, fit3_ml)
```

```
##      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
## test_ml    1 18 1625.441 1695.088 -794.7205
## fit3_ml    2 15 1621.839 1679.879 -795.9196 1 vs 2 2.398306    0.494
```

The AIC for this model is higher than our previous interaction model, and the p-value is quite high. We will discard this model # and consider the next case.

Case 3)

```
test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age * SES
  + nWBV * Age + nWBV * SES + Age * SES + nWBV * Age * EDUC
  + nWBV * EDUC + Age * EDUC,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```

## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1634.656 1703.63 -799.3278
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 32.223200 (Intr)
## nWBV        41.190244 -1
## Residual    1.874473
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3362815
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age * SES + nWBV * Age + nWBV * SES + Age * SES + nWBV * Age * EDUC + nWBV * EDUC + Age * EDUC
##           Value Std.Error DF   t-value p-value
## (Intercept)  -0.89866 203.38619 203 -0.0044185 0.9965
## nWBV          7.66569 271.75408 203 0.0282082 0.9775
## Age          -0.55584 2.61503 203 -0.2125551 0.8319
## M.FM         -0.62174 0.44078 138 -1.4105286 0.1606
## SES          60.91629 46.37888 138 1.3134488 0.1912
## EDUC        -10.96125 18.18232 138 -0.6028521 0.5476
## nWBV:Age      1.12066 3.52603 203 0.3178252 0.7509
## nWBV:SES     -74.31345 61.82199 203 -1.2020553 0.2307
## Age:SES      -0.63022 0.62037 203 -1.0158721 0.3109
## nWBV:EDUC    15.23772 24.31031 203 0.6268007 0.5315
## Age:EDUC      0.15972 0.23799 203 0.6711189 0.5029
## nWBV:Age:SES  0.75002 0.83332 203 0.9000381 0.3692
## nWBV:Age:EDUC -0.21923 0.32110 203 -0.6827583 0.4955
## Correlation:
##           (Intr) nWBV   Age    M.FM    SES    EDUC    nWBV:Ag nWBV:S Ag:SES
## nWBV      -0.998
## Age       -0.988 0.992
## M.FM       0.236 -0.236 -0.232
## SES        0.354 -0.356 -0.362 0.019
## EDUC      -0.889 0.889 0.885 -0.198 -0.727
## nWBV:Age   0.979 -0.988 -0.998 0.230 0.362 -0.879
## nWBV:SES  -0.353 0.356 0.363 -0.011 -0.998 0.725 -0.364
## Age:SES   -0.342 0.346 0.359 -0.006 -0.989 0.715 -0.361 0.993
## nWBV:EDUC  0.886 -0.890 -0.888 0.192 0.725 -0.998 0.886 -0.726 -0.718
## Age:EDUC  0.864 -0.870 -0.882 0.181 0.735 -0.988 0.883 -0.737 -0.739
## nWBV:Age:SES 0.339 -0.345 -0.358 -0.002 0.982 -0.709 0.362 -0.989 -0.998
## nWBV:Age:EDUC -0.856 0.866 0.880 -0.175 -0.729 0.979 -0.884 0.734 0.738
##           nWBV:E A:EDUC nWBV:A:S
## nWBV
## Age
## M.FM

```

```
## SES
## EDUC
## nWBV:Age
## nWBV:SES
## Age:SES
## nWBV:EDUC
## Age:EDUC      0.992
## nWBV:Age:SES  0.715  0.738
## nWBV:Age:EDUC -0.988 -0.998 -0.739
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.8954263 -0.2905445  0.1290028  0.4592727  2.7946119
##
## Number of Observations: 354
## Number of Groups: 142
```

```
# Switch to ML so we can use ANOVA:
test_ml <- update(test, method = "ML")

anova(test_ml, fit3_ml)
```

```
##      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
## test_ml    1 18 1626.827 1696.475 -795.4137
## fit3_ml    2 15 1621.839 1679.879 -795.9196 1 vs 2 1.011909  0.7984
```

```

# We again see a higher AIC and high p-value, so we discard this new model.

# Although our first interaction model performs relatively well in terms of AIC, as noted earlier, all coefficients are
# statistically insignificant (p-values > 0.05.) We will discard this model and instead add 2-way interaction terms to
# our original model one by one. We consider the following cases:

# 1) nWBV * Age : Brain volume changes with age.

# 2) nWBV * EDUC: Education acts as a buffer against brain volume loss.

# 3) nWBV * SES : Higher SES allows for better healthcare, improving preservation of brain volume.

# 4) nWBV * M.F : Sex affects brain volume.

# 5) Age * EDUC : Education acts as a buffer against age effects.

# 6) Age * M.F : Aging affects women and men differently.

# 7) Age * SES : Higher SES acts as a buffer against age effects by again allowing for better healthcare.

# Case 1)

dc$nWBV <- scale(dc$nWBV, center = TRUE, scale = FALSE)

test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)

```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1632.058 1678.25 -804.0289
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.258129 (Intr)
## nWBV        38.137276 -0.998
## Residual    1.928601
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3772774
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age
##           Value Std.Error DF   t-value p-value
## (Intercept) 22.13681   2.90423 209   7.622267  0.0000
## nWBV        35.66991  45.55343 209   0.783035  0.4345
## Age         0.04323   0.03349 209   1.290798  0.1982
## M.FM       -0.68406   0.43496 138  -1.572699  0.1181
## SES         0.06860   0.25805 138   0.265859  0.7907
## EDUC        0.14075   0.10239 138   1.374615  0.1715
## nWBV:Age    -0.08491   0.60439 209  -0.140492  0.8884
## Correlation:
##      (Intr) nWBV   Age    M.FM   SES    EDUC
## nWBV      -0.387
## Age      -0.916  0.401
## M.FM     -0.147  0.081  0.152
## SES       0.113  0.023  0.019  0.090
## EDUC     -0.363 -0.008  0.010 -0.131 -0.693
## nWBV:Age  0.305 -0.988 -0.317 -0.046 -0.007 -0.003
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.8465014 -0.2815776  0.1228742  0.4816521  2.8951217
##
## Number of Observations: 354
## Number of Groups: 142
```

We will compare this model to the same model with centered age in the random effects.

```
test2 <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age,
  data = dc,
  random = ~ 1 + dvar(nWBV, Subject.ID) | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test2)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1650.32 1696.512 -813.1601
##
## Random effects:
## Formula: ~1 + dvar(nWBV, Subject.ID) | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##
##              StdDev   Corr
## (Intercept)    2.744041 (Intr)
## dvar(nWBV, Subject.ID) 77.841761 -0.724
## Residual        1.582905
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.1776536
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * Age
##
##              Value Std.Error DF   t-value p-value
## (Intercept) 17.07768   2.97717 209   5.736218  0.0000
## nWBV         48.53674  47.71518 209   1.017218  0.3102
## Age          0.09352   0.03381 209   2.766037  0.0062
## M.FM        -0.38823   0.49634 138  -0.782185  0.4354
## SES          0.32460   0.30185 138   1.075346  0.2841
## EDUC         0.16333   0.11848 138   1.378536  0.1703
## nWBV:Age     -0.13075   0.63029 209  -0.207445  0.8359
## Correlation:
##      (Intr) nWBV   Age    M.FM   SES    EDUC
## nWBV      -0.156
## Age       -0.898  0.139
## M.FM      -0.195  0.075  0.193
## SES        0.196  0.013 -0.037  0.044
## EDUC       -0.421 -0.006  0.035 -0.096 -0.727
## nWBV:Age   0.079 -0.988 -0.051 -0.033 -0.004 -0.004
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.8973923 -0.2845950  0.0766139  0.3549071  3.2910612
##
## Number of Observations: 354
## Number of Groups: 142
```

```
anova(test, test2)
```

```
##      Model df      AIC      BIC    logLik
## test      1 12 1632.058 1678.250 -804.0289
## test2     2 12 1650.320 1696.512 -813.1601
```

```
# By AIC comparison, centering appears to worsen the model performance.
```

```
# switch to ML so we can use ANOVA:
```

```
ml_test <- update(test, method = "ML")
ml_fit2 <- update(fit2, method = "ML")
anova(ml_fit2, ml_test)
```

```
##           Model df      AIC      BIC    logLik  Test   L.Ratio p-value
## ml_fit2      1 11 1625.466 1668.028 -801.7328
## ml_test      2 12 1627.382 1673.814 -801.6911 1 vs 2 0.08334778 0.7728
```

```
# The p-value is high and the AIC increases comparing this model to our original model; we discard this interaction and proceed with the next one.
```

```
# Case 2)
```

```
test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * EDUC,
            data = dc,
            random = ~ 1 + nWBV | Subject.ID,
            correlation = corAR1(form = ~ 1 | Subject.ID),
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1627.14 1673.332 -801.5698
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.235843 (Intr)
## nWBV         37.637741 -0.998
## Residual     1.921957
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3727465
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * EDUC
##           Value Std.Error DF   t-value p-value
## (Intercept) 21.33381  2.803450 209   7.609842  0.0000
## nWBV         76.67934 30.664831 209   2.500563  0.0132
## Age          0.04213  0.031552 209   1.335127  0.1833
## M.FM        -0.74038  0.432846 138  -1.710503  0.0894
## SES          0.05012  0.256532 138   0.195394  0.8454
## EDUC         0.20641  0.109873 138   1.878623  0.0624
## nWBV:EDUC    -3.23246  2.043979 209  -1.581453  0.1153
## Correlation:
##      (Intr) nWBV   Age   M.FM   SES   EDUC
## nWBV      -0.324
## Age       -0.888  0.135
## M.FM      -0.121 -0.028  0.145
## SES        0.128 -0.024  0.018  0.094
## EDUC      -0.420  0.354  0.008 -0.152 -0.660
## nWBV:EDUC  0.200 -0.974  0.002  0.082  0.050 -0.380
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.9026426 -0.2673011  0.1522762  0.4732399  2.9009588
##
## Number of Observations: 354
## Number of Groups: 142
```

Switch to ML so we can use ANOVA:

```
ml_test <- update(test, method = "ML")
anova(ml_fit2, ml_test)
```


##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	ml_fit2	1 11	1625.466	1668.028	-801.7328			
##	ml_test	2 12	1624.777	1671.209	-800.3886	1 vs 2	2.688471	0.1011

There is very slight improvement in AIC but the p-value is high; we discard this interaction and consider the next one.

Case 3)

```
test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1623.105 1669.297 -799.5523
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.245722 (Intr)
## nWBV         38.073288 -0.998
## Residual     1.898225
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3594869
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES
##           Value Std.Error DF   t-value p-value
## (Intercept) 22.07925  2.729954 209   8.087774  0.0000
## nWBV         56.44612 14.318230 209   3.942255  0.0001
## Age          0.03952  0.031343 209   1.260746  0.2088
## M.FM        -0.73028  0.427867 138  -1.706787  0.0901
## SES          0.33657  0.283022 138   1.189205  0.2364
## EDUC         0.11852  0.101219 138   1.170898  0.2437
## nWBV:SES    -11.09394  5.090170 209  -2.179484  0.0304
## Correlation:
##      (Intr) nWBV   Age    M.FM   SES    EDUC
## nWBV      -0.321
## Age       -0.905  0.277
## M.FM      -0.136  0.072  0.145
## SES        0.087  0.434  0.008  0.061
## EDUC      -0.372 -0.121  0.011 -0.126 -0.662
## nWBV:SES   0.046 -0.873  0.020  0.045 -0.443  0.099
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -4.0002839 -0.2861112  0.1377110  0.4862570  2.9455455
##
## Number of Observations: 354
## Number of Groups: 142
```

```
# Switch to ML so we can use ANOVA:
```

```
ml_test <- update(test, method = "ML")
anova(ml_fit2, ml_test)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	ml_fit2	1 11	1625.466	1668.028	-801.7328			
##	ml_test	2 12	1622.468	1668.899	-799.2340	1 vs 2	4.997608	0.0254

We see an improvement in AIC with a low p-value, so we will keep this interaction. We give this model a proper name and consider the next interaction.

```
fit4 <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
```

Case 4)

```
test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES + nWBV * M.F,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1618.263 1668.267 -796.1315
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.234140 (Intr)
## nWBV         38.194751 -0.999
## Residual     1.907077
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3628227
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES + nWBV * M.F
##           Value Std.Error DF   t-value p-value
## (Intercept) 21.91989  2.758595 208   7.946035  0.0000
## nWBV         58.43954 15.511528 208   3.767491  0.0002
## Age          0.04134  0.031644 208   1.306232  0.1929
## M.FM        -0.64049  0.509738 138  -1.256506  0.2111
## SES          0.34137  0.283163 138   1.205570  0.2300
## EDUC         0.11572  0.101598 138   1.139024  0.2567
## nWBV:SES     -11.19309  5.106327 208  -2.192005  0.0295
## nWBV:M.FM    -3.61510 11.436460 208  -0.316103  0.7522
## Correlation:
##           (Intr) nWBV   Age   M.FM   SES   EDUC   nWBV:S
## nWBV        -0.348
## Age        -0.907  0.305
## M.FM       -0.191  0.263  0.196
## SES         0.080  0.416  0.013  0.075
## EDUC       -0.355 -0.143 -0.001 -0.152 -0.663
## nWBV:SES    0.052 -0.825  0.013  0.011 -0.442  0.102
## nWBV:M.FM   0.144 -0.381 -0.139 -0.543 -0.042  0.082  0.048
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.9954125 -0.2847773  0.1361222  0.4802547  2.9275793
##
## Number of Observations: 354
## Number of Groups: 142
```

```
# Switch to ML so we can use ANOVA:
```

```
ml_test <- update(test, method = "ML")
ml_fit4 <- update(fit4, method = "ML")
anova(ml_fit4, ml_test)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	ml_fit4	1 12	1622.468	1668.899	-799.2340			
##	ml_test	2 13	1624.373	1674.674	-799.1864	1 vs 2	0.09516363	0.7577

There is minimal improvement in AIC with a high p-value. We discard this interaction and proceed with the next one.

