Total and progressive motility over time

Marie Bisconti, Philippe Grosjean & Elise Hennebert

Introduction

The effect of DMSO and ethanol is evaluated over time.

Total motility

Here we examine the effect of time on total motility without the addition of DMSO or ethanol.

```
mtt_cont <- readxl::read_excel("../data/Table S3.xlsx", sheet = "total motility_control")
mtt_cont$treat <- "control"
mtt_dmso <- readxl::read_excel("../data/Table S3.xlsx", sheet = "total motility_2%DMSO")
mtt_dmso$treat <- "DMSO 2%"
mtt_etoh1 <- readxl::read_excel("../data/Table S3.xlsx", sheet = "total motility_1%EtOH")
mtt_etoh1$treat <- "EtOH 1%"
mtt_etoh2 <- readxl::read_excel("../data/Table S3.xlsx", sheet = "total motility_2%EtOH")
mtt_etoh2$treat <- "EtOH 2%"
mtt <- rbind(mtt_cont, mtt_dmso, mtt_etoh1, mtt_etoh2)
mtt$donor <- as.factor(mtt$donor)
mtt$treat <- as.factor(mtt$treat)
names(mtt) <- c("donor", "time", "motile", "total", "treat")
mtt$motile_frac <- mtt$motile / mtt$total
skimr::skim(mtt)
```

Table 1: Data summary

Name	mtt
Number of rows	64
Number of columns	6
Column type frequency:	

Table 1: Data summary

factor	2
numeric	4
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
donor	0	1	FALSE	4	6: 16, 7: 16, 8: 16, 9: 16
treat	0	1	FALSE	4	con: 16, DMS: 16, EtO: 16, EtO: 16

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
time	0	1	1.62	1.57	0.00	0.38	1.25	2.50	4.00	
motile	0	1	82.98	31.78	29.00	58.50	79.50	105.25	174.00	
total	0	1	101.83	33.86	47.00	74.75	100.50	124.25	201.00	
motile_frac	0	1	0.80	0.07	0.59	0.77	0.81	0.85	0.94	

There are four donors, four treatments and no missing data.

```
table(mtt$donor, as.factor(mtt$time), mtt$treat)
```

```
, ,  = control


   0 0.5 2 4
6 1   1 1 1
7 1   1 1 1
8 1   1 1 1
9 1   1 1 1

, ,  = DMSO 2%
```

```
    0 0.5 2 4
  6 1   1 1 1
  7 1   1 1 1
  8 1   1 1 1
  9 1   1 1 1

, ,  = EtOH 1%


    0 0.5 2 4
  6 1   1 1 1
  7 1   1 1 1
  8 1   1 1 1
  9 1   1 1 1

, ,  = EtOH 2%


    0 0.5 2 4
  6 1   1 1 1
  7 1   1 1 1
  8 1   1 1 1
  9 1   1 1 1
```

