

06_update

randy

2022-04-19

1. Redo table 1 with final sample

table1

Variable	testing, N = 457 ¹	training, N = 913 ¹
<i>Genotype</i>		
One allele F508del	151 (33%)	326 (36%)
Others or Unknown	58 (13%)	92 (10%)
Two alleles F508del	248 (54%)	495 (54%)
<i>Gender</i>		
Female	224 (49%)	469 (51%)
Male	233 (51%)	444 (49%)
<i>Race</i>		
Other	21 (4.6%)	41 (4.5%)
White	436 (95%)	872 (96%)
<i>Ethnicity</i>		
Hispanic	18 (4.1%)	25 (2.8%)
Non-Hispanic	419 (96%)	855 (97%)
Unknown	20	33
<i>Visit number</i>	44 (32, 61)	45 (31, 59)
<i>Age at registration, years</i>	3.14 (3.07, 3.23)	3.13 (3.05, 3.22)
<i>Age at end of follow-up, years</i>	13.1 (10.3, 15.9)	12.7 (10.2, 16.0)
<i>Follow-up, years</i>	9.9 (7.2, 12.7)	9.4 (7.0, 12.9)
<i>Height at registration, cm</i>	94.0 (91.7, 97.2)	94.0 (91.2, 96.8)
<i>Weight at registration, kg</i>	14.10 (13.20, 15.30)	14.10 (13.10, 15.30)

¹n (%); Median (IQR)

```
## save pptx -----
## flextable can be saved directly to powerpoints
flextable::save_as_pptx(
  table1,
  path = "figure/01_table1.pptx")
```

2. Refine spline model for training dataset.

Try spline model with 3 inner knots set at percentiles (quartiles) too and compare to the splines you had tried before.

```
train <- here::here("data", "epic_clean_randy.csv") %>%
  read.csv(row.names = 1) %>%
  filter(group == "training")

test <- here::here("data", "epic_clean_randy.csv") %>%
  read.csv(row.names = 1) %>%
  filter(group == "testing")

baseline <- train %>%
  group_by(id) %>%
  arrange(time) %>%
  slice(1L) %>%
  rename(ht_base = ht,
         time_base = time) %>%
  dplyr::select(id, ht_base, time_base)
```

```
# lmeControl
ctrl <- lmeControl(opt='optim')

## the original mod2_2_4 in file "04_model_fit.Rmd"
fit2_0 <- lme(ht ~ bs(time, knots = c(10, 15),
                     degree = 3,
                     intercept = FALSE) * sex + genotype - 1,
             control = ctrl,
             random = ~ 1 | id,
             data = train)

fit2_1 <- lme(ht ~ bs(time, knots = c(10, 12),
                     degree = 3,
                     intercept = FALSE) * sex + genotype - 1,
             control = ctrl,
             random = ~ 1 | id,
             data = train)

fit2_2 <- lme(ht ~ bs(time, df = 5,
                     degree = 3,
                     intercept = FALSE) * sex + genotype - 1,
             control = ctrl,
             random = ~ 1 | id,
             data = train)
```

```
fit3_0 <- lme(ht ~ bs(time, df = 6,
                    degree = 3,
                    intercept = FALSE) * sex + genotype - 1,
            control = ctrl,
            random = ~ 1 | id,
            data = train)
```

```
get_tag <- function(fit, data1 = train) {
  tidy <- broom.mixed::tidy(fit)
  glance <- broom.mixed::glance(fit)
  augment <- broom.mixed::augment(fit) %>%
    dplyr::select(time, .fitted, sex) %>%
    unique()

  plot1 <- ggplot() +
    geom_smooth(data = data1,
               aes(x = time, y = ht,
                   group = sex, color = sex)) +
    geom_smooth(data = augment,
               aes(x = time, y = .fitted,
                   group = sex),
               linetype = 2) +
    xlab("Age") +
    theme_bw()

  return(list(tidy = tidy,
              glance = glance,
              augment = augment,
              plot = plot1))
}
```