Case 5)

```
test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES + Age * EDUC,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```

## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1630.289 1680.292 -802.1444
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.210867 (Intr)
## nWBV         39.249905 -0.998
## Residual     1.895464
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3566176
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES + Age * EDUC
##           Value Std.Error DF   t-value p-value
## (Intercept)  5.41231 11.301323 208   0.478909  0.6325
## nWBV         64.76430 15.311973 208   4.229651  0.0000
## Age          0.25792  0.147047 208   1.753991  0.0809
## M.FM        -0.71275  0.424297 138  -1.679837  0.0953
## SES          0.39576  0.283159 138   1.397671  0.1645
## EDUC         1.24566  0.750252 138   1.660319  0.0991
## nWBV:SES     -14.43882  5.538408 208  -2.607034  0.0098
## Age:EDUC     -0.01492  0.009834 208  -1.517331  0.1307
## Correlation:
##           (Intr) nWBV   Age    M.FM   SES    EDUC   nWBV:S
## nWBV       -0.403
## Age        -0.995  0.388
## M.FM       -0.045  0.076  0.044
## SES        -0.103  0.451  0.125  0.069
## EDUC       -0.974  0.323  0.969 -0.004  0.037
## nWBV:SES    0.379 -0.890 -0.367  0.032 -0.457 -0.364
## Age:EDUC    0.971 -0.341 -0.977 -0.014 -0.126 -0.991  0.379
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.9727841 -0.3093041  0.1325530  0.4856013  2.9055208
##
## Number of Observations: 354
## Number of Groups: 142

```

```
# Switch to ML so we can use ANOVA:
```

```
ml_test <- update(test, method = "ML")
anova(ml_fit4, ml_test)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	ml_fit4	1 12	1622.468	1668.899	-799.2340			
##	ml_test	2 13	1622.148	1672.448	-798.0738	1 vs 2	2.320445	0.1277

Again, there is minimal improvement in AIC with a high p-value. We discard this interaction and proceed with the next one.

#Case 6)

```
test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES + Age * M.F,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```

## Linear mixed-effects model fit by REML
##   Data: dc
##           AIC      BIC    logLik
##  1627.686 1677.69 -800.8431
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev    Corr
## (Intercept)  2.191887 (Intr)
## nWBV         37.465819 -0.999
## Residual     1.926228
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##           Phi
## 0.3756498
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES + Age * M.F
##           Value Std.Error  DF   t-value p-value
## (Intercept) 23.40150  3.014104 208  7.763999  0.0000
## nWBV         59.67923 14.474649 208  4.123017  0.0001
## Age          0.02138  0.035801 208  0.597183  0.5510
## M.FM        -5.53193  4.096055 138 -1.350550  0.1791
## SES          0.34278  0.281684 138  1.216903  0.2257
## EDUC         0.11801  0.101378 138  1.164083  0.2464
## nWBV:SES     -12.22701  5.174199 208 -2.363073  0.0190
## Age:M.FM      0.06499  0.055023 208  1.181069  0.2389
## Correlation:
##           (Intr) nWBV  Age    M.FM  SES    EDUC  nWBV:S
## nWBV       -0.210
## Age        -0.923  0.155
## M.FM       -0.437 -0.158  0.496
## SES         0.088  0.417  0.000 -0.004
## EDUC       -0.340 -0.118  0.012 -0.011 -0.666
## nWBV:SES   -0.045 -0.877  0.111  0.194 -0.423  0.096
## Age:M.FM   0.426  0.167 -0.485 -0.995  0.011 -0.003 -0.190
##
## Standardized Within-Group Residuals:
##           Min      Q1      Med      Q3      Max
## -4.0247421 -0.2983120  0.1319903  0.4828876  2.9435176
##
## Number of Observations: 354
## Number of Groups: 142

```

```
# Switch to ML so we can use ANOVA:
```

```
ml_test <- update(test, method = "ML")
anova(ml_fit4, ml_test)
```


##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	ml_fit4	1 12	1622.468	1668.899	-799.234			
##	ml_test	2 13	1623.126	1673.427	-798.563	1 vs 2	1.341973	0.2467

Once again, there is minimal improvement in AIC with a high p-value. We discard this interaction and proceed with the next one.

#Case 7)

```
test <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1623.763 1673.767 -798.8816
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.165124 (Intr)
## nWBV         40.299769 -0.999
## Residual     1.861253
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3270035
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept)  6.57180  6.378684 208  1.030275  0.3041
## nWBV         81.48911 16.917555 208  4.816838  0.0000
## Age          0.23832  0.080255 208  2.969533  0.0033
## M.FM        -0.62003  0.414320 138 -1.496502  0.1368
## SES          6.17838  2.195653 138  2.813917  0.0056
## EDUC         0.13927  0.097947 138  1.421880  0.1573
## nWBV:SES     -20.80074  6.174351 208 -3.368895  0.0009
## Age:SES      -0.07664  0.028633 208 -2.676626  0.0080
## Correlation:
##           (Intr) nWBV   Age    M.FM   SES    EDUC   nWBV:S
## nWBV       -0.599
## Age        -0.985  0.581
## M.FM       -0.127  0.111  0.130
## SES        -0.899  0.578  0.920  0.091
## EDUC       -0.234 -0.053  0.088 -0.140  0.008
## nWBV:SES   0.532 -0.913 -0.517 -0.018 -0.606  0.028
## Age:SES    0.911 -0.535 -0.926 -0.082 -0.992 -0.091  0.563
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.9577092 -0.3023488  0.1509074  0.4464815  2.8116261
##
## Number of Observations: 354
## Number of Groups: 142
```

```
# Switch to ML so we can use ANOVA:
```

```
ml_test <- update(test, method = "ML")
anova(ml_fit4, ml_test)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	ml_fit4	1 12	1622.468	1668.899	-799.234			
##	ml_test	2 13	1617.520	1667.821	-795.760	1 vs 2	6.947978	0.0084

We see a decrease in AIC with a low p-value, so we will keep this interaction term. We give this model a name.

```
fit5 <- lme(MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(fit5)
```

```

## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1623.763 1673.767 -798.8816
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.165124 (Intr)
## nWBV         40.299769 -0.999
## Residual     1.861253
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3270035
## Fixed effects: MMSE ~ nWBV + Age + M.F + SES + EDUC + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept)  6.57180  6.378684 208  1.030275  0.3041
## nWBV         81.48911 16.917555 208  4.816838  0.0000
## Age          0.23832  0.080255 208  2.969533  0.0033
## M.FM        -0.62003  0.414320 138 -1.496502  0.1368
## SES          6.17838  2.195653 138  2.813917  0.0056
## EDUC         0.13927  0.097947 138  1.421880  0.1573
## nWBV:SES     -20.80074  6.174351 208 -3.368895  0.0009
## Age:SES      -0.07664  0.028633 208 -2.676626  0.0080
## Correlation:
##           (Intr) nWBV   Age    M.FM   SES    EDUC   nWBV:S
## nWBV       -0.599
## Age        -0.985  0.581
## M.FM       -0.127  0.111  0.130
## SES        -0.899  0.578  0.920  0.091
## EDUC       -0.234 -0.053  0.088 -0.140  0.008
## nWBV:SES   0.532 -0.913 -0.517 -0.018 -0.606  0.028
## Age:SES    0.911 -0.535 -0.926 -0.082 -0.992 -0.091  0.563
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.9577092 -0.3023488  0.1509074  0.4464815  2.8116261
##
## Number of Observations: 354
## Number of Groups: 142

```

Notice that, unlike in our earlier models, many coefficients in this model are statistically significant. Since sex and education appear to be insignificant, we will see if dropping them improves the model, starting with sex:

```
test <- lme(MMSE ~ nWBV + Age + SES + EDUC + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1624.086 1670.278 -800.0429
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.159381 (Intr)
## nWBV        40.654858 -0.999
## Residual    1.866020
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.327664
## Fixed effects: MMSE ~ nWBV + Age + SES + EDUC + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept)  5.28295  6.332527 208  0.834256  0.4051
## nWBV        84.46475 16.851531 208  5.012289  0.0000
## Age         0.25495  0.079634 208  3.201558  0.0016
## SES         6.50376  2.187894 139  2.972611  0.0035
## EDUC        0.11825  0.097035 139  1.218676  0.2250
## nWBV:SES    -21.01217  6.188756 208 -3.395218  0.0008
## Age:SES     -0.08048  0.028553 208 -2.818629  0.0053
## Correlation:
##      (Intr) nWBV   Age    SES    EDUC  nWBV:S
## nWBV      -0.591
## Age      -0.984  0.572
## SES      -0.899  0.571  0.919
## EDUC      -0.257 -0.038  0.108  0.021
## nWBV:SES  0.532 -0.916 -0.517 -0.605  0.025
## Age:SES   0.911 -0.528 -0.927 -0.992 -0.104  0.561
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.9117203 -0.2884104  0.1526157  0.4640838  2.7912355
##
## Number of Observations: 354
## Number of Groups: 142
```

Switch to ML so we can use ANOVA:

```
ml_fit5 <- update(fit5, method = "ML")
ml_test <- update(test, method = "ML")
anova(ml_fit5, ml_test)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	ml_fit5	1 13	1617.520	1667.821	-795.7600			
##	ml_test	2 12	1617.799	1664.231	-796.8995	1 vs 2	2.279012	0.1311

Since the p-value is high with minimal difference in AIC, we will drop sex. Now we see if we can drop education:

```
fit5 <- lme(MMSE ~ nWBV + Age + SES + EDUC + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1624.086 1670.278 -800.0429
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.159381 (Intr)
## nWBV         40.654858 -0.999
## Residual     1.866020
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.327664
## Fixed effects: MMSE ~ nWBV + Age + SES + EDUC + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept)  5.28295  6.332527 208  0.834256  0.4051
## nWBV         84.46475 16.851531 208  5.012289  0.0000
## Age          0.25495  0.079634 208  3.201558  0.0016
## SES          6.50376  2.187894 139  2.972611  0.0035
## EDUC         0.11825  0.097035 139  1.218676  0.2250
## nWBV:SES     -21.01217  6.188756 208 -3.395218  0.0008
## Age:SES      -0.08048  0.028553 208 -2.818629  0.0053
## Correlation:
##      (Intr) nWBV   Age    SES    EDUC   nWBV:S
## nWBV      -0.591
## Age       -0.984  0.572
## SES       -0.899  0.571  0.919
## EDUC      -0.257 -0.038  0.108  0.021
## nWBV:SES   0.532 -0.916 -0.517 -0.605  0.025
## Age:SES    0.911 -0.528 -0.927 -0.992 -0.104  0.561
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.9117203 -0.2884104  0.1526157  0.4640838  2.7912355
##
## Number of Observations: 354
## Number of Groups: 142
```

```
test <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```



```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1620.725 1663.099 -799.3626
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.153931 (Intr)
## nWBV         40.941276 -0.999
## Residual     1.866858
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3262077
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept)  7.18955  6.113428 208  1.176026  0.2409
## nWBV         85.40487 16.848709 208  5.068926  0.0000
## Age          0.24541  0.079073 208  3.103606  0.0022
## SES          6.47719  2.184955 140  2.964449  0.0036
## nWBV:SES     -21.27314  6.190681 208 -3.436317  0.0007
## Age:SES      -0.07724  0.028365 208 -2.722972  0.0070
## Correlation:
##           (Intr) nWBV   Age    SES   nWBV:S
## nWBV      -0.620
## Age       -0.996  0.578
## SES       -0.925  0.571  0.923
## nWBV:SES  0.556 -0.916 -0.521 -0.604
## Age:SES   0.920 -0.534 -0.926 -0.996  0.565
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.8980981 -0.2841556  0.1564629  0.4760757  2.7854975
##
## Number of Observations: 354
## Number of Groups: 142
```

#Switch to ML so we can use ANOVA:

```
ml_test <- update(test, method = "ML")
ml_fit5 <- update(fit5, method = "ML")
anova(ml_fit5, ml_test)
```

```
##           Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## ml_fit5      1 12 1617.799 1664.231 -796.8995
## ml_test      2 11 1617.342 1659.905 -797.6712 1 vs 2 1.543347  0.2141
```

