2018 Q1

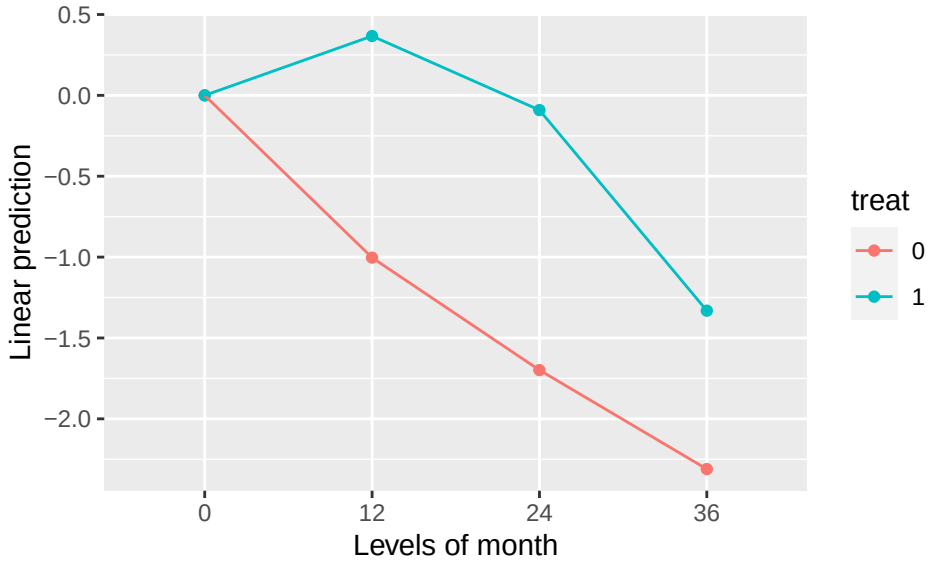
codename

```
library(gmodels)
library(MASS)
library(car)
library(nlme)
library(emmeans)
library(dplyr)
```

(1)

```
d <- read.csv("q1.csv", header = T)
d <- within(d, {
  treat <- as.factor(treat)
  gender <- as.factor(gender)
  race <- as.factor(race)
  scanner <- as.factor(scanner)
  clinic <- as.factor(clinic)
})
```

```
m <- lm(bmd.chg ~ treat * factor(month), data = d)
emmip(m, treat ~ factor(month), style = "factor")
```



For treat0, namely, placebo group, the mean percent change in BMD keeps going down, while for the active drug group, it first goes up, then goes down after month 12. Also, the differences between the two treatment are not the same in the follow-up months.

(2)

This is a longitudinal data, so we need to fit a linear mixed model. The model including all two-way interactions suffers singularity issue. So I then starting with the largest plausible model, and do manual selection. Keep the items with p-value $<$ 0.1. Also, for the same reason as that in 2015 Q1, it is justifiable to use identity as the correlation matrix.

```
d2 <- d %>% filter(month > 0)
m.lme1 <- lme(bmd.chg ~ treat + factor(month) + bmd0 + age + gender +
                          race + scanner + clinic + smoking + bmibl +
                          il6bl + crpbl + cd4bl, random = ~1|id,
              data = d2, method = "ML")
anova(m.lme1)
m.lme2 <- update(m.lme1, fixed = . ~ treat + factor(month) +
                   bmd0 + scanner)
Anova(m.lme2)
m.lme3 <- update(m.lme2, fixed = . ~ treat + factor(month) + scanner)
Anova(m.lme3)
```

```
anova(m.lme1, m.lme2, m.lme3)
```

```
##        Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## m.lme1     1 21 2292.178 2378.566 -1125.089
## m.lme2     2  8 2275.707 2308.616 -1129.853 1 vs 2 9.528595  0.7320
## m.lme3     3  7 2278.206 2307.002 -1132.103 2 vs 3 4.499149  0.0339
```