The data are balanced with one observation for each treatment, each time and each donor and no missing data.

```
mtt_m1 <- glmer(cbind(motile, total - motile) ~ treat * time + (treat | donor),
  data = mtt, family = binomial(link = "logit"))
```

```
boundary (singular) fit: see help('isSingular')
```

```
summary(mtt_m1)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula: cbind(motile, total - motile) ~ treat * time + (treat | donor)
   Data: mtt
```

```
     AIC      BIC   logLik deviance df.resid
   449.2    488.1   -206.6    413.2       46

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2570 -0.9485 -0.2177  0.5707  2.8790

Random effects:
 Groups Name         Variance Std.Dev. Corr
 donor  (Intercept)  0.000000 0.00000
        treatDMSO 2% 0.004591 0.06776   NaN
        treatEtOH 1% 0.002260 0.04754   NaN 0.93
        treatEtOH 2% 0.077276 0.27799   NaN 0.46 0.10
Number of obs: 64, groups:  donor, 4

Fixed effects:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        1.83866    0.10476  17.551  < 2e-16 ***
treatDMSO 2%       0.06953    0.15230   0.457 0.647998
treatEtOH 1%      -0.52907    0.14255  -3.711 0.000206 ***
treatEtOH 2%      -0.64370    0.19476  -3.305 0.000950 ***
time              -0.06591    0.04301  -1.533 0.125376
treatDMSO 2%:time -0.07126    0.06059  -1.176 0.239551
treatEtOH 1%:time  0.11258    0.06011   1.873 0.061072 .
treatEtOH 2%:time  0.03526    0.05760   0.612 0.540359
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) trDMSO2% trEOH1% trEOH2% time   tDMSO2%: tEOH1%:
treatDMSO2% -0.688
treatEtOH1% -0.735  0.538
treatEtOH2% -0.538  0.442    0.407
time        -0.748  0.514    0.550   0.402
trtDMSO2%:t  0.531 -0.732   -0.388  -0.285  -0.710
trtEtOH1%:t  0.535 -0.367   -0.728  -0.288  -0.716  0.509
trtEtOH2%:t  0.558 -0.384   -0.409  -0.518  -0.747  0.530    0.534
optimizer (Nelder_Mead) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

There is a singularity in the model fitting because the correlation between donor and time is close to -1. We should try to simplify it. The model is:

,

$$\text{motile}_i \sim \text{Binomial}(n = 1, \text{prob}_{\text{motile}=1} = \widehat{P})$$

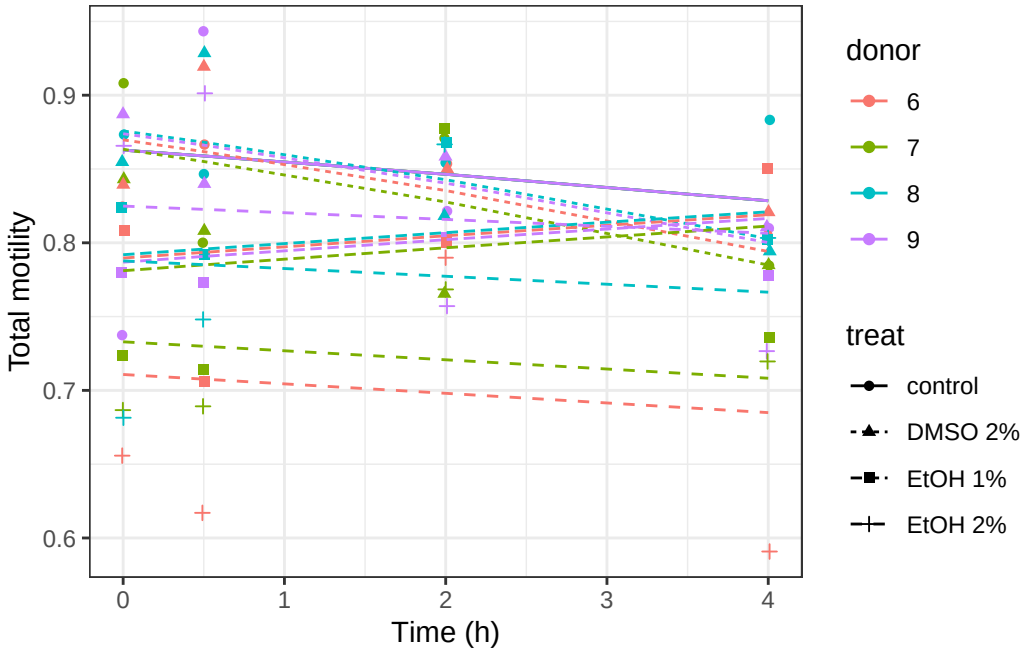
$$\log\left[\frac{\hat{P}}{1-\hat{P}}\right] = \alpha_{j[i]} + \beta_{1j[i]}(\text{treat}_{\text{DMSO 2\%}}) + \beta_{2j[i]}(\text{treat}_{\text{EtOH 1\%}}) + \beta_{3j[i]}(\text{treat}_{\text{EtOH 2\%}}) + \beta_4(\text{time}) + \beta_5(\text{time} \times$$

,

$$\begin{pmatrix} \alpha_j \\ \beta_{1j} \\ \beta_{2j} \\ \beta_{3j} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\alpha_j} \\ \mu_{\beta_{1j}} \\ \mu_{\beta_{2j}} \\ \mu_{\beta_{3j}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha_j} & \rho_{\alpha_j\beta_{1j}} & \rho_{\alpha_j\beta_{2j}} & \rho_{\alpha_j\beta_{3j}} \\ \rho_{\beta_{1j}\alpha_j} & \sigma^2_{\beta_{1j}} & \rho_{\beta_{1j}\beta_{2j}} & \rho_{\beta_{1j}\beta_{3j}} \\ \rho_{\beta_{2j}\alpha_j} & \rho_{\beta_{2j}\beta_{1j}} & \sigma^2_{\beta_{2j}} & \rho_{\beta_{2j}\beta_{3j}} \\ \rho_{\beta_{3j}\alpha_j} & \rho_{\beta_{3j}\beta_{1j}} & \rho_{\beta_{3j}\beta_{2j}} & \sigma^2_{\beta_{3j}} \end{pmatrix}\right), \text{ for donor j} = 1, \ldots, \text{J} \tag{1}$$

Here is a plot of this model:

```
ggplot(data = mtt) +
  geom_jitter(aes(x = time, y = motile_frac, shape = treat, col = donor), width = 0.01) +
  geom_line(aes(x = time, y = fitted(mtt_m1), linetype = treat, col = donor)) +
  labs(x = "Time (h)", y = "Total motility")
```



Here, we try to simplify the model so that singularity disappears. If we use `(1 | donor)`, we got this:

```
mtt_m2 <- glmer(cbind(motile, total - motile) ~ treat * time + (1 | donor),
  data = mtt, family = binomial(link = "logit"))
anova(mtt_m1, mtt_m2) # Despite the name, it is indeed a LR test
```