Again, there is minimal difference in AIC with a high p-value, so we drop education.

```
fit5 <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(fit5)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1620.725 1663.099 -799.3626
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.153931 (Intr)
## nWBV         40.941276 -0.999
## Residual     1.866858
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3262077
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept)  7.18955  6.113428 208  1.176026  0.2409
## nWBV         85.40487 16.848709 208  5.068926  0.0000
## Age          0.24541  0.079073 208  3.103606  0.0022
## SES          6.47719  2.184955 140  2.964449  0.0036
## nWBV:SES     -21.27314  6.190681 208 -3.436317  0.0007
## Age:SES      -0.07724  0.028365 208 -2.722972  0.0070
## Correlation:
##           (Intr) nWBV   Age    SES   nWBV:S
## nWBV        -0.620
## Age         -0.996  0.578
## SES         -0.925  0.571  0.923
## nWBV:SES     0.556 -0.916 -0.521 -0.604
## Age:SES     0.920 -0.534 -0.926 -0.996  0.565
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.8980981 -0.2841556  0.1564629  0.4760757  2.7854975
##
## Number of Observations: 354
## Number of Groups: 142
```

```
ml_fit5 <- update(fit5, method = "ML")

# We compare our latest model with our two earlier models using an ANOVA test:

anova(fit2, fit3, fit5)
```

```
## Warning in anova.lme(fit2, fit3, fit5): fitted objects with different fixed
## effects. REML comparisons are not meaningful.
```

```
##      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## fit2      1 11 1632.025 1674.400 -805.0128
## fit3      2 15 1679.897 1737.507 -824.9487 1 vs 2 39.87192 <.0001
## fit5      3 11 1620.725 1663.099 -799.3626 2 vs 3 51.17229 <.0001
```

```
# We see that our latest model has the lowest AIC and all terms are statistically significant, so we proceed with this model
# as our tentative "best" model.
```

```
dc$fit_vals2 <- fitted(fit2)
dc$fit_vals3 <- fitted(fit3)
dc$fit_vals5 <- fitted(fit5)
```

Diagnostics for Random Effects

```
summary(fit5)
```

```

## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1620.725 1663.099 -799.3626
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.153931 (Intr)
## nWBV         40.941276 -0.999
## Residual     1.866858
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3262077
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept)  7.18955  6.113428 208  1.176026  0.2409
## nWBV         85.40487 16.848709 208  5.068926  0.0000
## Age          0.24541  0.079073 208  3.103606  0.0022
## SES          6.47719  2.184955 140  2.964449  0.0036
## nWBV:SES     -21.27314  6.190681 208 -3.436317  0.0007
## Age:SES      -0.07724  0.028365 208 -2.722972  0.0070
## Correlation:
##           (Intr) nWBV   Age    SES   nWBV:S
## nWBV       -0.620
## Age        -0.996  0.578
## SES        -0.925  0.571  0.923
## nWBV:SES    0.556 -0.916 -0.521 -0.604
## Age:SES     0.920 -0.534 -0.926 -0.996  0.565
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.8980981 -0.2841556  0.1564629  0.4760757  2.7854975
##
## Number of Observations: 354
## Number of Groups: 142

```

```
# Notice that the random effects matrix is numerically unstable due to the correlation being very close to -1 (-0.999),  
# which suggests that the random slope and intercept are almost perfectly negatively correlated.  
We will attempt to  
# resolve this by centering the random slope within-subject.  
  
test <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES,  
            data = dc,  
            random = ~ 1 + dvar(nWBV, Subject.ID) | Subject.ID,  
            correlation = corAR1(form = ~ 1 | Subject.ID),  
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))  
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1643.24 1685.614 -810.6199
##
## Random effects:
## Formula: ~1 + dvar(nWBV, Subject.ID) | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##
##              StdDev   Corr
## (Intercept)    2.731644 (Intr)
## dvar(nWBV, Subject.ID) 79.552499 -0.733
## Residual        1.527502
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.1161679
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##
##              Value Std.Error   DF   t-value p-value
## (Intercept)  7.37509  6.358712 208  1.159841  0.2474
## nWBV        79.97830 17.194289 208  4.651445  0.0000
## Age         0.24054  0.083259 208  2.889080  0.0043
## SES         4.90998  2.247075 140  2.185055  0.0305
## nWBV:SES    -15.89945  6.389334 208 -2.488436  0.0136
## Age:SES     -0.05631  0.029439 208 -1.912847  0.0571
## Correlation:
##      (Intr) nWBV   Age    SES   nWBV:S
## nWBV    -0.568
## Age     -0.996  0.566
## SES     -0.919  0.530  0.917
## nWBV:SES 0.506 -0.916 -0.507 -0.562
## Age:SES  0.916 -0.531 -0.921 -0.996  0.563
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.77155309 -0.26352203  0.07824107  0.35319931  3.34942747
##
## Number of Observations: 354
## Number of Groups: 142
```

```
anova(fit5, test)
```

```
##      Model df      AIC      BIC    logLik
## fit5     1 11 1620.725 1663.099 -799.3626
## test     2 11 1643.240 1685.614 -810.6199
```

```
dc$test <- fitted(test)

# Centering resolves the correlation issue, but at the cost of a relatively higher AIC, as well
# as the Age * SES interaction term
# Losing significance. What if we revisit the idea of dropping random slope altogether?

test <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES,
            data = dc,
            random = ~ 1 | Subject.ID,
            correlation = corAR1(form = ~ 1 | Subject.ID),
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
##   Data: dc
##       AIC      BIC    logLik
## 1668.793 1703.463 -825.3965
##
## Random effects:
## Formula: ~1 | Subject.ID
##      (Intercept) Residual
## StdDev:      2.16451 2.438783
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.576772
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##              Value Std.Error DF   t-value p-value
## (Intercept)  9.59805  7.004764 208   1.370217  0.1721
## nWBV         79.00883 17.010051 208   4.644832  0.0000
## Age          0.21204  0.091277 208   2.323072  0.0211
## SES          3.45461  2.450397 140   1.409818  0.1608
## nWBV:SES     -12.90163  6.283797 208  -2.053158  0.0413
## Age:SES      -0.03787  0.031975 208  -1.184240  0.2377
## Correlation:
##      (Intr) nWBV   Age    SES   nWBV:S
## nWBV      -0.551
## Age       -0.996  0.560
## SES       -0.920  0.518  0.917
## nWBV:SES   0.491 -0.914 -0.502 -0.552
## Age:SES    0.916 -0.528 -0.921 -0.996  0.565
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -5.6694520 -0.2574970  0.1602427  0.4689268  1.4793152
##
## Number of Observations: 354
## Number of Groups: 142
```

```
anova(fit5, test)
```

```
##      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## fit5     1 11 1620.725 1663.099 -799.3626
## test     2  9 1668.793 1703.463 -825.3965 1 vs 2 52.06787 <.0001
```

AIC increases and we have a low p-value, so this suggests we should keep the random slope. What if we instead revisit the idea of Age as a random effect?

```
test <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES,
            data = dc,
            random = ~ 1 + Age | Subject.ID,
            correlation = corAR1(form = ~ 1 | Subject.ID),
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```



```

## Linear mixed-effects model fit by REML
##   Data: dc
##       AIC      BIC    logLik
## 1672.794 1715.168 -825.3968
##
## Random effects:
## Formula: ~1 + Age | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev    Corr
## (Intercept) 2.164598683 (Intr)
## Age          0.000511519 -0.006
## Residual     2.438551061
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.576684
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##              Value Std.Error  DF   t-value p-value
## (Intercept)  9.59782  7.004716 208   1.370194  0.1721
## nWBV         79.00912 17.009972 208   4.644871  0.0000
## Age          0.21205  0.091277 208   2.323115  0.0211
## SES          3.45468  2.450384 140   1.409852  0.1608
## nWBV:SES     -12.90157  6.283764 208  -2.053160  0.0413
## Age:SES      -0.03787  0.031975 208  -1.184269  0.2377
## Correlation:
##      (Intr) nWBV   Age    SES   nWBV:S
## nWBV      -0.551
## Age       -0.996  0.560
## SES       -0.920  0.518  0.917
## nWBV:SES  0.491 -0.914 -0.502 -0.552
## Age:SES   0.916 -0.528 -0.921 -0.996  0.565
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -5.6693886 -0.2574740  0.1602281  0.4688469  1.4793940
##
## Number of Observations: 354
## Number of Groups: 142

```

```
anova(fit5, test)
```

```

##      Model df      AIC      BIC    logLik
## fit5     1 11 1620.725 1663.099 -799.3626
## test     2 11 1672.794 1715.168 -825.3968

```

```
# AIC increases again; moreover, notice that the correlation between slope and intercept becomes
very close to 0. Perhaps
# centering Age resolves this?

test <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 + dvar(Age, Subject.ID) | Subject.ID,
  correlation = corAR1(form = ~ 1 | Subject.ID),
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
##   Data: dc
##       AIC       BIC    logLik
##  1661.663 1704.037 -819.8316
##
## Random effects:
## Formula: ~1 + dvar(Age, Subject.ID) | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev   Corr
## (Intercept)    2.5784902 (Intr)
## dvar(Age, Subject.ID) 0.3291202 0.984
## Residual        1.9116980
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3469412
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##              Value Std.Error DF   t-value p-value
## (Intercept)   8.79880   6.737038 208   1.306034  0.1930
## nWBV          82.50288  16.555845 208   4.983308  0.0000
## Age           0.22667   0.088602 208   2.558351  0.0112
## SES           4.69058   2.351697 140   1.994552  0.0480
## nWBV:SES      -16.34440   6.122761 208  -2.669450  0.0082
## Age:SES       -0.05460   0.030952 208  -1.763956  0.0792
## Correlation:
##      (Intr) nWBV   Age    SES   nWBV:S
## nWBV      -0.582
## Age       -0.996  0.582
## SES       -0.922  0.543  0.920
## nWBV:SES   0.514 -0.915 -0.516 -0.567
## Age:SES    0.919 -0.545 -0.923 -0.996  0.570
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -4.4600524 -0.2734584  0.0777182  0.4168110  2.4473895
##
## Number of Observations: 354
## Number of Groups: 142
```

```
anova(fit5, test)
```

##	Model	df	AIC	BIC	logLik
##	fit5	1 11	1620.725	1663.099	-799.3626
##	test	2 11	1661.663	1704.037	-819.8316

Centering the age again results in a higher AIC; moreover, notice that the correlation is now very close to 1.

What if we try including both nWBV and Age in random effects?

```
test <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES,  
            data = dc,  
            random = ~ nWBV + Age | Subject.ID,  
            correlation = corAR1(form = ~ 1 | Subject.ID),  
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))  
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1626.72 1680.651 -799.3601
##
## Random effects:
## Formula: ~nWBV + Age | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev      Corr
## (Intercept) 2.152796e+00 (Intr) nWBV
## nWBV         4.095088e+01 -0.999
## Age          1.576016e-05  0.000  0.000
## Residual     1.869205e+00
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3287629
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept)  7.19564  6.117339 208   1.176270  0.2408
## nWBV         85.37349 16.857846 208   5.064318  0.0000
## Age          0.24533  0.079125 208   3.100513  0.0022
## SES          6.47188  2.186340 140   2.960142  0.0036
## nWBV:SES     -21.25592  6.194170 208  -3.431602  0.0007
## Age:SES      -0.07717  0.028383 208  -2.718739  0.0071
## Correlation:
##           (Intr) nWBV  Age    SES   nWBV:S
## nWBV       -0.620
## Age        -0.996  0.578
## SES        -0.925  0.571  0.923
## nWBV:SES    0.556 -0.916 -0.521 -0.604
## Age:SES     0.920 -0.534 -0.926 -0.996  0.565
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.8938152 -0.2833021  0.1567827  0.4766906  2.7811050
##
## Number of Observations: 354
## Number of Groups: 142
```

```
anova(test, fit5)
```

```
##      Model df      AIC      BIC    logLik  Test      L.Ratio p-value
## test      1 14 1626.720 1680.651 -799.3601
## fit5      2 11 1620.725 1663.099 -799.3626 1 vs 2 0.004908021 0.9999
```

```
# For nWBV the correlation is very close to -1 and for Age the correlation is very close to 0. And we again see a higher AIC  
# with a high p-value. What if we try centering both terms?  
  
