Biological Ages

A New Perspective on Aging

Delin Zhao

May 10, 2022

Introduction

Background

It has been widely accepted that the aging process is the major risk factor for disease and death. However, the aging rate varies for different people at the same chronological age (CA). The concept of biological age (BA) was thus developed to assess the true aging rate.

Contributions

Our study developed a novel nonparametric procedure to select age-related features. Specifically, we utilized questionnaire data, including aspects of diseases, organ functions, cognitive and physical tests, and living habits to estimate the BAs from the selected features. Several rank-based robust aging indexes were then computed, on which we conducted a cluster analysis to demonstrate the usefulness of the constructed BAs.

Exploratory Data Analysis

Data Description

A total of $17,359$ records (46% males) aged between 20 and 90 years were considered. Specific to our study, in total 57 features were included. These were classified into four categories, namely, disease-related, organ-related, cognitive and physical-related and habit-related. we use a random 75% of the observations to serve as the training dataset. The remaining 25% will serve as the test dataset.

- Since a large proportion of features had sex-specific effects, the following construction of BAs was conducted for male and female groups separately.

```
library(dplyr)
library(tidyverse)
library(labelled)
library(data.table)
library(corrplot)
library(ggplot2)
```

```
library(ggpubr)
library(XICOR)
library(FOCI)
library(randomForest)
library(ComplexHeatmap)
source("plot_foci.R")
source("plot_age.R")
df <- readRDS("BJdata20220404.RDS")
question <- readRDS("question.RDS")


subset <- df %>% select(A2, A4, C2_new, C4_new, C5_new:C10_new, C8,
                        B11_new, B12:B13, B17_new:B18_new, B15.2_new:B15.9_new,
                        J1, K1, K2:K5.B, K2.2, K4.2, K6.1:K6.2, K10, K11.A, K12.A,
                        K7.A.1:K7.2.B.1, K8.1,
                        D1.1, D1.10, D1.2:D1.9, D1.11:D1.17)

subset <- subset %>% filter(A4 <= 90, A4 >= 20) %>% unlabelled()

subset_male <- subset %>% filter(A2==1) %>% select(-A2)
subset_female <- subset %>% filter(A2==2) %>% select(-A2)

set.seed(100)
train_test_female <- sample(1:nrow(subset_female), 0.75*nrow(subset_female))
train_female <- subset_female[train_test_female, ]
test_female <- subset_female[-train_test_female, ]
set.seed(100)
train_test_male <- sample(1:nrow(subset_male), 0.75*nrow(subset_male))
train_male <- subset_male[train_test_male, ]
test_male <- subset_male[-train_test_male, ]
```

Preliminary Analysis

- We simply deleted the records containing missing values or outliers.

```
preprocessing <- function(df){
  df <- df %>% unlabelled()
  for (i in grep("^D", colnames(df), value = TRUE)) {
    df[is.na(df[, i]), i] <- 2
  }
  for(i in colnames(df)){
    x <- getElement(df, i)
    ind <- rep(TRUE, length(x))
    if(!is.null(attr(x, "labels"))) ind <- (x %in% attr(x, "labels"))
    if(!is.null(levels(x))) ind <- (x %in% levels(x))
    df <- df[ind, ]
    df[, i] <- as.numeric(x[ind])
  }
  df <- df %>% na.omit()
```

```
  return(df)
}
```

Age-related Features

```
subset_Dfemale <- select(train_female, A4, starts_with("D")) %>%
  preprocessing()
subset_byD <- tibble(rep(subset_Dfemale$A4, ncol(subset_Dfemale)-1),
                     as.factor(c(as.matrix(subset_Dfemale[, -1]))),
                     rep(substr(question$colname[match(colnames(subset_Dfemale[-1]),
                                                       question$question_id)], 1, 4),
                         each = nrow(subset_Dfemale)))
setnames(subset_byD, new = c("age", "level", "disease"))
subset_byD$variable <-
  factor(subset_byD$disease,
         levels = substr(question$colname[match(colnames(subset_Dfemale[-1]),
                                                question$question_id)], 1, 4))

box_byD <- ggplot(data = subset_byD, aes(y = age, x = level)) + geom_boxplot(fill = "gray")
box_byD + facet_wrap(~disease) + theme_bw() +
  theme(axis.line=element_line(size=0.5,colour="black")) +
  labs(x = "Level", y = "Age")
```