```
Data: mtt
Models:
mtt_m2: cbind(motile, total - motile) ~ treat * time + (1 | donor)
mtt_m1: cbind(motile, total - motile) ~ treat * time + (treat | donor)
       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
mtt_m2    9 440.88 460.31 -211.44   422.88
mtt_m1   18 449.23 488.09 -206.62   413.23 9.6475  9     0.3798
```

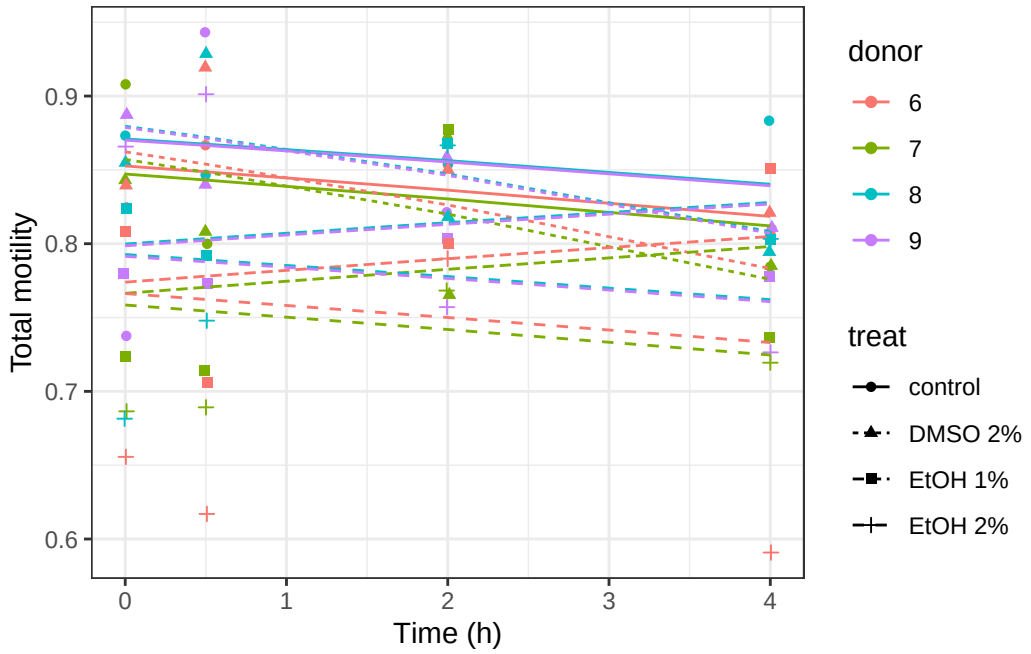
The likelihood ratio test does not detects significant differences between the full and simplified models at $\alpha = 5\%$. We could thus use the simplest `mtt_m2` model with only a shift in the slope per donor. This model is:

,

$$
\begin{aligned}
\text{motile}_i &\sim \text{Binomial}(n = 1, \text{prob}_{\text{motile}=1} = \widehat{P}) \\
, \log\left[\frac{\hat{P}}{1-\hat{P}}\right] &= \alpha_{j[i]} + \beta_1(\text{treat}_{\text{DMSO 2\%}}) + \beta_2(\text{treat}_{\text{EtOH 1\%}}) + \beta_3(\text{treat}_{\text{EtOH 2\%}}) + \beta_4(\text{time}) + \beta_5(\text{time} \times \text{treat}_{\text{DM} \\
\alpha_j &\sim N\left(\mu_{\alpha_j}, \sigma^2_{\alpha_j}\right), \text{for donor j} = 1, \ldots, \text{J}
\end{aligned}
\tag{2}
$$

Here is a plot of this model that forces the differences in total motility for samples at time = 0 to be the same for all treatments:

```
ggplot(data = mtt) +
  geom_jitter(aes(x = time, y = motile_frac, shape = treat, col = donor), width = 0.01) +
  geom_line(aes(x = time, y = fitted(mtt_m2), linetype = treat, col = donor)) +
  labs(x = "Time (h)", y = "Total motility")
```




```
summary(mtt_m2)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula: cbind(motile, total - motile) ~ treat * time + (1 | donor)
   Data: mtt

     AIC      BIC   logLik deviance df.resid
   440.9    460.3   -211.4    422.9       55

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5284 -0.9564 -0.2955  0.6484  3.4233

Random effects:
 Groups Name        Variance Std.Dev.
 donor  (Intercept) 0.01063  0.1031
Number of obs: 64, groups:  donor, 4

Fixed effects:
                  Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)        1.82055    0.11710  15.546  < 2e-16 ***
treatDMSO 2%       0.07873    0.14775   0.533 0.594119
treatEtOH 1%      -0.52463    0.13991  -3.750 0.000177 ***
treatEtOH 2%      -0.56825    0.13521  -4.203 2.64e-05 ***
time              -0.06231    0.04310  -1.446 0.148217
treatDMSO 2%:time -0.07509    0.06055  -1.240 0.214884
treatEtOH 1%:time  0.10889    0.06014   1.810 0.070224 .
treatEtOH 2%:time  0.01808    0.05726   0.316 0.752211
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) trDMSO2% trEOH1% trEOH2% time   tDMSO2%: tEOH1%:
treatDMSO2% -0.637
treatEtOH1% -0.672  0.532
treatEtOH2% -0.694  0.552    0.583
time        -0.672  0.530    0.561   0.579
trtDMSO2%:t  0.479 -0.752   -0.399  -0.413  -0.712
trtEtOH1%:t  0.482 -0.381   -0.742  -0.417  -0.717  0.511
trtEtOH2%:t  0.504 -0.400   -0.423  -0.739  -0.752  0.535    0.540
```

The Z test indicates that the slope for `time` and the difference of slope for DMSO and ethanol are not significantly different from zero at $\alpha = 5\%$. Intercept is significantly different for ethanol at $\alpha = 5\%$, but not for DMSO. However, it is not the best test in the case of a mixed model like here. We prefer to rely on the 95% confidence interval calculated either on the profile, or via parametric bootstrap (and especially the later one):

```
confint(mtt_m2, level = 0.95) # 95% CI based on profile
```

```
Computing profile confidence intervals ...

                         2.5 %      97.5 %
.sig01              0.024401012  0.29167216
(Intercept)         1.586116236  2.05696046
treatDMSO 2%       -0.211213957  0.36854622
treatEtOH 1%       -0.800361359 -0.25146492
treatEtOH 2%       -0.835226442 -0.30477199
time               -0.146519459  0.02260298
treatDMSO 2%:time  -0.193920744  0.04357366
treatEtOH 1%:time  -0.008972294  0.22694610
treatEtOH 2%:time  -0.094287781  0.13031820
```

```
set.seed(1643)
# 1000x parameter bootstrap
(mtt_m2_conf <- confint(mtt_m2, level = 0.95, method = "boot", nsim  = 1000L))
```

```
Computing bootstrap confidence intervals ...
```

```
65 message(s): boundary (singular) fit: see help('isSingular')
798 warning(s): Model failed to converge with max|grad| = 0.00200921 (tol = 0.002, component
```