test <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES,  
            data = dc,  
            random = ~ dvar(nWBV, Subject.ID) + dvar(Age, Subject.ID) | Subject.ID,  
            correlation = corAR1(form = ~ 1 | Subject.ID),  
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))  
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1646.086 1700.016 -809.0428
##
## Random effects:
## Formula: ~dvar(nWBV, Subject.ID) + dvar(Age, Subject.ID) | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##
##              StdDev      Corr
## (Intercept)      2.7420095 (Intr) d(WBVS
## dvar(nWBV, Subject.ID) 104.7624108 -0.670
## dvar(Age, Subject.ID)   0.2115454 -0.453  0.956
## Residual              1.4849687
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.09029946
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##
##              Value Std.Error DF   t-value p-value
## (Intercept)  7.11388  6.334009 208   1.123124  0.2627
## nWBV          79.88284 17.332811 208   4.608764  0.0000
## Age           0.24287  0.082746 208   2.935076  0.0037
## SES           4.75004  2.241707 140   2.118940  0.0359
## nWBV:SES     -15.37532  6.450907 208  -2.383436  0.0181
## Age:SES      -0.05396  0.029318 208  -1.840493  0.0671
## Correlation:
##      (Intr) nWBV   Age    SES   nWBV:S
## nWBV      -0.574
## Age       -0.996  0.573
## SES       -0.919  0.538  0.916
## nWBV:SES   0.513 -0.917 -0.515 -0.572
## Age:SES    0.915 -0.539 -0.920 -0.996  0.574
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.64938008 -0.28535264  0.07313079  0.36355414  3.24872759
##
## Number of Observations: 354
## Number of Groups: 142
```

```
anova(fit5, test)
```

```
##      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## fit5      1 11 1620.725 1663.099 -799.3626
## test      2 14 1646.085 1700.016 -809.0428 1 vs 2 19.36039 2e-04
```

```
# Centering stabilizes the correlations at the cost of a higher AIC.
```

```
# What happens if we center nWBV and remove autocorrelation?
```

```
test <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 + dvar(nWBV, Subject.ID) | Subject.ID,
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1641.699 1680.221 -810.8494
##
## Random effects:
## Formula: ~1 + dvar(nWBV, Subject.ID) | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev   Corr
## (Intercept)    2.769654 (Intr)
## dvar(nWBV, Subject.ID) 82.002320 -0.717
## Residual        1.446901
##
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##              Value Std.Error DF   t-value p-value
## (Intercept)   7.43309   6.299383 208   1.179972  0.2394
## nWBV          80.43414  17.176870 208   4.682701  0.0000
## Age           0.23984   0.082514 208   2.906714  0.0040
## SES           4.95410   2.229860 140   2.221709  0.0279
## nWBV:SES      -16.16945   6.385101 208  -2.532371  0.0121
## Age:SES       -0.05686   0.029225 208  -1.945618  0.0530
## Correlation:
##      (Intr) nWBV   Age    SES   nWBV:S
## nWBV      -0.569
## Age       -0.996  0.567
## SES       -0.919  0.531  0.916
## nWBV:SES   0.507 -0.917 -0.508 -0.563
## Age:SES    0.916 -0.531 -0.921 -0.996  0.564
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.80003509 -0.27107276  0.06960314  0.35341122  3.58985504
##
## Number of Observations: 354
## Number of Groups: 142
```

```
anova(test, fit5)
```

```
##      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
## test      1 10 1641.699 1680.221 -810.8494
## fit5      2 11 1620.725 1663.099 -799.3626 1 vs 2 22.97373 <.0001
```

We again see that AIC increases with a small p-value.

What if we remove autocorrelation from the model with centered nWBV and centered Age?

```
test <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES,
            data = dc,
            random = ~ dvar(nWBV, Subject.ID) + dvar(Age, Subject.ID) | Subject.ID,
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1644.405 1694.483 -809.2023
##
## Random effects:
## Formula: ~dvar(nWBV, Subject.ID) + dvar(Age, Subject.ID) | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev      Corr
## (Intercept)    2.7685243 (Intr) d(WBVS
## dvar(nWBV, Subject.ID) 105.9809344 -0.660
## dvar(Age, Subject.ID)   0.2149892 -0.419  0.947
## Residual              1.4291804
##
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##              Value Std.Error DF   t-value p-value
## (Intercept)   7.07284  6.293365 208   1.123857  0.2624
## nWBV          80.61840 17.317939 208   4.655196  0.0000
## Age           0.24351  0.082255 208   2.960358  0.0034
## SES           4.81620  2.228848 140   2.160847  0.0324
## nWBV:SES      -15.73080  6.443269 208  -2.441431  0.0155
## Age:SES       -0.05481  0.029165 208  -1.879258  0.0616
##
## Correlation:
##      (Intr) nWBV   Age    SES   nWBV:S
## nWBV      -0.576
## Age       -0.996  0.575
## SES       -0.919  0.540  0.916
## nWBV:SES   0.515 -0.917 -0.517 -0.574
## Age:SES    0.915 -0.541 -0.920 -0.996  0.576
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.63933720 -0.26812543  0.06750717  0.36084662  3.40249604
##
## Number of Observations: 354
## Number of Groups: 142
```



```
anova(fit5, test)
```

```
##      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## fit5      1 11 1620.725 1663.099 -799.3626
## test      2 13 1644.405 1694.483 -809.2023 1 vs 2 19.67938 1e-04
```

*# AIC is higher again, but the p-value is small. What if we drop the insignificant Age * SES term?*

```
test <- lme(MMSE ~ nWBV + Age + SES + nWBV * SES,
            data = dc,
            random = ~ dvar(nWBV, Subject.ID) + dvar(Age, Subject.ID) | Subject.ID,
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1640.893 1687.154 -808.4467
##
## Random effects:
## Formula: ~dvar(nWBV, Subject.ID) + dvar(Age, Subject.ID) | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev      Corr
## (Intercept)  2.8004334 (Intr) d(WBVS
## dvar(nWBV, Subject.ID) 111.8261765 -0.637
## dvar(Age, Subject.ID)  0.2379695 -0.514  0.961
## Residual      1.4091921
##
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES
##              Value Std.Error DF   t-value p-value
## (Intercept) 17.73978  2.572587 209   6.895698  0.0000
## nWBV         63.10880 14.750335 209   4.278466  0.0000
## Age          0.10322  0.032594 209   3.166869  0.0018
## SES          0.64070  0.205909 140   3.111560  0.0023
## nWBV:SES     -8.64260  5.333343 209  -1.620484  0.1066
## Correlation:
##      (Intr) nWBV   Age    SES
## nWBV    -0.234
## Age     -0.975  0.233
## SES     -0.204  0.002  0.001
## nWBV:SES -0.043 -0.880  0.044  0.011
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.67450896 -0.27754616  0.06606324  0.34477523  3.43467951
##
## Number of Observations: 354
## Number of Groups: 142
```

Switch to ML so we can use ANOVA:

```
fit5_ml <- update(fit5, method = "ML")
test_ml <- update(test, method = "ML")
anova(fit5_ml, test_ml)
```

```
##           Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## fit5_ml      1 11 1617.342 1659.905 -797.6712
## test_ml      2 12 1643.574 1690.006 -809.7870 1 vs 2 24.23172  <.0001
```

*# Again, the AIC is higher. What if we drop nWBV*SES?*

```
test <- lme(MMSE ~ nWBV + Age + SES,
            data = dc,
            random = ~ dvar(nWBV, Subject.ID) + dvar(Age, Subject.ID) | Subject.ID,
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1646.396 1688.833 -812.1978
##
## Random effects:
## Formula: ~dvar(nWBV, Subject.ID) + dvar(Age, Subject.ID) | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##
##              StdDev      Corr
## (Intercept)      2.8127886 (Intr) d(WBVS
## dvar(nWBV, Subject.ID) 109.4525372 -0.649
## dvar(Age, Subject.ID)   0.2265192 -0.557  0.982
## Residual              1.4189235
##
## Fixed effects: MMSE ~ nWBV + Age + SES
##
##              Value Std.Error DF  t-value p-value
## (Intercept) 17.55329  2.579116 210  6.805934  0.0000
## nWBV         42.00858  7.025723 210  5.979254  0.0000
## Age          0.10546  0.032663 210  3.228630  0.0014
## SES          0.64771  0.206862 140  3.131125  0.0021
## Correlation:
##      (Intr) nWBV   Age
## nWBV -0.575
## Age  -0.975  0.574
## SES  -0.204  0.024  0.001
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.69671065 -0.28487403  0.06913295  0.36127854  3.43519690
##
## Number of Observations: 354
## Number of Groups: 142
```

```
#ML for anova
test_ml <- update(test, method = "ML")
anova(fit5_ml, test_ml)
```

```
##      Model df      AIC      BIC    logLik
## fit5_ml    1 11 1617.342 1659.905 -797.6712
## test_ml    2 11 1644.384 1686.947 -811.1923
```

```
# AIC is still higher.
```

```
# At this point, our options are a trade-off between lower AIC and an extra interaction term with our earlier model (fit5) versus
```

```
# relatively stable random effects with a centered nWBV term.
```

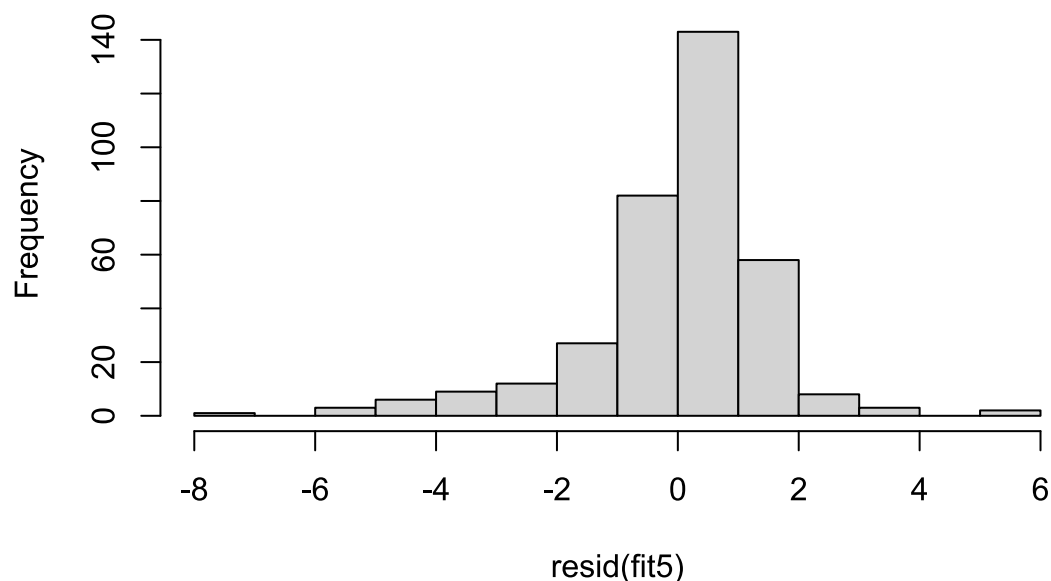
```
# Another possible remedy is to consider incorporating non-linear terms into the model. We will consider this approach in the
```

```
# next section on residuals.
```

Diagnostics for Residuals

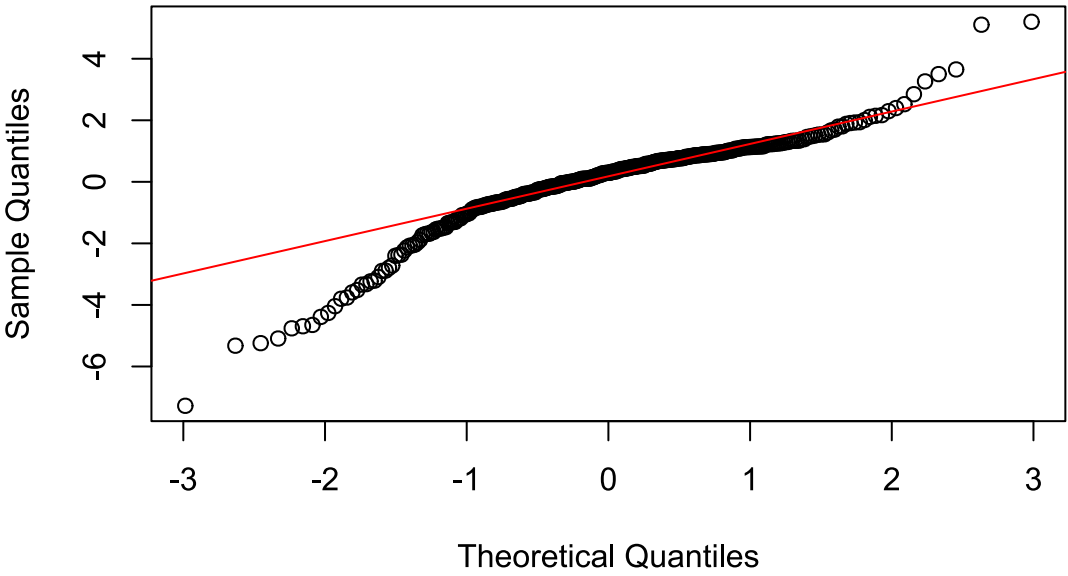
```
hist(resid(fit5), main = "Histogram of Residuals (fit5)")
```

Histogram of Residuals (fit5)

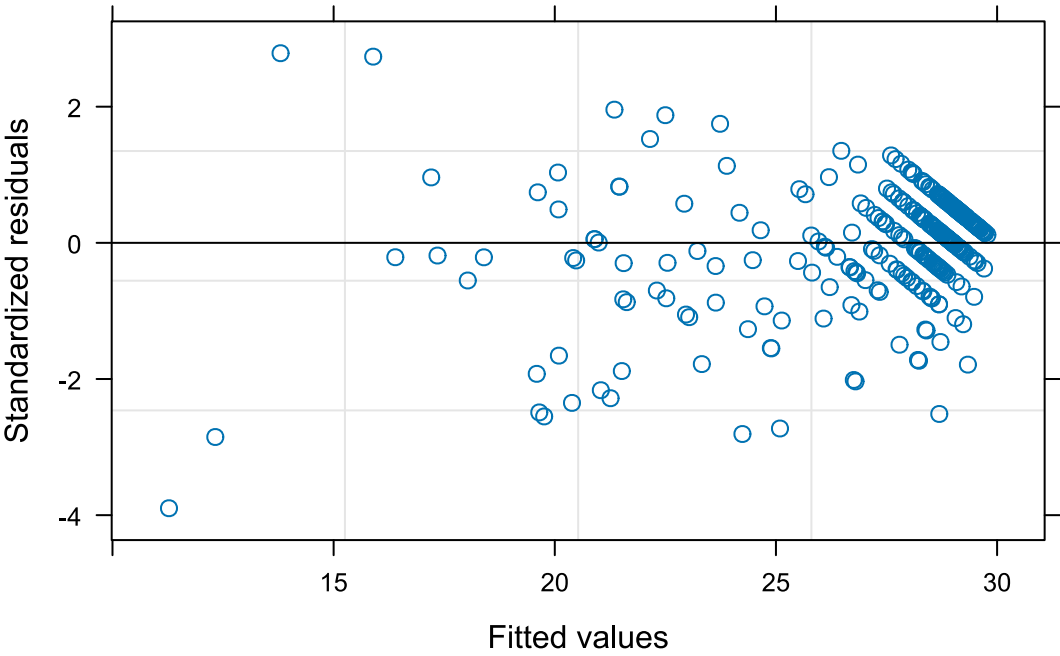