m.lme4 is the selected model. Checking the residuals, as shown in Figure 1a-1c in the Appendix, no transformation is needed.

```
m.lme4 <- update(m.lme2, method = "REML")
fit.contrast(m.lme4, "treat", coeff = c(1, -1), conf.int = 0.95)
```

```
##                    Estimate Std. Error   t-value    Pr(>|t|)  lower CI
## treat c=( 1 -1 ) -1.378845  0.4326775 -3.186772 0.001673193 -2.232119
##                     upper CI
## treat c=( 1 -1 ) -0.5255704
## attr(,"class")
## [1] "fit_contrast"
```

```
intervals(m.lme4)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                        lower       est.       upper
## (Intercept)     -1.6693674  1.5299496  4.72926666
## treat1           0.5255704  1.3788446  2.23211887
## factor(month)24 -0.9694324 -0.4998415 -0.03025054
## factor(month)36 -1.9549666 -1.3871094 -0.81925218
## bmd0            -6.2463261 -3.2297186 -0.21311109
## scannerType2     0.5472170  1.4066934  2.26616990
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: id
##                     lower     est.    upper
## sd((Intercept)) 2.330979 2.650146 3.013014
##
##  Within-group standard error:
##    lower     est.    upper
## 1.976879 2.153265 2.345389
```

The placebo group has BMD measure 1.39 lower than the active drug group, with the p-value being 0.0015. The 95% CI is [-2.25, -0.54].

(3)

It is the treatment difference after adjusting for `scanner` and `month`, namely, the effect of treatment after excluding the effect of `scanner` and `month`. And in the plot in (1), the mean percentage change in BMD is raw mean, without forcing the difference among `month` to be the same.

```
xtabs(~treat+month, d2)
```

```
##      month
## treat  12  24  36
##     0  92  73  45
##     1 107  84  51
```

(4)

Firstly, there might be model bias in our proposed model. Secondly, augmenting the missing data with fitted values would reduce the variance, and reduce the width of the confidence interval. In summary, the new CI would have shorter coverage, as well as possible biased center point.

(5)

Now the data is grouped under `clinic`, instead of `id`, so we need to fit another linear mixed model. Again, start with the largest plausible model and do manual model selection.

```
d5 <- d %>% filter(month == 0)
m5.lme1 <- lme(bmd ~ age + gender + race + scanner + smoking +
                                  bmibl + il6bl + crpbl + cd4bl,
               random = ~1|clinic, data = d5, method = "ML")
Anova(m5.lme1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: bmd
##           Chisq Df Pr(>Chisq)
## age      1.3372  1  0.2475258
## gender   6.3565  1  0.0116950 *
## race    14.6242  3  0.0021676 **
## scanner 13.2566  1  0.0002716 ***
## smoking  0.2475  1  0.6188063
## bmibl   38.8908  1  4.482e-10 ***
## il6bl    5.1939  1  0.0226660 *
## crpbl    0.9944  1  0.3186631
## cd4bl    0.9088  1  0.3404293
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m5.lme2 <- update(m5.lme1, . ~ gender + race + scanner + bmibl + il6bl)
Anova(m5.lme2)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: bmd
##           Chisq Df Pr(>Chisq)
## gender   7.4305  1  0.0064128 **
## race    15.3174  3  0.0015645 **
## scanner 14.5440  1  0.0001369 ***
## bmibl   39.0579  1  4.114e-10 ***
```

```
## il6bl   11.1122  1  0.0008576 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(m5.lme1, m5.lme2)
```

```
##         Model df       AIC       BIC   logLik   Test  L.Ratio p-value
## m5.lme1     1 14 -338.6556 -288.4867 183.3278
## m5.lme2     2 10 -342.8162 -306.9813 181.4081 1 vs 2 3.839411  0.4282
```

```
m5.lme2 <- update(m5.lme2, method = "REML")
fixef(m5.lme2)
```

```
##  (Intercept)      gender2        race2        race3        race4 scannerType2
##  0.804434743 -0.055813565 -0.057781692 -0.023362277 -0.081163322  0.065781308
##        bmibl        il6bl
##  0.009691822 -0.006214962
```

Based on the estimated effects, (1) female would have lower BMD than male; (2) the ranking of race effects is race4 < race2 < race3 < race1; (3) scanner Type 2 would have higher BMD than Type 1; (4) higher `bmibl` is associated with higher BMD; (5) higher `il6bl` is associated with lower BMD. There is no interaction among the covariates in the selected model, so all of these effects are independent.

```
summary(m5.lme2)
```

```
...
## Random effects:
##  Formula: ~1 | clinic
##         (Intercept)  Residual
## StdDev:  0.01399878 0.1238445
...
```

```
intervals(m5.lme2)$reStruct
```

```
...
## [1] "Random Effects:"
NA
NA
NA
...
```