```
                          2.5 %      97.5 %
.sig01              1.620473e-05  0.17576137
(Intercept)         1.594022e+00  2.06630632
treatDMSO 2%       -1.955345e-01  0.35919461
treatEtOH 1%       -8.179399e-01 -0.24784984
treatEtOH 2%       -8.490808e-01 -0.31262716
time               -1.485760e-01  0.02281249
treatDMSO 2%:time -1.883411e-01  0.03719578
treatEtOH 1%:time -7.159468e-03  0.22637425
treatEtOH 2%:time -9.910036e-02  0.13883909
```

All 95%ICs are not significantly different from zero at $\alpha = 5\%$ (they contain zero), except for the standanrd deviation of the random term (donor, .sig01), the intercept and for the shift in intercept for ethanol. This means we cannot detect an effect of time, or a significantly different effect of time in presence of DMSO or ethanol. Keep in mind, however, that we have few data for such a complex model. Even if it is perfectly balanced, prediction power is probably rather low. On the other hand, drop of motility after 4h is only a few percents, even for DMSO.

Additional verifications

We could double-check the significance of the difference in slope for `treat:time` by looking at a likelihood ratio test when dropping `conc` from the model:

```
#drop1(mtt_m2, scope = "time")
mtt_m3 <- glmer(cbind(motile, total - motile) ~ treat + time + (1 | donor),
  data = mtt, family = binomial(link = "logit"))
anova(mtt_m2, mtt_m3, refit = TRUE)
```

```
Data: mtt
Models:
mtt_m3: cbind(motile, total - motile) ~ treat + time + (1 | donor)
mtt_m2: cbind(motile, total - motile) ~ treat * time + (1 | donor)
       npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
mtt_m3    6 444.55 457.51 -216.28   432.55
mtt_m2    9 440.88 460.31 -211.44   422.88 9.672  3    0.02157 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is not significant difference between the two models at α level of 5%. This means that we do not detect significant differences in slopes between models.

We also double-check convergence of the model by trying different optimisation engines (just to make sure). First, is there a singularity in the model?

```
isSingular(mtt_m2)
```

```
[1] FALSE
```

… then, a report about the model convergence:

```
mtt_m2_all <- allFit(mtt_m2)
```

```
Loading required namespace: dfoptim

Loading required namespace: optimx

bobyqa : [OK]
Nelder_Mead : [OK]
nlminbwrap : [OK]
nmkbw :

Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
Model failed to converge with max|grad| = 0.00236931 (tol = 0.002, component 1)

[OK]
optimx.L-BFGS-B : [OK]
nloptwrap.NLOPT_LN_NELDERMEAD : [OK]
nloptwrap.NLOPT_LN_BOBYQA : [OK]
```

```r
summary(mtt_m2_all)
```

```
$which.OK
                        bobyqa                    Nelder_Mead
                          TRUE                           TRUE
                    nlminbwrap                          nmkbw
                          TRUE                           TRUE
              optimx.L-BFGS-B nloptwrap.NLOPT_LN_NELDERMEAD
                          TRUE                           TRUE
    nloptwrap.NLOPT_LN_BOBYQA
                          TRUE

$msgs
$msgs$bobyqa
NULL

$msgs$Nelder_Mead
NULL

$msgs$nlminbwrap
NULL

$msgs$nmkbw
[1] "Model failed to converge with max|grad| = 0.00236931 (tol = 0.002, component 1)"

$msgs$`optimx.L-BFGS-B`
NULL

$msgs$nloptwrap.NLOPT_LN_NELDERMEAD
NULL

$msgs$nloptwrap.NLOPT_LN_BOBYQA
NULL


$fixef
                            (Intercept) treatDMSO 2% treatEtOH 1%
bobyqa                         1.820546   0.07872758   -0.5246395
Nelder_Mead                    1.820557   0.07873472   -0.5246489
nlminbwrap                     1.820535   0.07873918   -0.5246279
nmkbw                          1.820537   0.07865544   -0.5245985
```

```
optimx.L-BFGS-B                  1.820314   0.07892879   -0.5243293
nloptwrap.NLOPT_LN_NELDERMEAD    1.820557   0.07872363   -0.5246579
nloptwrap.NLOPT_LN_BOBYQA        1.820601   0.07866186   -0.5247314
                              treatEtOH 2%        time treatDMSO 2%:time
bobyqa                         -0.5682546 -0.06231021       -0.07509064
Nelder_Mead                    -0.5682510 -0.06231212       -0.07509766
nlminbwrap                     -0.5682422 -0.06230728       -0.07509404
nmkbw                          -0.5682864 -0.06231234       -0.07505189
optimx.L-BFGS-B                -0.5679358 -0.06223314       -0.07515646
nloptwrap.NLOPT_LN_NELDERMEAD  -0.5682516 -0.06231243       -0.07509466
nloptwrap.NLOPT_LN_BOBYQA      -0.5683158 -0.06231565       -0.07508766
                              treatEtOH 1%:time treatEtOH 2%:time
bobyqa                               0.1088936        0.01807771
Nelder_Mead                          0.1089035        0.01807609
nlminbwrap                           0.1088898        0.01807356
nmkbw                                0.1088859        0.01808685
optimx.L-BFGS-B                      0.1087882        0.01797594
nloptwrap.NLOPT_LN_NELDERMEAD        0.1088969        0.01806394
nloptwrap.NLOPT_LN_BOBYQA            0.1089151        0.01808990

$llik
                       bobyqa                     Nelder_Mead
                    -211.4401                       -211.4401
                   nlminbwrap                            nmkbw
                    -211.4401                       -211.4401
              optimx.L-BFGS-B nloptwrap.NLOPT_LN_NELDERMEAD
                    -211.4401                       -211.4401
    nloptwrap.NLOPT_LN_BOBYQA
                    -211.4401

$sdcor
                              donor.(Intercept)
bobyqa                                0.1031111
Nelder_Mead                           0.1031028
nlminbwrap                            0.1031113
nmkbw                                 0.1031153
optimx.L-BFGS-B                       0.1031130
nloptwrap.NLOPT_LN_NELDERMEAD         0.1030984
nloptwrap.NLOPT_LN_BOBYQA             0.1031118

$theta
                              donor.(Intercept)
bobyqa                                0.1031111
```

```
Nelder_Mead                          0.1031028
nlminbwrap                           0.1031113
nmkbw                                0.1031153
optimx.L-BFGS-B                      0.1031130
nloptwrap.NLOPT_LN_NELDERMEAD        0.1030984
nloptwrap.NLOPT_LN_BOBYQA            0.1031118

$times
                             user.self sys.self elapsed user.child sys.child
bobyqa                           0.110    0.000   0.110          0         0
Nelder_Mead                      0.167    0.000   0.168          0         0
nlminbwrap                       0.187    0.000   0.187          0         0
nmkbw                            0.169    0.000   0.170          0         0
optimx.L-BFGS-B                  0.452    0.001   0.455          0         0
nloptwrap.NLOPT_LN_NELDERMEAD    0.144    0.001   0.146          0         0
nloptwrap.NLOPT_LN_BOBYQA        0.086    0.001   0.088          0         0

$feval
                         bobyqa                     Nelder_Mead
                            324                             573
                      nlminbwrap                           nmkbw
                             NA                             446
               optimx.L-BFGS-B nloptwrap.NLOPT_LN_NELDERMEAD
                             18                             408
    nloptwrap.NLOPT_LN_BOBYQA
                            101

attr(,"class")
[1] "summary.allFit"
```