```
qqnorm(resid(fit5)); qqline(resid(fit5), col = "red")
```

Normal Q-Q Plot



```
plot(fit5)
```

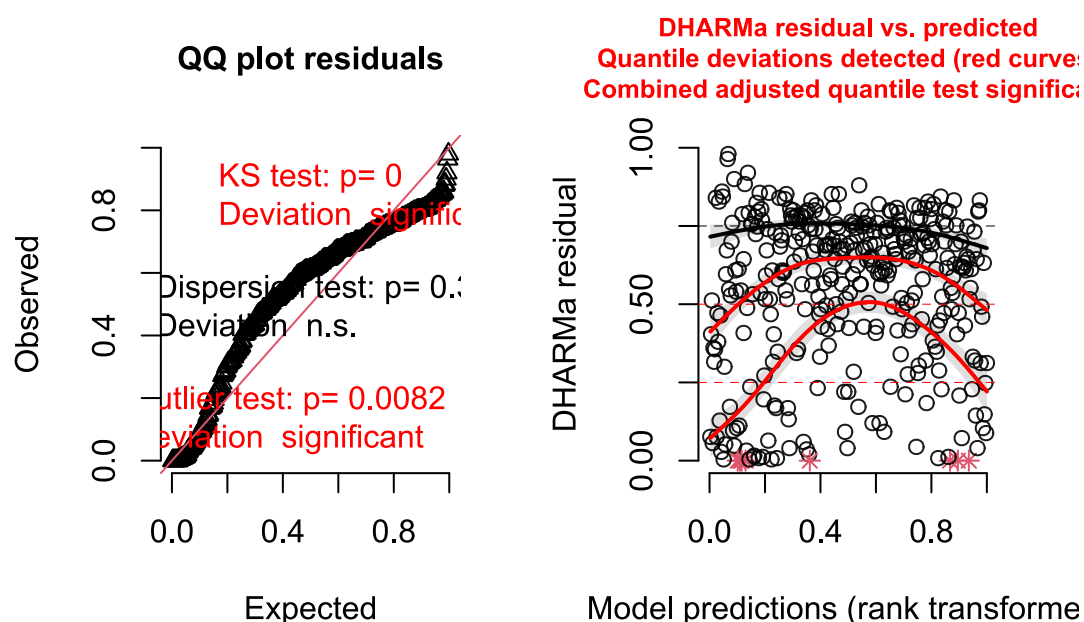


```
# We see kurtosis in the QQ plot. The standardized residuals plot is not well-suited for longitudinal models; instead, we will
# need to reconstruct our model using glmmTMB so that we can then use DHARMa to plot residuals.

fit5_tmb <- glmmTMB(
  MMSE ~ nWBV + Age + SES + nWBV:SES + Age:SES + (1 + nWBV | Subject.ID),
  data = dc,
  REML = TRUE
)

res <- simulateResiduals(fittedModel = fit5_tmb)
plot(res)
```

DHARMa residual



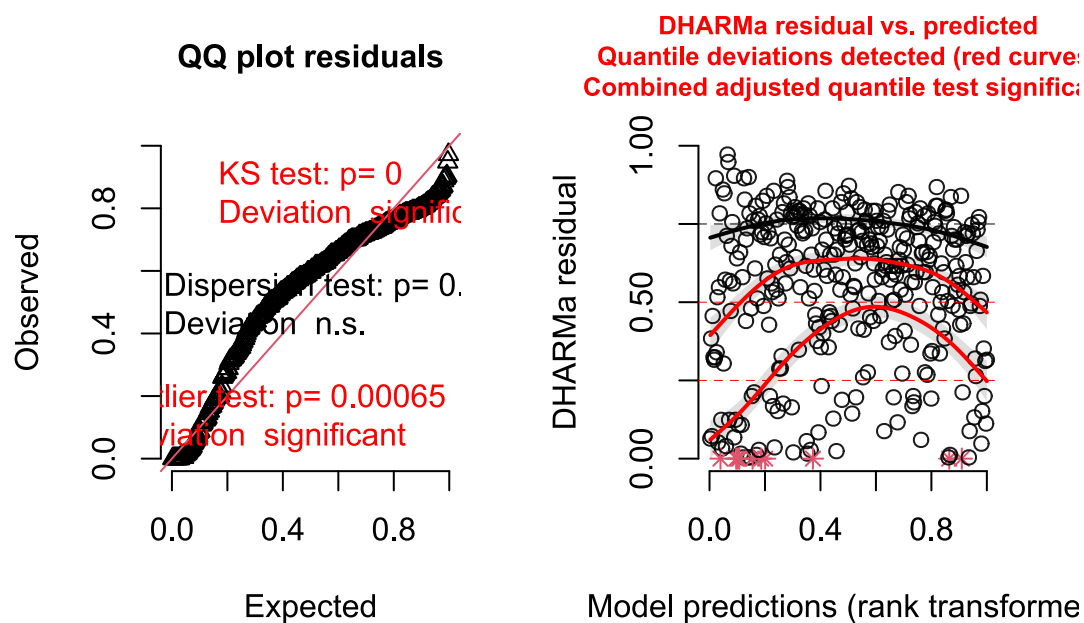
```
# Failing the KS test indicates that our residuals are not uniformly distributed, and failing the outlier test suggests that
# our model handles outliers poorly. The curves in the residual plot being highly non-linear suggests that the model handles
# heteroskedasticity poorly.
```

```
# We will try incorporating non-linear terms into the model. Beginning with a square term for Age:
```

```
fit_age_poly <- glmmTMB(
  MMSE ~ poly(Age, 2) + nWBV + SES + nWBV:SES + poly(Age, 2):SES + (1 + nWBV | Subject.ID),
  data = dc,
  REML = TRUE
)

sim_age_poly <- simulateResiduals(fit_age_poly)
plot(sim_age_poly)
```

DHARMA residual



We see no improvement in the residual plots.

```
dc$nwBV_sq <- dc$nwBV^2
```

```
dc$Age_sq <- dc$Age^2
```

What if we consider a square term for nwBV?

```
fit_nwbv_poly_raw <- glmmTMB(
```

```
  MMSE ~ nwBV + nwBV_sq + Age + SES + nwBV:SES + nwBV_sq:SES + Age:SES + (1 + nwBV | Subject.ID),
```

```
  data = dc,
```

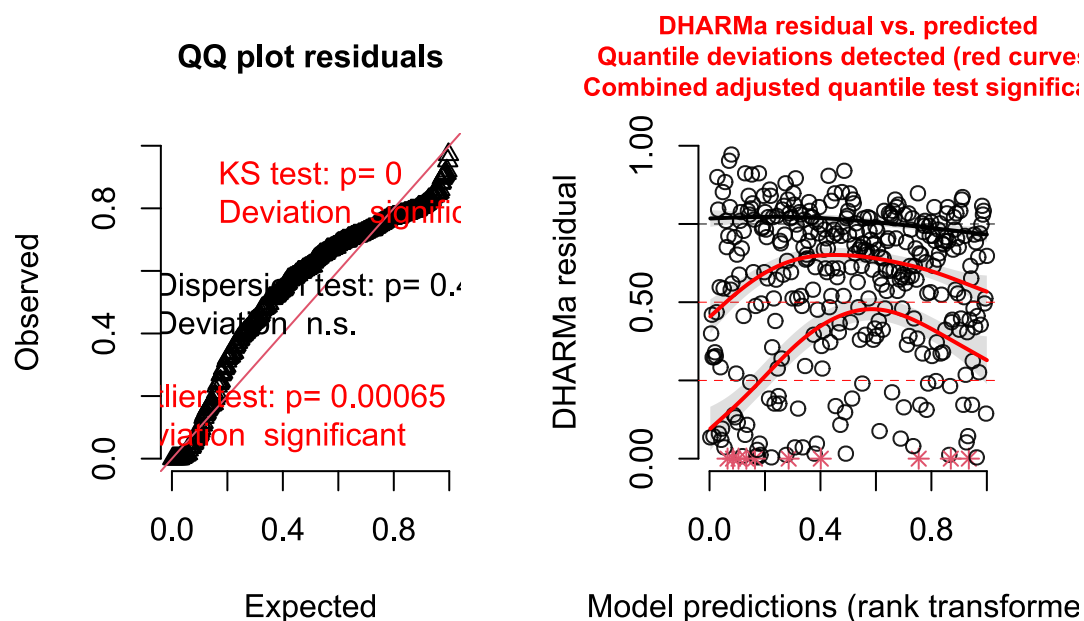
```
  REML = TRUE
```

```
)
```

```
sim_nwbv_poly <- simulateResiduals(fit_nwbv_poly_raw)
```

```
plot(sim_nwbv_poly)
```

DHARMA residual



We again see no improvement in the residual plots.

What if we drop slope from the random effects?

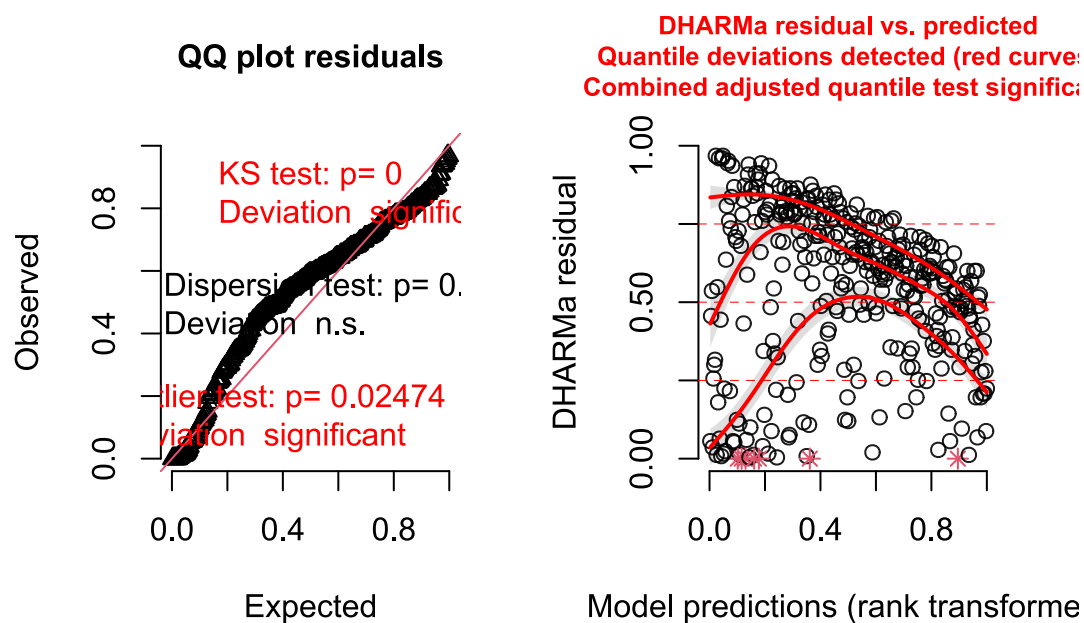
```
fit_simple_re <- glmmTMB(
  MMSE ~ nWBV + Age + SES + nWBV:SES + Age:SES + (1 | Subject.ID),
  data = dc,
  REML = TRUE
)
```

Our deviations get even worse; this validates our earlier tests that suggested keeping the random slope.

Let's compare AIC values:

```
sim_simple <- simulateResiduals(fit_simple_re)
plot(sim_simple)
```


DHARMA residual



```
AIC(fit5_tmb, fit_simple_re)
```

```
##           df      AIC
## fit5_tmb    10 1628.256
## fit_simple_re 8 1691.105
```

We see that AIC is lower if we keep nWBV as a random effect.