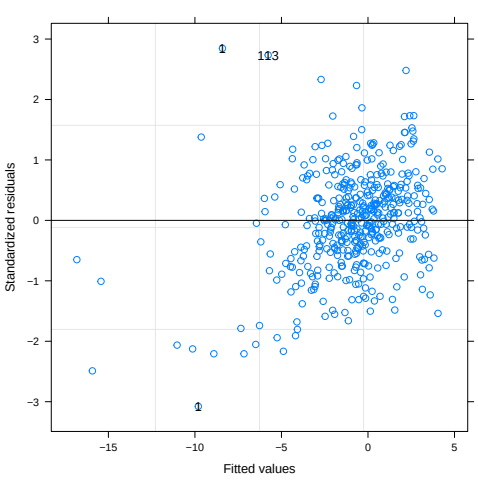
```
(2.504619e-06)^2 / ((2.504619e-06)^2 + 0.1223433^2)
```

```
...
NA
NA
NA
NA
...
```

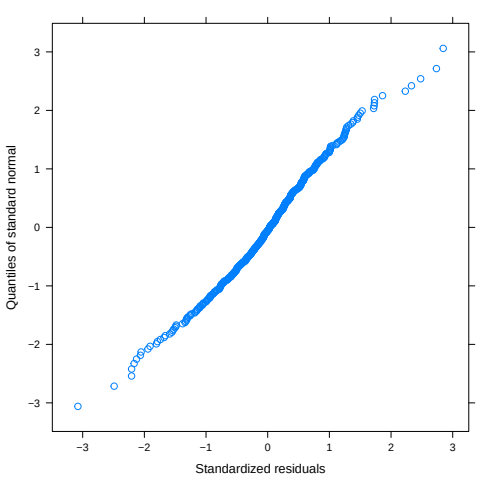
The variability in BMD that is due to the clinic random effect is 4.191051e-10, very small.

Appendix

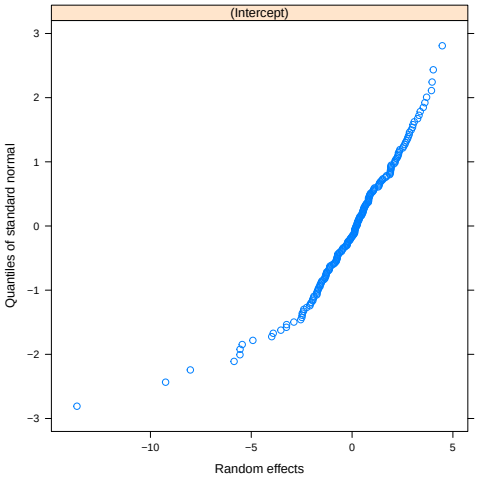
Figures



(a) residual plot of m.lme4



(b) qqnorm of residuals



(c) qqnorm of random effects

Figure 1: Model Diagnostics

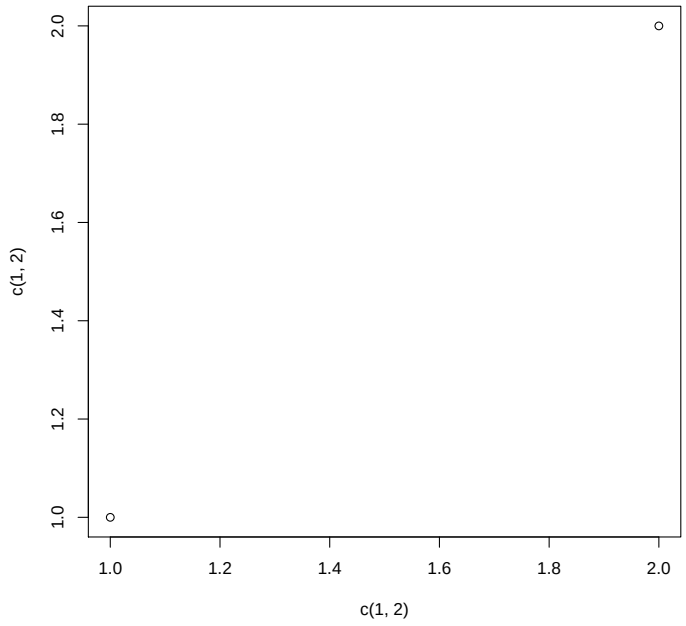


Figure 2: Scatter Plot Age vs. np.chg