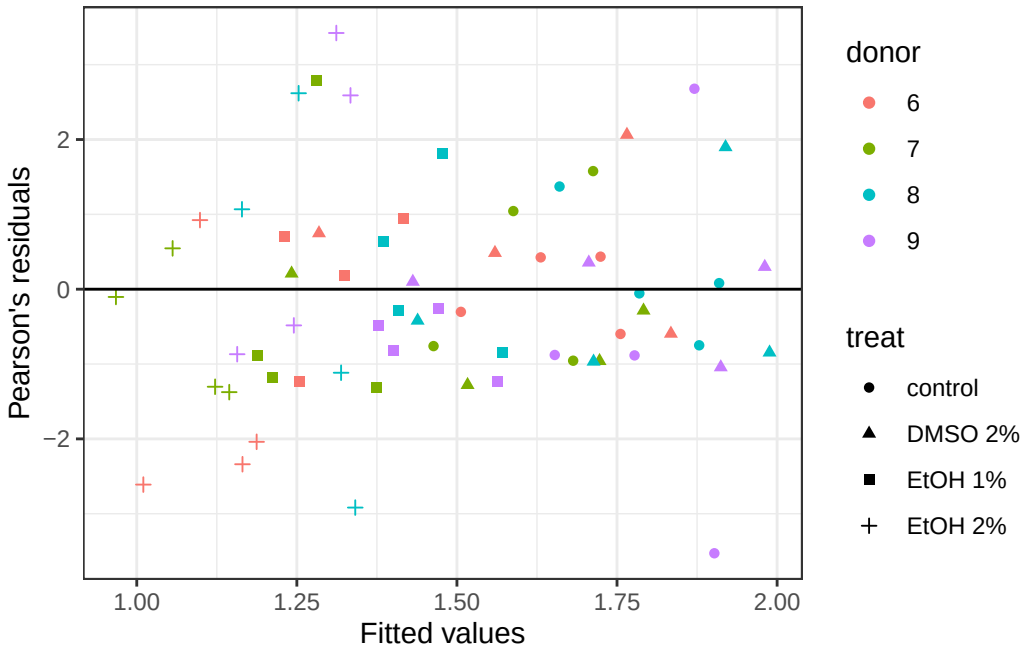
The model failed to converge with the nmkbw algorithm, but otherwise, results are consistent between the other optimisation algorithms.

Analysis of the residuals

Let's check how the residuals distribute and if there is homoscedasticity.

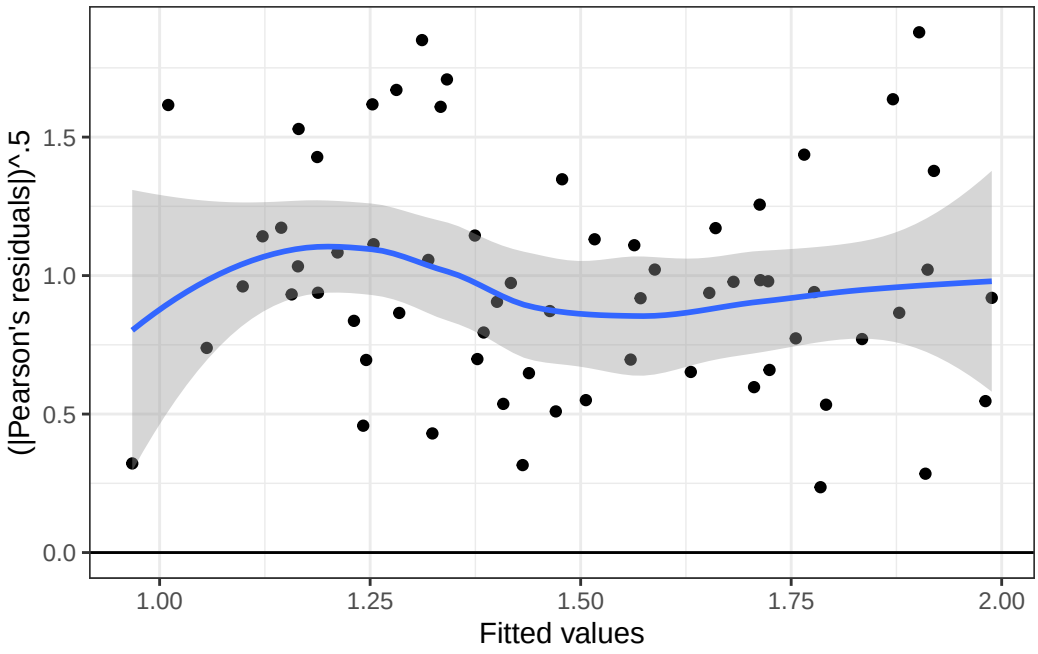
```
mtt <- fortify.merMod(mtt_m2)
ggplot(data = mtt, aes(x = .fitted, y = .scresid, shape = treat, col = donor)) +
  geom_point() +
  geom_hline(yintercept = 0) +
```

```
  labs(x = "Fitted values", y = "Pearson's residuals")
```



There is good.