```
AIC(fit_age_poly, fit_nwbv_poly_raw)
```

```
##           df      AIC
## fit_age_poly    12 1602.622
## fit_nwbv_poly_raw 12 1601.715
```

AIC is lower if we have an nWBV squared term than if we have an age squared term.

```
AIC(fit_nwbv_poly_raw, fit5_tmb)
```

```
##           df      AIC
## fit_nwbv_poly_raw 12 1601.715
## fit5_tmb          10 1628.256
```

AIC is lower if we incorporate the squared nWBV term into our linear model.

```
AIC(fit_age_poly, fit5_tmb)
```

```
##           df      AIC
## fit_age_poly 12 1602.622
## fit5_tmb     10 1628.256
```

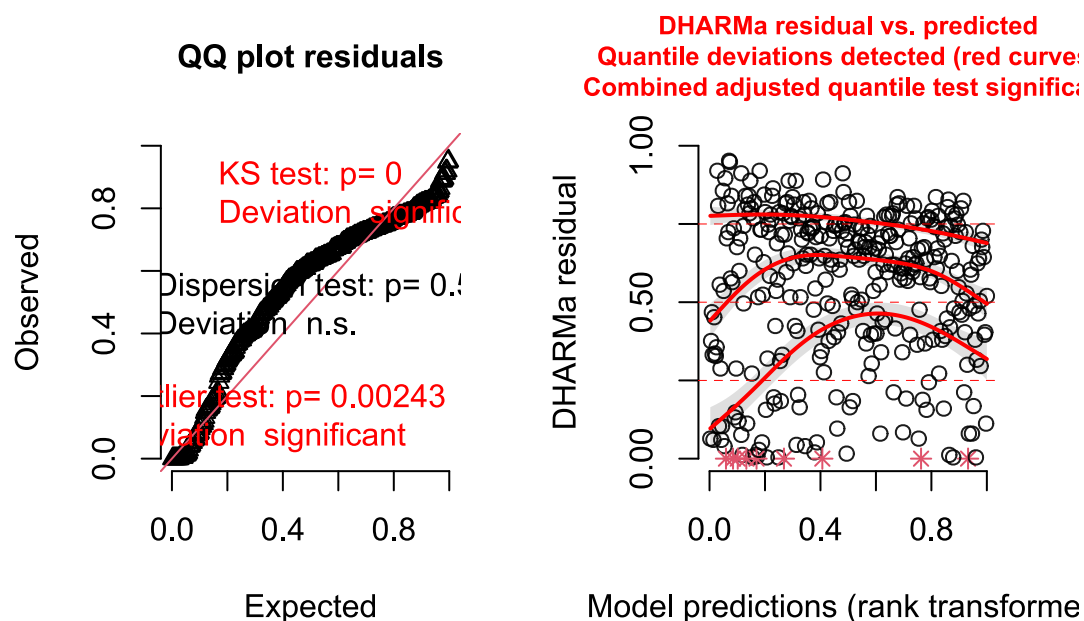
AIC is also lower if we incorporate the squared age term into our linear model.

What if we incorporate both squared age and squared nWBV?

```
fit_both <- glmmTMB(
  MMSE ~ nWBV + nWBV_sq + Age + Age_sq + SES + nWBV:SES + nWBV_sq:SES + Age:SES + (1 + nWBV | Subject.ID),
  data = dc,
  REML = TRUE
)

sim_both <- simulateResiduals(fit_both)
plot(sim_both)
```

DHARMa residual



We see that our residuals are even more poorly behaved.

```
AIC(fit_both, fit_nwbv_poly_raw)
```

```
##           df      AIC
## fit_both      13 1610.703
## fit_nwbv_poly_raw 12 1601.715
```

```
# Moreover, our AIC is worse than if we just have a squared nWBV term.

# Let's add the squared nWBV term to our LME.

test <- lme(MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1613.7 1656.042 -795.8498
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.312403 (Intr)
## nWBV         42.612419 -0.961
## Residual     1.630889
##
## Fixed effects: MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept)  7.78611   5.88642 207   1.322724  0.1874
## nWBV         95.00952  16.80813 207   5.652592  0.0000
## nWBV_sq      -239.24946 101.25160 207  -2.362920  0.0191
## Age           0.24044   0.07588 207   3.168804  0.0018
## SES           6.92424   2.09134 140   3.310911  0.0012
## nWBV:SES      -23.20838   6.05597 207  -3.832315  0.0002
## Age:SES       -0.08314   0.02711 207  -3.067214  0.0024
## Correlation:
##           (Intr) nWBV   nWBV_s Age    SES   nWBV:S
## nWBV         -0.598
## nWBV_sq      -0.110 -0.176
## Age          -0.995  0.556  0.097
## SES          -0.924  0.561  0.050  0.923
## nWBV:SES     0.559 -0.907  0.026 -0.521 -0.609
## Age:SES      0.919 -0.519 -0.048 -0.926 -0.995  0.565
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -4.1472977 -0.2784109  0.1098665  0.4405279  3.2477387
##
## Number of Observations: 354
## Number of Groups: 142
```

```
test_ml <- update(test, method = "ML")
fit5_ml <- update(fit5, method = "ML")
anova(test_ml, fit5_ml)
```

```
##           Model df      AIC      BIC    logLik
## test_ml      1 11 1620.705 1663.268 -799.3527
## fit5_ml      2 11 1617.342 1659.905 -797.6712
```

Note that all terms are still statistically significant (p -values < 0.02) in this new model. This model performs slightly worse than fit5 in terms of AIC, but correlation between random slope and intercept is more stable (-0.961).

```
test <- lme(MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 | Subject.ID,
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1674.695 1709.339 -828.3475
##
## Random effects:
## Formula: ~1 | Subject.ID
##      (Intercept) Residual
## StdDev:      2.704679 1.752774
##
## Fixed effects: MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES + Age * SES
##              Value Std.Error DF   t-value p-value
## (Intercept)  11.17831   6.82254 207   1.638438  0.1028
## nWBV          79.97651  16.88235 207   4.737286  0.0000
## nWBV_sq       -292.69477 108.43562 207  -2.699249  0.0075
## Age           0.19728   0.08868 207   2.224620  0.0272
## SES           3.24244   2.39221 140   1.355416  0.1775
## nWBV:SES      -12.58452   6.22289 207  -2.022297  0.0444
## Age:SES       -0.03516   0.03122 207  -1.126266  0.2614
## Correlation:
##      (Intr) nWBV  nWBV_s Age    SES  nWBV:S
## nWBV      -0.553
## nWBV_sq   -0.112 -0.035
## Age       -0.996  0.564  0.086
## SES       -0.917  0.524  0.046  0.916
## nWBV:SES   0.494 -0.914  0.060 -0.508 -0.564
## Age:SES    0.913 -0.535 -0.042 -0.919 -0.996  0.577
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -5.79989989 -0.26696609  0.07316919  0.36723603  3.17528506
##
## Number of Observations: 354
## Number of Groups: 142
```

Again, what if we drop nWBV from the random effects?

```
test_ml <- update(test, method = "ML")
anova(fit5_ml, test_ml)
```

```
##      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
## fit5_ml    1 11 1617.342 1659.905 -797.6712
## test_ml    2  9 1683.988 1718.812 -832.9942 1 vs 2 70.64602 <.0001
```

Dropping nWBV from random effects significantly reduces performance of this new model.

What happens if we add squared age to our LME?

```
test <- lme(MMSE ~ nWBV + Age_sq + Age + SES + nWBV * SES + Age * SES,
  data = dc,
  random = ~ 1 + nWBV | Subject.ID,
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1639.983 1682.326 -808.9915
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.334330 (Intr)
## nWBV        43.347984 -0.976
## Residual    1.649195
##
## Fixed effects: MMSE ~ nWBV + Age_sq + Age + SES + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept) 16.55367 14.314723 207  1.156409  0.2488
## nWBV        88.19691 16.576059 207  5.320741  0.0000
## Age_sq      0.00180  0.002276 207  0.789149  0.4309
## Age        -0.01498  0.354078 207 -0.042320  0.9663
## SES         7.17944  2.085791 140  3.442068  0.0008
## nWBV:SES    -23.06783  6.062503 207 -3.805001  0.0002
## Age:SES     -0.08651  0.027015 207 -3.202164  0.0016
## Correlation:
##           (Intr) nWBV   Age_sq Age    SES    nWBV:S
## nWBV        -0.301
## Age_sq      0.913 -0.046
## Age        -0.978  0.170 -0.977
## SES        -0.415  0.585 -0.041  0.237
## nWBV:SES    0.277 -0.917  0.049 -0.160 -0.616
## Age:SES     0.407 -0.540  0.035 -0.231 -0.995  0.569
##
## Standardized Within-Group Residuals:
##           Min      Q1      Med      Q3      Max
## -4.2755822 -0.2784075  0.1268521  0.4262244  3.2416607
##
## Number of Observations: 354
## Number of Groups: 142
```

```
test_ml <- update(test, method = "ML")
anova(fit5_ml, test_ml)
```

##	Model	df	AIC	BIC	logLik
##	fit5_ml	1 11	1617.342	1659.905	-797.6712
##	test_ml	2 11	1625.521	1668.083	-801.7606

Squared age also sees improvement in correlation and performs slightly worse than fit5; not as much improvement as squared

nWBV.

#We will keep the model that includes squared nWBV; call it fit6.

```
fit6 <- lme(MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES + Age * SES,  
           data = dc,  
           random = ~ 1 + nWBV | Subject.ID,  
           control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))  
summary(fit6)
```

```

## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1613.7 1656.042 -795.8498
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.312403 (Intr)
## nWBV         42.612419 -0.961
## Residual     1.630889
##
## Fixed effects: MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES + Age * SES
##           Value Std.Error DF   t-value p-value
## (Intercept)  7.78611   5.88642 207   1.322724  0.1874
## nWBV         95.00952  16.80813 207   5.652592  0.0000
## nWBV_sq      -239.24946 101.25160 207  -2.362920  0.0191
## Age           0.24044   0.07588 207   3.168804  0.0018
## SES           6.92424   2.09134 140   3.310911  0.0012
## nWBV:SES      -23.20838   6.05597 207  -3.832315  0.0002
## Age:SES       -0.08314   0.02711 207  -3.067214  0.0024
## Correlation:
##           (Intr) nWBV   nWBV_s Age    SES    nWBV:S
## nWBV        -0.598
## nWBV_sq     -0.110 -0.176
## Age         -0.995  0.556  0.097
## SES         -0.924  0.561  0.050  0.923
## nWBV:SES     0.559 -0.907  0.026 -0.521 -0.609
## Age:SES      0.919 -0.519 -0.048 -0.926 -0.995  0.565
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -4.1472977 -0.2784109  0.1098665  0.4405279  3.2477387
##
## Number of Observations: 354
## Number of Groups: 142

```

#What if we do within subject-centering for nWBV?

```

fit7 <- lme(MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES + Age * SES,
            data = dc,
            random = ~ 1 + dvar(nWBV, Subject.ID) | Subject.ID,
            control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(fit7)

```



```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1629.191 1671.534 -803.5955
##
## Random effects:
## Formula: ~1 + dvar(nWBV, Subject.ID) | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##
##              StdDev   Corr
## (Intercept)    2.763116 (Intr)
## dvar(nWBV, Subject.ID) 79.683931 -0.688
## Residual        1.440242
##
## Fixed effects: MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES + Age * SES
##
##              Value Std.Error DF   t-value p-value
## (Intercept)   8.79702   6.34494 207   1.386463  0.1671
## nWBV          82.37691  17.23383 207   4.779953  0.0000
## nWBV_sq      -203.09752 109.74305 207  -1.850664  0.0656
## Age           0.22624   0.08292 207   2.728584  0.0069
## SES           4.68921   2.24017 140   2.093235  0.0381
## nWBV:SES      -15.96121   6.38986 207  -2.497897  0.0133
## Age:SES       -0.05375   0.02935 207  -1.831549  0.0685
## Correlation:
##      (Intr) nWBV   nWBV_s Age    SES    nWBV:S
## nWBV      -0.559
## nWBV_sq   -0.086 -0.073
## Age       -0.996  0.561  0.059
## SES       -0.918  0.527  0.037  0.916
## nWBV:SES   0.506 -0.914  0.002 -0.508 -0.563
## Age:SES    0.915 -0.528 -0.031 -0.920 -0.996  0.565
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.78096232 -0.28073089  0.08416253  0.36061394  3.58145250
##
## Number of Observations: 354
## Number of Groups: 142
```

Correlation improves again, but Age:SES becomes insignificant. Can we drop it?