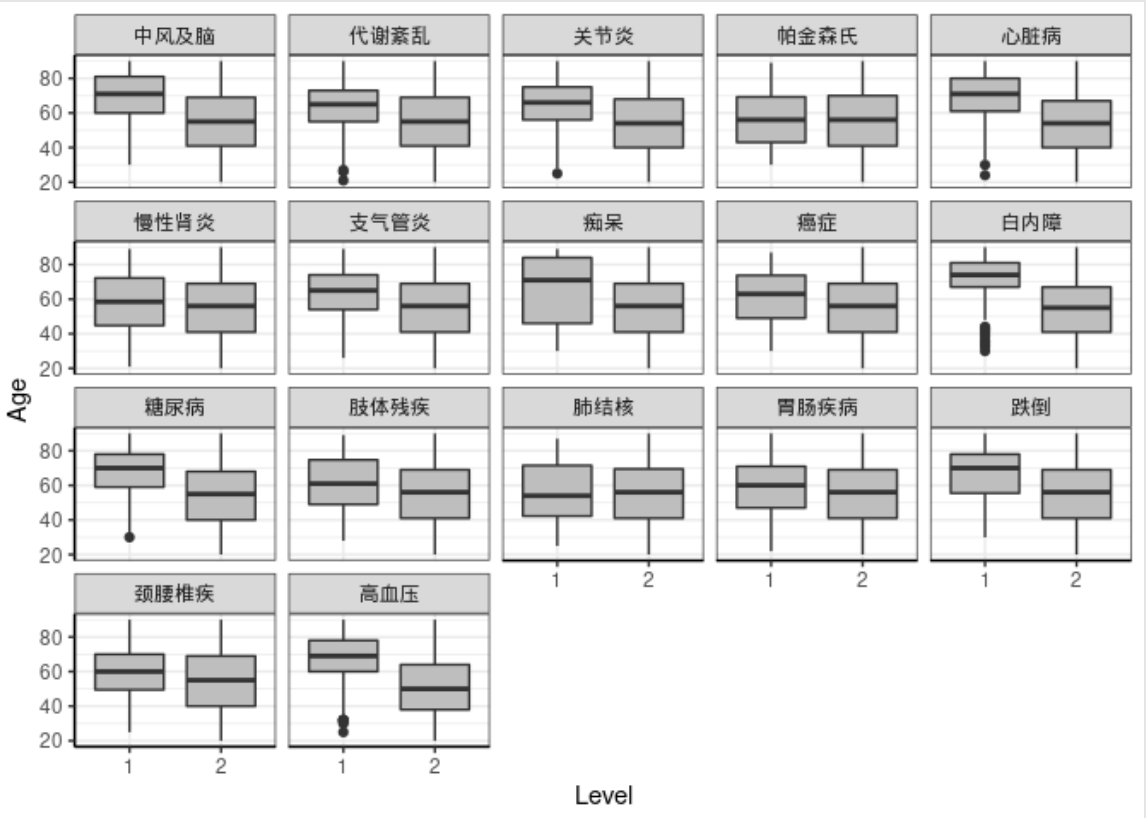

Figure: boxplots of the chronological age versus the level of disease, where 1 represents illness.

- Take disease-related features as an example, it can be observed that most diseases are strongly age-dependent.

Inter-correlations among Features

```
subset_Kfemale <- select(train_female, A4, starts_with("K")) %>%
  preprocessing()
par(mfrow=c(2,2), bg = "lightgray")
subset_Kfemale <- subset_Kfemale %>% apply(2, function(x) (x-mean(x))/sd(x)) %>%
         as_tibble()
subset_Kfemale <- subset_Kfemale %>% select(-A4, -K2.2, -K4.2)
corrplot(abs(cor(subset_Kfemale)), is.corr = F, diag = FALSE, tl.col = "black", tl.cex =
         0.75)
corrplot(abs(cor(subset_Kfemale, method = "spearman")), is.corr = F, diag = FALSE,
         tl.col = "black", tl.cex = 0.75)
corrplot(xicor(subset_Kfemale), is.corr = F, diag = FALSE, tl.col = "black", tl.cex = 0.75)
```