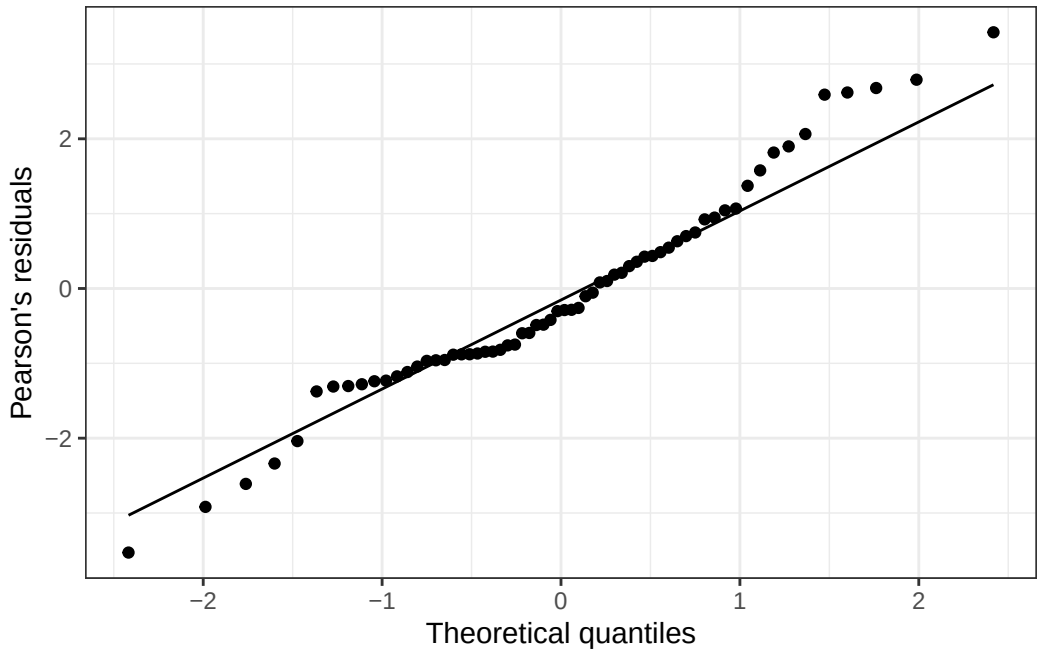
```
ggplot(data = mtt, aes(x = .fitted, y = sqrt(abs(.scresid)))) +
  geom_point() +
  geom_smooth(method = "loess", formula = y ~ x) +
  geom_hline(yintercept = 0) +
  labs(x = "Fitted values", y = "(|Pearson's residuals|)^.5")
```



Homoscedasticity is checked here too.

Note: according to Gauss-Markov theorem that indicates that linearity, random sample, non-collinearity between predictors, non-correlation between predictors and error term and homoscedasticity are the only requirements for our GLMM, we do not have to check it. Also the model is robust to departure of Normality, and none of the tests we made depend on a Normal distribution of the residuals (we said we don't trust z/t tests and replace them by likelihood ratio tests and parameterized bootstrapped confidence intervals). However, for completeness, here is the quantile-quantile plot of the residuals:

```
ggplot(data = mtt, aes(sample = .scresid)) +
  geom_qq() +
  geom_qq_line() +
  labs(x = "Theoretical quantiles", y = "Pearson's residuals")
```



It appears not really good. A Shapiro-Wilk test also indicates no Normality of the residuals at $\alpha = 5\%$.

```
shapiro.test(mtt$.scresid)
```

```
    Shapiro-Wilk normality test

data:  mtt$.scresid
W = 0.97053, p-value = 0.129
```

Progressive motility

Here we examine the effect of time on progressive motility without the addition of DMSO or ethanol.

```
mpt_cont <- readxl::read_excel("../data/Table S3.xlsx", sheet = "progressive motility_cont
mpt_cont$treat <- "control"
mpt_dmso <- readxl::read_excel("../data/Table S3.xlsx", sheet = "progressive motility_2%DM
mpt_dmso$treat <- "DMSO 2%"
mpt_etoh1 <- readxl::read_excel("../data/Table S3.xlsx", sheet = "progressive motility_1%E
```

```r
mpt_etoh1$treat <- "EtOH 1%"
mpt_etoh2 <- readxl::read_excel("../data/Table S3.xlsx", sheet = "progressive motility_2%E
mpt_etoh2$treat <- "EtOH 2%"
mpt <- rbind(mpt_cont, mpt_dmso, mpt_etoh1, mpt_etoh2)
mpt$donor <- as.factor(mpt$donor)
mpt$treat <- as.factor(mpt$treat)
names(mpt) <- c("donor", "time", "prog", "total", "treat")
mpt$prog_frac <- mpt$prog / mpt$total
skimr::skim(mpt)
```

Table 4: Data summary

Name	mpt
Number of rows	64
Number of columns	6
_____	
Column type frequency:	
factor	2
numeric	4
_____	
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
donor	0	1	FALSE	4	6: 16, 7: 16, 8: 16, 9: 16
treat	0	1	FALSE	4	con: 16, DMS: 16, EtO: 16, EtO: 16