```
test <- lme(MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES,
  data = dc,
  random = ~ 1 + dvar(nWBV, Subject.ID) | Subject.ID,
  control = lmeControl(opt = "optim", maxIter = 200, msMaxIter = 200))
summary(test)
```

```
## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1625.162 1663.684 -802.581
##
## Random effects:
## Formula: ~1 + dvar(nWBV, Subject.ID) | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##
##              StdDev   Corr
## (Intercept)    2.794652 (Intr)
## dvar(nWBV, Subject.ID) 80.806040 -0.652
## Residual        1.427964
##
## Fixed effects: MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES
##
##              Value Std.Error DF   t-value p-value
## (Intercept)  19.56408   2.59781 208   7.530981  0.0000
## nWBV          66.00156  14.78470 208   4.464179  0.0000
## nWBV_sq       -216.96136 110.70469 208  -1.959821  0.0514
## Age           0.08509   0.03283 208   2.592136  0.0102
## SES           0.59595   0.20723 140   2.875719  0.0047
## nWBV:SES       -9.42023   5.32735 208  -1.768278  0.0785
## Correlation:
##      (Intr) nWBV   nWBV_s Age    SES
## nWBV      -0.220
## nWBV_sq   -0.143 -0.104
## Age       -0.973  0.224  0.079
## SES       -0.202  0.003  0.063 -0.004
## nWBV:SES  -0.036 -0.879  0.024  0.038  0.003
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.86990092 -0.29727772  0.08939817  0.35156032  3.63580799
##
## Number of Observations: 354
## Number of Groups: 142
```

```
ml_test <- update(test, method = "ML")
ml_fit7 <- update(fit7, method = "ML")
anova(ml_test, ml_fit7)
```

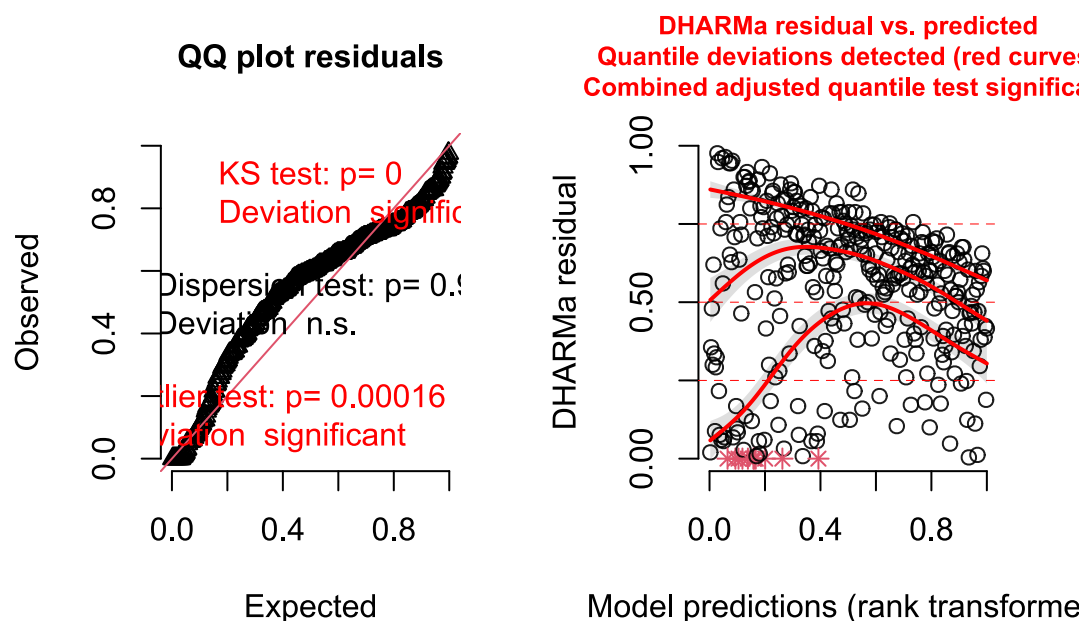
```
##      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
## ml_test    1 10 1639.445 1678.138 -809.7224
## ml_fit7    2 11 1638.075 1680.637 -808.0375 1 vs 2 3.369639  0.0664
```

```
# Although the correlation improves, we again see that the model has a relatively higher AIC, and we also see that
# squared nWBV and nWBV * SES become statistically insignificant (p-values > 0.05). Centering seems to be the wrong
# choice here. What happens if we look at the residual plots?
```

```
fit7_tmb <- glmmTMB(
  MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES + Age * SES +
  (1 + dvar(nWBV, Subject.ID) | Subject.ID),
  data = dc,
  REML = TRUE
)

res7 <- simulateResiduals(fittedModel = fit7_tmb)
plot(res7)
```

DHARMA residual



```
# Residuals worsen significantly with centered nWBV.
```

```
# We have three models to consider so far; fit5, fit6, fit7. Let's do an ANOVA test.
```

```
ml_fit5 <- update(fit5, method = "ML")
ml_fit6 <- update(fit6, method = "ML")
ml_fit7 <- update(fit7, method = "ML")
anova(ml_fit5, ml_fit6, ml_fit7)
```

##	Model	df	AIC	BIC	logLik
##	ml_fit5	1 11	1617.342	1659.905	-797.6712
##	ml_fit6	2 11	1620.705	1663.268	-799.3527
##	ml_fit7	3 11	1638.075	1680.637	-808.0375

```
# We see that our linear model, fit5, has the lowest AIC, closely followed by our squared nWBV model, fit6. Just as earlier,  
# centering lowers AIC.  
  
# Since non-linearity shows improvement, let's incorporate splines.  
  
dc$ns_nWBV <- ns(dc$nWBV, df = 4)  
ns_basis <- ns(dc$nWBV, df = 4)  
ns_df <- as.data.frame(ns_basis)  
colnames(ns_df) <- paste0("ns_nWBV_", 1:ncol(ns_df))  
  
dc <- bind_cols(dc, ns_df)  
  
fit_spline <- lme(MMSE ~ ns_nWBV_1 + ns_nWBV_2 + ns_nWBV_3 + Age + SES + nWBV * SES + Age * SES,  
  random = ~ 1 + nWBV | Subject.ID,  
  data = dc,  
  method = "REML"  
)  
  
summary(fit_spline)
```

```

## Linear mixed-effects model fit by REML
## Data: dc
##      AIC      BIC    logLik
## 1614.129 1664.095 -794.0644
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.347488 (Intr)
## nWBV        43.558128 -0.953
## Residual    1.594280
##
## Fixed effects: MMSE ~ ns_nWBV_1 + ns_nWBV_2 + ns_nWBV_3 + Age + SES + nWBV *      SES + Age
* SES
##           Value Std.Error  DF   t-value p-value
## (Intercept)  2.38331  5.996885 205   0.397425  0.6915
## ns_nWBV_1    4.29860  1.261824 205   3.406656  0.0008
## ns_nWBV_2    1.62393  1.423238 205   1.141010  0.2552
## ns_nWBV_3    8.33811  2.759514 205   3.021585  0.0028
## Age          0.25381  0.076411 205   3.321585  0.0011
## SES          7.11215  2.100483 140   3.385960  0.0009
## nWBV         98.39750 18.197478 205   5.407206  0.0000
## SES:nWBV     -23.90276  6.118227 205  -3.906811  0.0001
## Age:SES      -0.08572  0.027222 205  -3.148981  0.0019
## Correlation:
##           (Intr) n_WBV_1 n_WBV_2 n_WBV_3 Age    SES    nWBV    SES:WB
## ns_nWBV_1 -0.156
## ns_nWBV_2 -0.086  0.651
## ns_nWBV_3 -0.188  0.878  0.712
## Age       -0.969 -0.061 -0.096 -0.032
## SES       -0.907 -0.030 -0.033 -0.008  0.921
## nWBV      -0.580  0.000 -0.297 -0.048  0.562  0.538
## SES:nWBV  0.559 -0.045 -0.014 -0.039 -0.517 -0.605 -0.850
## Age:SES   0.902  0.028  0.034  0.007 -0.925 -0.995 -0.500  0.561
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.8820696 -0.3098540  0.1314801  0.4338532  3.1968448
##
## Number of Observations: 354
## Number of Groups: 142

```

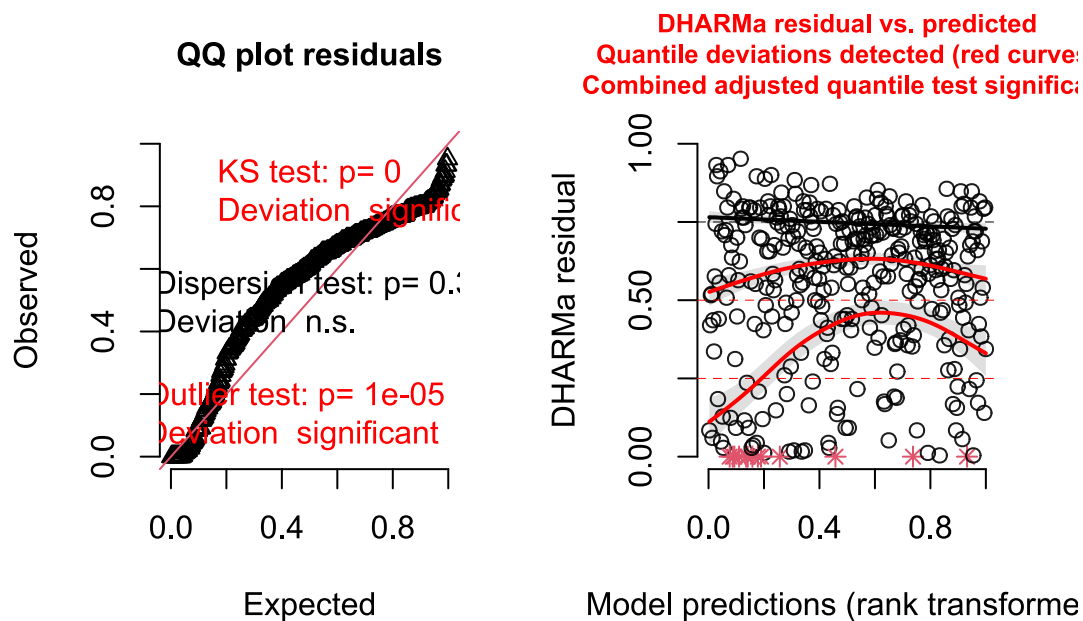
```
# We see further improvement in the random effects correlation. Note that we have omitted the fourth basis function from the
# model as it is collinear with the others.

# What happens if we look at residuals for our spline model?

fit_spline_tmb <- glmmTMB(
  MMSE ~ ns_nWBV_1 + ns_nWBV_2 + ns_nWBV_3 + Age + SES + nWBV * SES + Age * SES
  + (1 + nWBV | Subject.ID),
  data = dc,
  REML = TRUE
)

res_spline <- simulateResiduals(fit_spline_tmb)
plot(res_spline)
```

DHARMa residual



```
# We see the same issues in residuals as with our earlier models.
```

Possible Solutions for Residuals

At this point, it is reasonable to conclude that the issues with residuals arise from the data being highly non-normal; this gives us an idea of where the limitations in LMEs applied to non-normal data arise.

A possible solution is to instead construct a glmmTMB and then use its zero-inflation feature. Most of the MMSE scores are at or very close to 30, so we can reverse the ordering to have most scores be at or close to 0 and then account for zero-inflation.

As a simple example, let's apply this approach to fit5 (note that we omit the random slope to resolve convergence issues:

```
dc_c <- dc
dc_c$MMSE <- 30 - dc_c$MMSE
dc_c$nWBV_scaled <- scale(dc_c$nWBV)[,1]
dc_c$Age_scaled <- scale(dc_c$Age)[,1]
dc_c$SES_scaled <- scale(dc_c$SES)[,1]

fit5_tmb <- glmmTMB(
  MMSE ~ nWBV_scaled + Age_scaled + SES_scaled +
    nWBV_scaled * SES_scaled + Age_scaled * SES_scaled +
    (1 | Subject.ID),
  data = dc_c,
  ziformula = ~ Age_scaled,
  family = poisson,
  REML = TRUE
)