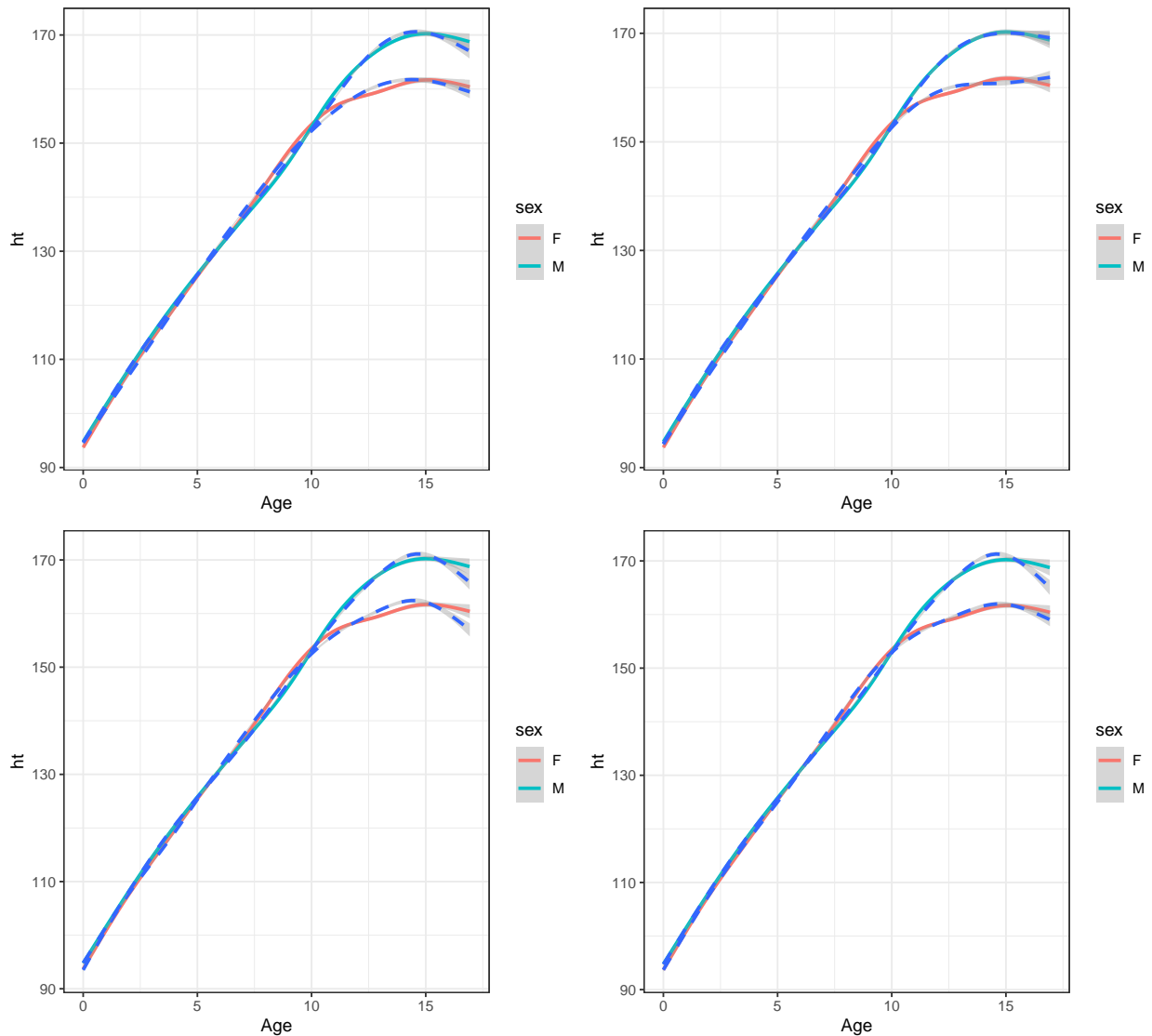
```
tag2_0 <- get_tag(fit = fit2_0)
tag2_1 <- get_tag(fit = fit2_1)
tag2_2 <- get_tag(fit = fit2_2)
tag3_0 <- get_tag(fit = fit3_0)

glance_all <- map_dfr(list(tag2_0, tag2_1,
                          tag2_2, tag3_0),
                     "glance") %>%
  mutate(model = c("fit2_0", "fit2_1", "fit2_2", "fit3_0")) %>%
  dplyr::select(model, everything())

glance_all
```

```
## # A tibble: 4 x 5
##   model  sigma  logLik    AIC    BIC
##   <chr>  <dbl>   <dbl>  <dbl>  <dbl>
## 1 fit2_0  2.47 -104493. 209019. 209158.
## 2 fit2_1  2.44 -103942. 207916. 208055.
## 3 fit2_2  2.46 -104212. 208456. 208595.
## 4 fit3_0  2.44 -103908. 207851. 208007.
```

3. plot the predicted/expected means.