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
time	0	1	1.62	1.57	0.00	0.38	1.25	2.50	4.00	
prog	0	1	72.81	29.67	18.00	49.50	68.50	91.00	167.00	
total	0	1	101.83	33.86	47.00	74.75	100.50	124.25	201.00	
prog_frac	0	1	0.70	0.10	0.38	0.64	0.71	0.77	0.92	

There are four donors, four treatments and no missing data.

```
table(mpt$donor, as.factor(mpt$time), mpt$treat)
```

```
, ,  = control


   0 0.5 2 4
6 1   1 1 1
7 1   1 1 1
8 1   1 1 1
9 1   1 1 1

, ,  = DMSO 2%


   0 0.5 2 4
6 1   1 1 1
7 1   1 1 1
8 1   1 1 1
9 1   1 1 1

, ,  = EtOH 1%


   0 0.5 2 4
6 1   1 1 1
7 1   1 1 1
8 1   1 1 1
9 1   1 1 1

, ,  = EtOH 2%


   0 0.5 2 4
6 1   1 1 1
7 1   1 1 1
8 1   1 1 1
9 1   1 1 1
```

The data are balanced with one observation for each treatment, each time and each donor and no missing data.

```r
mpt_m1 <- glmer(cbind(prog, total - prog) ~ treat * time + (treat | donor),
  data = mpt, family = binomial(link = "logit"))
```

```
boundary (singular) fit: see help('isSingular')
```

```r
summary(mpt_m1)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula: cbind(prog, total - prog) ~ treat * time + (treat | donor)
   Data: mpt

     AIC      BIC   logLik deviance df.resid
   481.3    520.2   -222.7    445.3       46

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.4380 -1.0019 -0.0271  0.8145  3.8196

Random effects:
 Groups Name         Variance Std.Dev. Corr
 donor  (Intercept)  0.01305  0.1142
        treatDMSO 2% 0.03076  0.1754   -0.81
        treatEtOH 1% 0.02011  0.1418   -0.98  0.68
        treatEtOH 2% 0.07006  0.2647    0.82 -0.32 -0.91
Number of obs: 64, groups:  donor, 4

Fixed effects:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        1.21141    0.10458  11.583  < 2e-16 ***
treatDMSO 2%       0.06926    0.15108   0.458  0.64665
treatEtOH 1%      -0.37111    0.13993  -2.652  0.00800 **
treatEtOH 2%      -0.51461    0.17695  -2.908  0.00363 **
time              -0.02341    0.03683  -0.636  0.52510
treatDMSO 2%:time -0.09487    0.05162  -1.838  0.06612 .
treatEtOH 1%:time  0.03516    0.05174   0.680  0.49675
treatEtOH 2%:time -0.07634    0.05016  -1.522  0.12800
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) trDMSO2% trEOH1% trEOH2% time   tDMSO2%: tEOH1%:
treatDMSO2% -0.745
treatEtOH1% -0.796  0.566
treatEtOH2% -0.080  0.147   -0.037
time        -0.615  0.427    0.460   0.363
trtDMSO2%:t  0.438 -0.596   -0.328  -0.258  -0.714
trtEtOH1%:t  0.438 -0.304   -0.633  -0.258  -0.712  0.508
trtEtOH2%:t  0.452 -0.314   -0.338  -0.486  -0.734  0.524    0.523
optimizer (Nelder_Mead) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')
```

There is also a singularity in the model fitting here. The model is:

,

$$\text{prog}_i \sim \text{Binomial}(n = 1, \text{prob}_{\text{prog=1}} = \widehat{P})$$

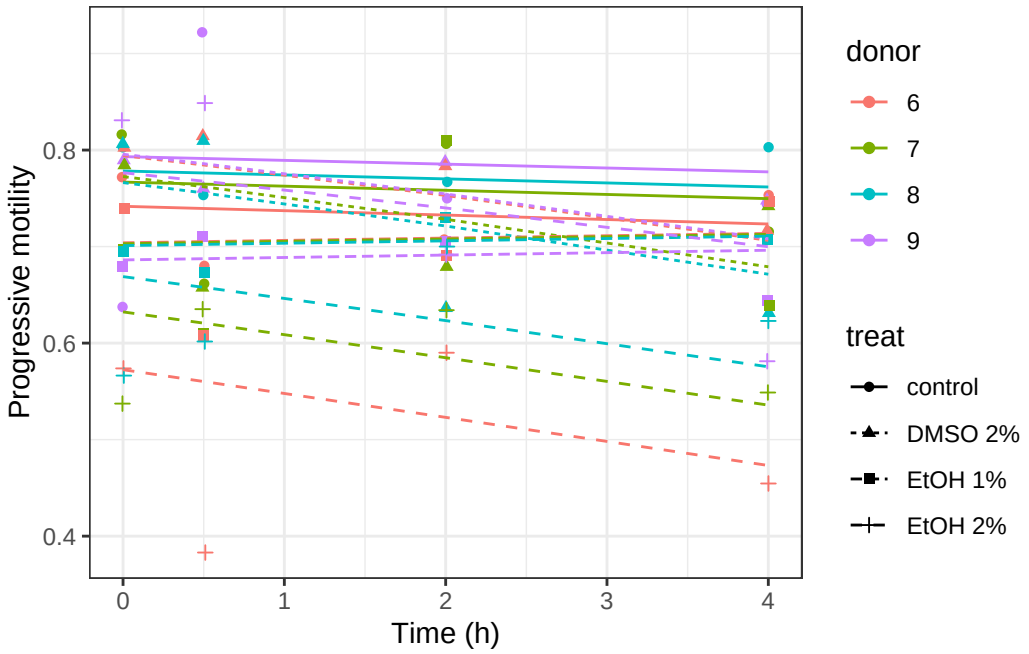
$$\log\left[\frac{\hat{P}}{1-\hat{P}}\right] = \alpha_{j[i]} + \beta_{1j[i]}(\text{treat}_{\text{DMSO 2\%}}) + \beta_{2j[i]}(\text{treat}_{\text{EtOH 1\%}}) + \beta_{3j[i]}(\text{treat}_{\text{EtOH 2\%}}) + \beta_4(\text{time}) + \beta_5(\text{time} \times t$$