AIC(fit5, fit5_tmb)
```

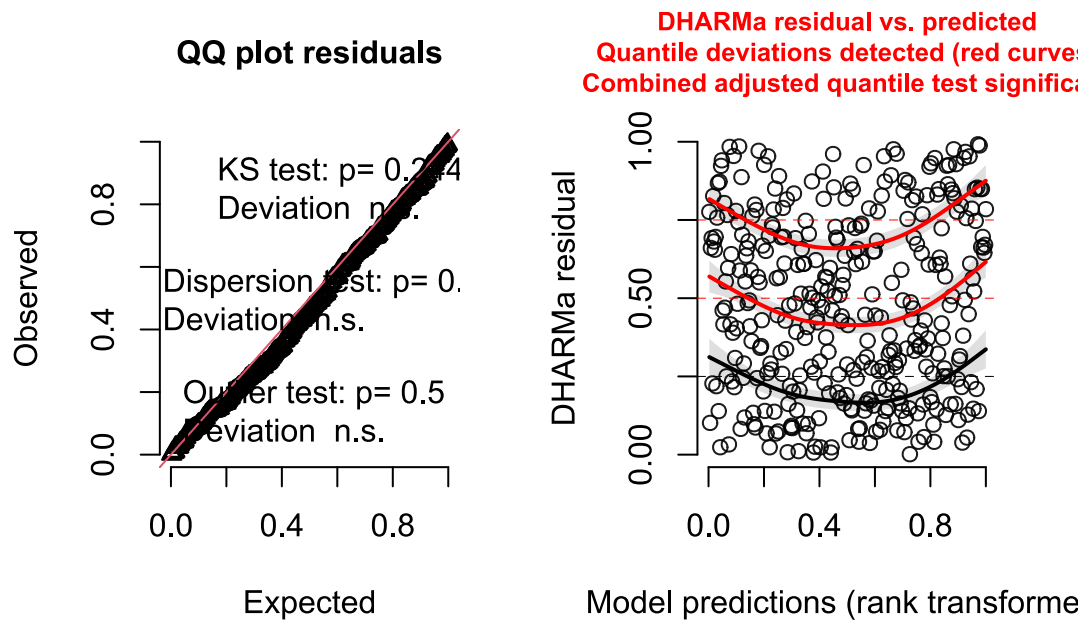
```
## Warning in AIC.default(fit5, fit5_tmb): models are not all fitted to the same
## number of observations
```

```
##           df      AIC
## fit5       11 1620.725
## fit5_tmb    9 1305.129
```

We immediately see a significantly lower AIC. Moreover, we can plot the residuals:

```
res1 <- simulateResiduals(fit5_tmb)
plot(res1)
```

DHARMA residual



We immediately see improvement in the residual plots; the curves are relatively flatter, and this model now passes the KS test and outlier test, indicating that the residuals are uniformly distributed and the model is less affected by outliers.

It was my intention to further explore this approach for `fit6` and `fit_spline`, but I ran into convergence issues that I was unable to resolve due to time constraints for the project. However, from the improvement in this example, I believe this approach to be very promising.

Model Comparison

At this point we have three models to consider; `fit5`, `fit6`, and `fit_spline`. Let's do an ANOVA test:

```
ml_fit5 <- update(fit5, method = "ML")
ml_fit6 <- update(fit6, method = "ML")
ml_fit_spline <- update(fit_spline, method = "ML")
anova(ml_fit5, ml_fit6, ml_fit_spline)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	ml_fit5	1 11	1617.342	1659.905	-797.6712			
##	ml_fit6	2 11	1620.705	1663.268	-799.3527			
##	ml_fit_spline	3 13	1616.616	1666.917	-795.3078	2 vs 3	8.089748	0.0175

We see that `fit_spline` has the lowest AIC, followed closely by `fit5`. The low p-value (< 0.02) for `fit_spline` suggests that the extra terms provide a statistically significant improvement to model fit.

Let's compare the mean slopes of the trajectories generated by these models to the mean slopes of the actual trajectories that we computed earlier.

Here are the slopes we originally computed:


```
print(avg_slopes_by_CDR_old)
```

```
## # A tibble: 3 × 3
##   CDR avg_slope    n
##   <dbl>    <dbl> <int>
## 1  0     -0.0306    76
## 2  0.5   -0.464     38
## 3  1     -0.826     9
```

Now we compute the slopes generated by these models.

For fit5:

```
dc$fit_vals5 <- fitted(fit5)

dc_slope <- dc %>%
  select(Subject.ID, Age, fit_vals5, CDR) %>%
  group_by(Subject.ID, CDR) %>%
  arrange(Age, .by_group = TRUE) %>%
  summarise(
    slope = if (n() >= 2) coef(lm(fit_vals5 ~ Age))[2] else NA_real_,
    .groups = "drop"
  )

avg_slopes_by_CDR_5 <- dc_slope %>%
  group_by(CDR) %>%
  summarise(
    avg_slope = mean(slope, na.rm = TRUE),
    n = n()
  )

print(avg_slopes_by_CDR_5)
```

```
## # A tibble: 4 × 3
##   CDR avg_slope    n
##   <dbl>    <dbl> <int>
## 1  0     0.00282    86
## 2  0.5   -0.161     61
## 3  1     -0.953     22
## 4  2     NaN       3
```

For fit6:

```
dc$fit_vals6 <- fitted(fit6)

dc_slope <- dc %>%
  select(Subject.ID, Age, fit_vals6, CDR) %>%
  group_by(Subject.ID, CDR) %>%
  arrange(Age, .by_group = TRUE) %>%
  summarise(
    slope = if (n() >= 2) coef(lm(fit_vals6 ~ Age))[2] else NA_real_,
    .groups = "drop"
  )

avg_slopes_by_CDR_6 <- dc_slope %>%
  group_by(CDR) %>%
  summarise(
    avg_slope = mean(slope, na.rm = TRUE),
    n = n()
  )

print(avg_slopes_by_CDR_6)
```

```
## # A tibble: 4 × 3
##   CDR avg_slope    n
##   <dbl>   <dbl> <int>
## 1  0     -0.0135   86
## 2  0.5   -0.227    61
## 3  1     -1.20    22
## 4  2      NaN     3
```

For fit_spline:

```
dc$fit_spline <- fitted(fit_spline)

dc_slope <- dc %>%
  select(Subject.ID, Age, fit_spline, CDR) %>%
  group_by(Subject.ID, CDR) %>%
  arrange(Age, .by_group = TRUE) %>%
  summarise(
    slope = if (n() >= 2) coef(lm(fit_spline ~ Age))[2] else NA_real_,
    .groups = "drop"
  )

avg_slopes_by_CDR <- dc_slope %>%
  group_by(CDR) %>%
  summarise(
    avg_slope = mean(slope, na.rm = TRUE),
    n = n()
  )

print(avg_slopes_by_CDR)
```

```
## # A tibble: 4 × 3
##   CDR avg_slope    n
##   <dbl>   <dbl> <int>
## 1     0   -0.0198    86
## 2   0.5   -0.302    61
## 3     1   -1.29     22
## 4     2    NaN      3
```

We see that fit5 has the closest mean slope for CDR = 1 to the actual mean trajectory slope while fit_spline has the closest mean slopes for CDR = 0 and CDR = 0.5. Note that the mean slope for fit5 is positive, suggesting an average increase in MMSE for those with CDR = 0; we would expect MMSE to decrease with age, even for those with CDR = 0. Note also that some subjects were lost in attempting to compute the mean slopes for the actual trajectories, so these estimates are limited compared to our model estimates, particularly for CDR = 1 which only has 9 subjects.

Model Interpretation

Finally, we interpret the fixed and random effects of our models.

For fit5:

```
summary(fit5)
```

```

## Linear mixed-effects model fit by REML
##   Data: dc
##       AIC      BIC    logLik
## 1620.725 1663.099 -799.3626
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.153931 (Intr)
## nWBV         40.941276 -0.999
## Residual     1.866858
##
## Correlation Structure: AR(1)
## Formula: ~1 | Subject.ID
## Parameter estimate(s):
##      Phi
## 0.3262077
## Fixed effects: MMSE ~ nWBV + Age + SES + nWBV * SES + Age * SES
##              Value Std.Error DF   t-value p-value
## (Intercept)  7.18955  6.113428 208   1.176026  0.2409
## nWBV         85.40487 16.848709 208   5.068926  0.0000
## Age          0.24541  0.079073 208   3.103606  0.0022
## SES          6.47719  2.184955 140   2.964449  0.0036
## nWBV:SES     -21.27314  6.190681 208  -3.436317  0.0007
## Age:SES      -0.07724  0.028365 208  -2.722972  0.0070
## Correlation:
##      (Intr) nWBV   Age    SES   nWBV:S
## nWBV      -0.620
## Age       -0.996  0.578
## SES       -0.925  0.571  0.923
## nWBV:SES   0.556 -0.916 -0.521 -0.604
## Age:SES    0.920 -0.534 -0.926 -0.996  0.565
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.8980981 -0.2841556  0.1564629  0.4760757  2.7854975
##
## Number of Observations: 354
## Number of Groups: 142

```

We will not attempt to interpret random effects here since the correlation is very close to -1.

For the fixed effects, the coefficient of nWBV (~85.40) suggests that increased nWBV is associated with increased MMSE, and this effect is dampened by an interaction with SES, as nWBV * SES has coefficient (~-21.27). The coefficient of Age (~0.25) is associated with increased MMSE, but this effect is dampened by an interaction with SES, as Age * SES has coefficient (~-0.08). The coefficient of SES is ~6.48, suggesting that higher SES is associated with higher MMSE. These coefficients are all statistically significant (p-values < 0.01).

For fit6:

```
summary(fit6)
```

```
## Linear mixed-effects model fit by REML
##   Data: dc
##       AIC      BIC    logLik
##   1613.7 1656.042 -795.8498
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev   Corr
## (Intercept)  2.312403 (Intr)
## nWBV         42.612419 -0.961
## Residual     1.630889
##
## Fixed effects: MMSE ~ nWBV + nWBV_sq + Age + SES + nWBV * SES + Age * SES
##              Value Std.Error  DF   t-value p-value
## (Intercept)   7.78611   5.88642 207   1.322724  0.1874
## nWBV          95.00952  16.80813 207   5.652592  0.0000
## nWBV_sq      -239.24946 101.25160 207  -2.362920  0.0191
## Age           0.24044   0.07588 207   3.168804  0.0018
## SES           6.92424   2.09134 140   3.310911  0.0012
## nWBV:SES      -23.20838   6.05597 207  -3.832315  0.0002
## Age:SES       -0.08314   0.02711 207  -3.067214  0.0024
## Correlation:
##      (Intr) nWBV   nWBV_s Age    SES   nWBV:S
## nWBV      -0.598
## nWBV_sq   -0.110 -0.176
## Age       -0.995  0.556  0.097
## SES       -0.924  0.561  0.050  0.923
## nWBV:SES   0.559 -0.907  0.026 -0.521 -0.609
## Age:SES    0.919 -0.519 -0.048 -0.926 -0.995  0.565
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -4.1472977 -0.2784109  0.1098665  0.4405279  3.2477387
##
## Number of Observations: 354
## Number of Groups: 142
```

For random effects, the correlation between the slope term nWBV and intercept is -0.961; this suggests that those with a higher baseline MMSE experience a greater decline in MMSE with loss of brain volume.

For the fixed effects, the coefficients in fit6 that are present in fit5 are all similar, and are all to be interpreted similarly, except for the additional squared nWBV coefficient (~ -239.25), which is suggestive of a non-linear relationship between nWBV and MMSE, and that there are diminishing returns in MMSE as nWBV increases.

For fit_spline:

```
summary(fit_spline)
```

```

## Linear mixed-effects model fit by REML
##   Data: dc
##       AIC      BIC    logLik
##  1614.129 1664.095 -794.0644
##
## Random effects:
## Formula: ~1 + nWBV | Subject.ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 2.347488 (Intr)
## nWBV         43.558128 -0.953
## Residual     1.594280
##
## Fixed effects: MMSE ~ ns_nWBV_1 + ns_nWBV_2 + ns_nWBV_3 + Age + SES + nWBV *      SES + Age
* SES
##           Value Std.Error  DF   t-value p-value
## (Intercept)  2.38331  5.996885 205   0.397425  0.6915
## ns_nWBV_1    4.29860  1.261824 205   3.406656  0.0008
## ns_nWBV_2    1.62393  1.423238 205   1.141010  0.2552
## ns_nWBV_3    8.33811  2.759514 205   3.021585  0.0028
## Age          0.25381  0.076411 205   3.321585  0.0011
## SES          7.11215  2.100483 140   3.385960  0.0009
## nWBV         98.39750 18.197478 205   5.407206  0.0000
## SES:nWBV     -23.90276  6.118227 205  -3.906811  0.0001
## Age:SES      -0.08572  0.027222 205  -3.148981  0.0019
## Correlation:
##           (Intr) n_WBV_1 n_WBV_2 n_WBV_3 Age    SES    nWBV    SES:WB
## ns_nWBV_1 -0.156
## ns_nWBV_2 -0.086  0.651
## ns_nWBV_3 -0.188  0.878  0.712
## Age       -0.969 -0.061 -0.096 -0.032
## SES       -0.907 -0.030 -0.033 -0.008  0.921
## nWBV      -0.580  0.000 -0.297 -0.048  0.562  0.538
## SES:nWBV  0.559 -0.045 -0.014 -0.039 -0.517 -0.605 -0.850
## Age:SES   0.902  0.028  0.034  0.007 -0.925 -0.995 -0.500  0.561
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -3.8820696 -0.3098540  0.1314801  0.4338532  3.1968448
##
## Number of Observations: 354
## Number of Groups: 142

```

For random effects, the interpretation is the same as in fit6.

For fixed effects, the coefficients in fit_spline that are present in fit5 are again similar and are again to be interpreted similarly. The coefficients for the spline terms (~4.30, ~1.62, ~8.34) are suggestive of a non-linear, positive relationship between nWBV and MMSE.

On the use of AI

ChatGPT and Claude were very useful for many aspects of this project, and I again want to give credit to both LLMs for their assistance. I have outlined the details of their use in the Discussion section of the report.