4. For brokenstick model try time for a anchoring of 10 and 12

```
bks1 <- brokenstick(formula = ht ~ time | id,
                    data = train,
                    subjid = train$id,
                    knots = c(10, 15))
est1_knots <- predict(bks1,
                     x = "knots",
                     group = train$id)
est1_all <- predict(bks1,
                   group = train$id)

bks2 <- brokenstick(formula = ht ~ time | id,
                    data = train,
```

```

      subjid = train$id,
      knots = c(10, 12))
est2_knots <- predict(bks2,
      x = "knots",
      group = train$id)
est2_all <- predict(bks2,
      group = train$id)

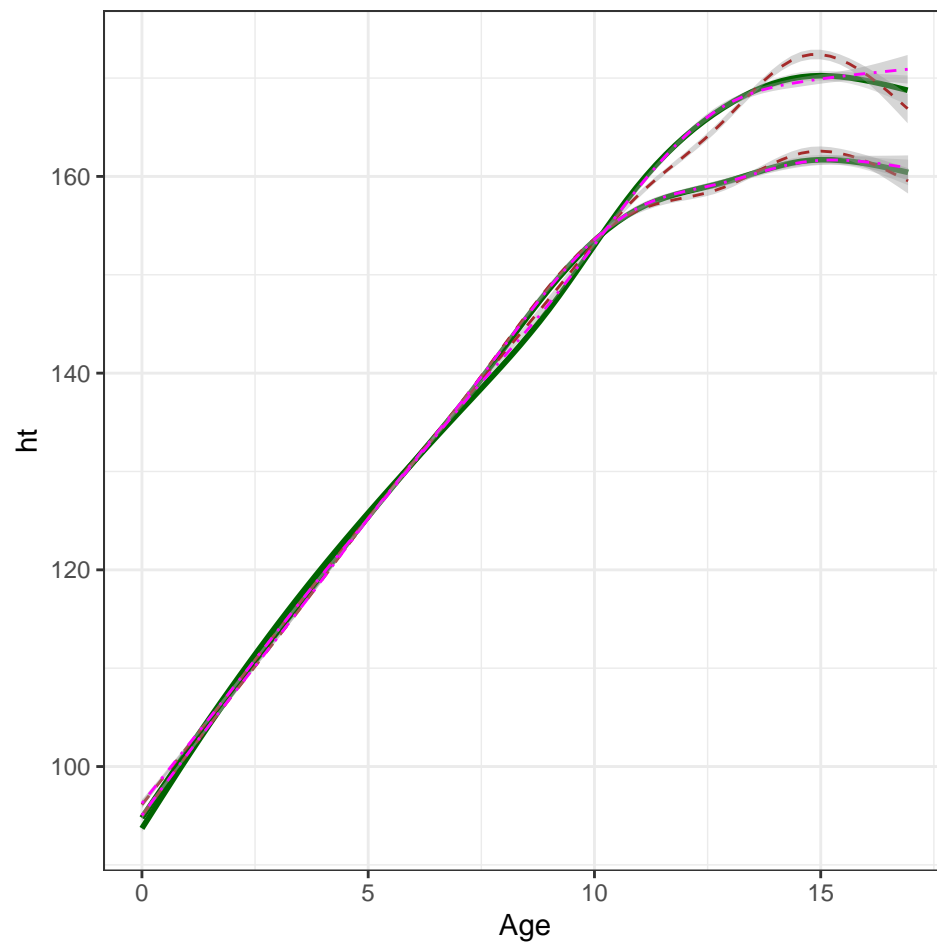
```

```

plot_bks1 <- ggplot() +
  geom_smooth(data = train,
    aes(x = time, y = ht,
      group = sex),
    alpha = 0.3,
    size = 1,
    color = "darkgreen") +
  geom_smooth(data = est1_all,
    aes(x = time, y = .pred,
      group = sex),
    color = "brown",
    linetype = 2,
    size = 0.5) +
  geom_smooth(data = est2_all,
    aes(x = time, y = .pred,
      group = sex),
    color = "magenta",
    linetype = 4,
    size = 0.5) +
  xlab("Age") +
  theme_bw()

plot_bks1

```



5. Make progress, as able with the latex file of the paper. Start drafting a description of the dataset using Moss and Rice's papers: one paragraph to describe the study and a second to include descriptive statistics.