,

$$\begin{pmatrix} \alpha_j \\ \beta_{1j} \\ \beta_{2j} \\ \beta_{3j} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\alpha_j} \\ \mu_{\beta_{1j}} \\ \mu_{\beta_{2j}} \\ \mu_{\beta_{3j}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{\alpha_j} & \rho_{\alpha_j\beta_{1j}} & \rho_{\alpha_j\beta_{2j}} & \rho_{\alpha_j\beta_{3j}} \\ \rho_{\beta_{1j}\alpha_j} & \sigma^2_{\beta_{1j}} & \rho_{\beta_{1j}\beta_{2j}} & \rho_{\beta_{1j}\beta_{3j}} \\ \rho_{\beta_{2j}\alpha_j} & \rho_{\beta_{2j}\beta_{1j}} & \sigma^2_{\beta_{2j}} & \rho_{\beta_{2j}\beta_{3j}} \\ \rho_{\beta_{3j}\alpha_j} & \rho_{\beta_{3j}\beta_{1j}} & \rho_{\beta_{3j}\beta_{2j}} & \sigma^2_{\beta_{3j}} \end{pmatrix}\right), \text{ for donor j} = 1, \ldots, \text{J} \tag{3}$$

Here is a plot of this model:

```
ggplot(data = mpt) +
  geom_jitter(aes(x = time, y = prog_frac, shape = treat, col = donor), width = 0.01) +
  geom_line(aes(x = time, y = fitted(mpt_m1), linetype = treat, col = donor)) +
  labs(x = "Time (h)", y = "Progressive motility")
```



Here again, we simplify the model, so that the shift in intercept is the same for each treatment as for total motility, in order to eliminate the singularity.

```
mpt_m2 <- glmer(cbind(prog, total - prog) ~ treat * time + (1 | donor),
  data = mpt, family = binomial(link = "logit"))
anova(mpt_m1, mpt_m2) # Despite the name, it is indeed a LR test
```

```
Data: mpt
Models:
mpt_m2: cbind(prog, total - prog) ~ treat * time + (1 | donor)
mpt_m1: cbind(prog, total - prog) ~ treat * time + (treat | donor)
       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
mpt_m2    9 494.60 514.03 -238.30   476.60
mpt_m1   18 481.32 520.18 -222.66   445.32 31.288  9  0.0002641 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The likelihood ratio test detects significant differences between the full and simplified models at $\alpha = 5\%$. We cannot use the simplest `mpt_m2` model with only a shift in the slope per donor. Keep our complete model, we cannot do much with it, including the calculation of profile or parametric bootstrapped 95%CI on the parameters that do not proceed well.

We could split into three separate models, one for control, one for DMSO, and one for ethanol. Yet, the difference in slopes between the three treatments is what we are looking for, and it is not possible to do it with three separate models.

We need more data to fit such a model with three explanatory variables.

General informations

```
sessionInfo()
```

```
R version 4.1.3 (2022-03-10)
Platform: x86_64-apple-darwin17.0 (64-bit)
Running under: macOS Big Sur/Monterey 10.16

Matrix products: default
LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] ggplot2_3.3.5 lme4_1.1-29   Matrix_1.4-1

loaded via a namespace (and not attached):
 [1] tidyr_1.2.0         jsonlite_1.8.0      splines_4.1.3
 [4] equatiomatic_0.3.1  shiny_1.7.1         assertthat_0.2.1
 [7] highr_0.9           broom.mixed_0.2.9.4 cellranger_1.1.0
[10] yaml_2.3.5          globals_0.14.0      numDeriv_2016.8-1.1
[13] pillar_1.7.0        backports_1.4.1     lattice_0.20-45
[16] glue_1.6.2          digest_0.6.29       promises_1.2.0.1
[19] minqa_1.2.4         colorspace_2.0-3    dfoptim_2020.10-1
[22] htmltools_0.5.2     httpuv_1.6.5        pkgconfig_2.0.3
[25] broom_0.8.0         listenv_0.8.0       purrr_0.3.4
[28] xtable_1.8-4        scales_1.2.0        later_1.3.0
[31] tibble_3.1.6        mgcv_1.8-40         generics_0.1.2
[34] farver_2.1.0        ellipsis_0.3.2      withr_2.5.0
[37] furrr_0.2.3         repr_1.1.4          skimr_2.1.4
[40] cli_3.2.0           magrittr_2.0.3      crayon_1.5.1
```

```
[43] readxl_1.4.0        mime_0.12           evaluate_0.15
[46] fansi_1.0.3         future_1.24.0       parallelly_1.31.0
[49] nlme_3.1-157        MASS_7.3-56         forcats_0.5.1
[52] tools_4.1.3         lifecycle_1.0.1     stringr_1.4.0
[55] munsell_0.5.0       compiler_4.1.3      rlang_1.0.2
[58] grid_4.1.3          nloptr_2.0.0        rstudioapi_0.13
[61] base64enc_0.1-3     labeling_0.4.2      rmarkdown_2.13
[64] boot_1.3-28         gtable_0.3.0        codetools_0.2-18
[67] DBI_1.1.2           R6_2.5.1            knitr_1.38
[70] dplyr_1.0.8         optimx_2021-10.12   fastmap_1.1.0
[73] utf8_1.2.2          stringi_1.7.6       parallel_4.1.3
[76] Rcpp_1.0.8.3        vctrs_0.4.1         tidyselect_1.1.2
[79] xfun_0.30
```