2013 Q2

codename

```
library(gmodels)
library(MASS)
# library(car)
```

(a)

Since the patients are randomized with treatment, it would fine to use a model without adjustment to compare the treatment effects.

```
da <- read.table("http://users.stat.umn.edu/~wangx346/artstudy-widedata.txt",
                 header = T)

da <- within(da, {
  treat <- as.factor(treat)
  country <- as.factor(country)
  gender <- as.factor(gender)
  np.chg <- np.8 - 0.5 * (np.0 + np.2)
})

ma <- lm(np.chg ~ treat, data = da)
fit.contrast(ma, "treat", coeff = c(1, -1), conf.int = 0.95)
```

```
##                   Estimate Std. Error  t value    Pr(>|t|)  lower CI  upper CI
## treat c=( 1 -1 ) -2.073679  0.6986437 -2.96815 0.003177036 -3.447172 -0.700186
## attr(,"class")
## [1] "fit_contrast"
```

In the analysis, participants on continuous ART have lower NP score than those on intermittent ART, by 2.07 points(p-value = 0.003). The 95% confidence of the difference is [-3.447172, -0.700186].

(b)

```
pdf("images/scatter_age_np.pdf")
with(da, scatter.smooth(x = age, y = np.chg))
dev.off()
```

As shown in Figure 1 in Appendix, there is a non-linear relationship between `age` and `np.chg`, so I include the quadratic term. Also, for easier interpretation, I center `age` around 40.

```
mb <- lm(np.chg ~ (treat + I(age-40) + I((age - 40)^2) + gender + country +
                     art.pre0 + artdur2)^2, data = da)
stepAIC(mb, scope = list(lower = ~ treat), direction = "backward", trace = F)$call
```

```
## lm(formula = np.chg ~ treat + I(age - 40) + I((age - 40)^2) +
##     gender + country + art.pre0 + treat:I((age - 40)^2) + I(age -
##     40):I((age - 40)^2) + I((age - 40)^2):art.pre0 + gender:country,
##     data = da)
```

Since the interaction of the treatment effect with the quadratic effect of age is included, to get a hierarchical model, I repeat the model selection, forcing the interaction between the treatment effect and the linear effect of age included.

```
stepAIC(mb, scope = list(lower = ~ treat * I(age - 40)),
        direction = "backward", trace = F)$call
```

```
## lm(formula = np.chg ~ treat + I(age - 40) + I((age - 40)^2) +
##     gender + country + art.pre0 + treat:I(age - 40) + treat:I((age -
##     40)^2) + I(age - 40):I((age - 40)^2) + I((age - 40)^2):art.pre0 +
##     gender:country, data = da)
```

Fit the selected model and check the summary results.

```
mb.2 <- lm(np.chg ~ (treat + art.pre0) * (I(age - 40) + I((age - 40)^2)) +
             I((age - 40)^3) + gender * country, data = da)
summary(mb.2)
```

```
...
## Call:
## lm(formula = np.chg ~ (treat + art.pre0) * (I(age - 40) + I((age -
##     40)^2)) + I((age - 40)^3) + gender * country, data = da)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.9221  -3.1022  -0.0879   3.0640  12.7128
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.0433726  0.8284734   1.259  0.20865
## treat2                     1.3147355  0.6827635   1.926  0.05489 .
## art.pre0                   2.3829338  0.8346416   2.855  0.00454 **
## I(age - 40)               -0.7062102  0.0759048  -9.304  < 2e-16 ***
## I((age - 40)^2)           -0.0098514  0.0050513  -1.950  0.05187 .
## I((age - 40)^3)            0.0012366  0.0002283   5.416 1.08e-07 ***
## gender2                    0.9828502  0.9598437   1.024  0.30649
## country2                   1.2801042  0.7851565   1.630  0.10384
## country3                  -0.5521595  0.8225370  -0.671  0.50244
## treat2:I(age - 40)        -0.0155535  0.0551045  -0.282  0.77790
## treat2:I((age - 40)^2)     0.0167229  0.0051727   3.233  0.00133 **
## art.pre0:I(age - 40)      -0.0386780  0.0646655  -0.598  0.55011
## art.pre0:I((age - 40)^2) -0.0107209  0.0056883  -1.885  0.06022 .
## gender2:country2          -3.4958794  1.3707794  -2.550  0.01115 *
## gender2:country3          -1.2171009  1.4334416  -0.849  0.39637
...
```

From the results, we can see that there is no evidence that the treatment effect depends on gender, country, ART use at study entry, and duration of ART use. But the treatment effect does depend on age.

(c)

Due to the existence of the interaction between the treatment effect and age, we cannot simply use `fit.contrast` to compare the two treatment effects. But from the estimated coefficients above, we can see that at every level of `age`, treatment 2, namely, intermittent ART has higher NP score than continuous ART. Also, from the results in (a), we can also conclude that intermittent ART is better(p-value = 0.003).

(d)

i. The treatment effect is confounded with duration of ART use during the study, so the model with all covariates is not suitable for estimating the treatment effect.

ii.

iii. This model doesn't recognize that the treatment effect would change with age.

Appendix

Figures

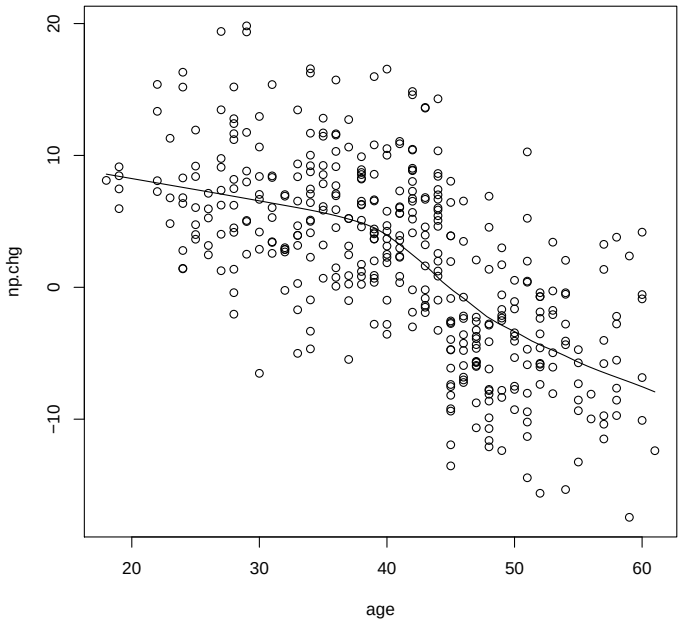

Figure 1: Scatter Plot Age vs. np.chg