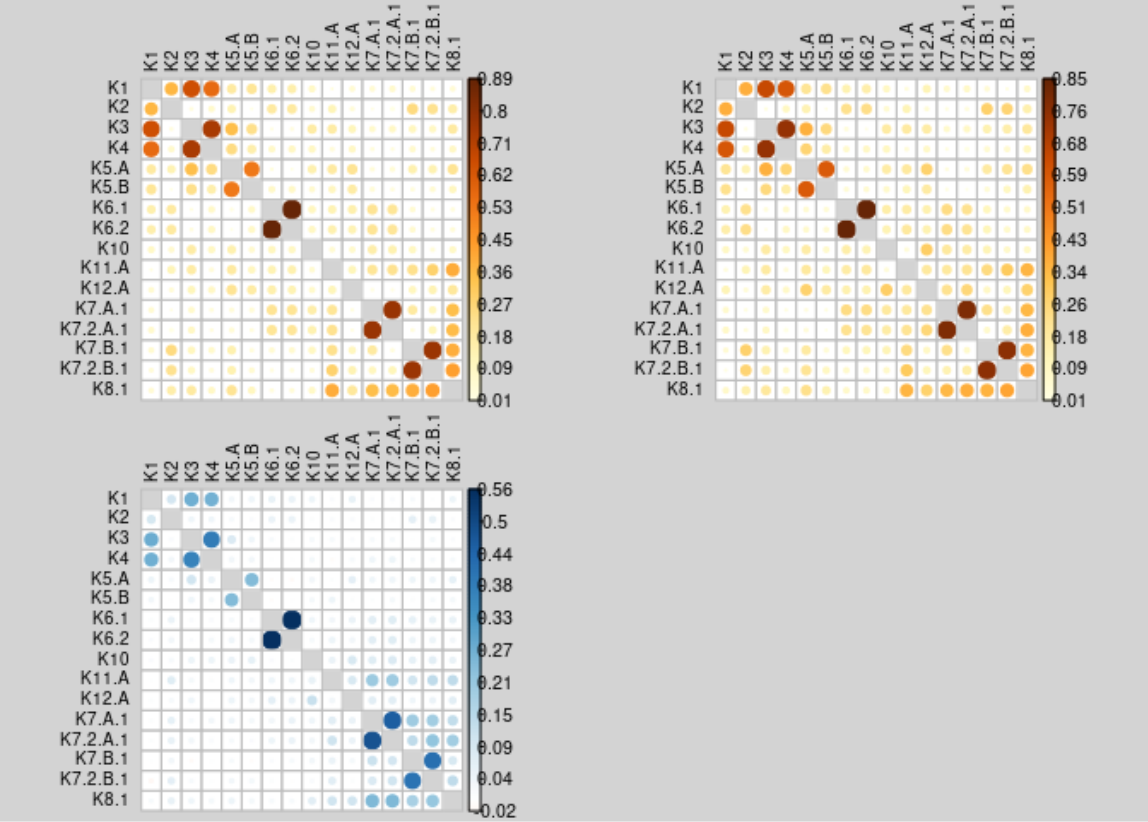

Figure: correlation plots among cognitive-and-physical-related features. The plots on the top left and top right are based on Pearson's r (Pearson 1895) and Spearman's ρ (Spearman 1904), which detect linear and monotonic relationships, respectively. The plot on the bottom left displays Chatterjee's ξ coefficient (Chatterjee 2021), which, however, can capture arbitrary forms of dependence between two variables.

- Take cognitive-and-physical-related features as an example, it can be observed that most indexes are reasonably inter-correlated.
- These two observations inspired us to identify age-related features and screen out redundant ones simultaneously, so as to achieve better model interpretation and lower computational complexity and avoid over-fitting.

Methods

- A conditional dependence coefficient (Azadkia and Chatterjee 2021):

$$T = T(Y, \mathbf{Z} \mid \mathbf{X}) := \frac{\int \mathrm{E}[\mathrm{Var}\{\mathrm{P}(Y \geq t \mid \mathbf{Z}, \mathbf{X}) \mid \mathbf{X}\}] d\mu(t)}{\int \mathrm{E}[\mathrm{Var}\{1(Y \geq t) \mid \mathbf{X}\}] d\mu(t)}.$$

  When Y is not almost surely equal to a measurable function of $\mathbf{X}$ (when $p = 0$, this means that Y is not almost surely a constant), T is well-defined and $0 \leq T \leq 1$. Moreover, $T = 0$ if and only if Y and $\mathbf{Z}$ are conditionally independent given $\mathbf{X}$, and $T = 1$ if and only if Y is almost surely equal to a measurable function of $\mathbf{Z}$ given $\mathbf{X}$. When $p = 0$, conditional independence given $\mathbf{X}$ simply means unconditional independence. The estimate of T is denoted as T_n, whose details are omitted here.

- We implemented a nonparametric forward stepwise algorithm that uses the conditional dependence coefficient at each step to select a subset of features: First, choose j_1 to be the index j that maximizes $T_n(Y, X_j)$. Having obtained $j_1, \ldots, j_k$, choose j_{k+1} to be the index $j \notin \{j_1, \ldots, j_k\}$ that maximizes $T_n(Y, X_j \mid X_{j_1}, \ldots, X_{j_k})$. We remark that features with redundant information were filtered out in the process of feature ordering.

- To be in line with this forward stepwise manner, we utilized random forest modeling for the generation of multiple BAs with the selected features.

- To deeply study the characteristics of the aging processes of different aspects, we calculated an aging index with the constructed BAs, which took the difference between ranks of BA and CA and divided it by the sample size, i.e., the possible maximum rank difference. This aging index represented the relative aging acceleration or deceleration. For instance, if the rank of BA is larger than that of CA, the person is aging faster than the population average.

Main Results

Disease Ages

Feature Selection

```
par(mfrow=c(1,1))
subset_Dfemale <- select(train_female, A4, starts_with("D")) %>%
  preprocessing()
test_Dfemale <- select(test_female, A4, starts_with("D")) %>%
  preprocessing()
X <- apply(subset_Dfemale[, -1], 2, function(x) (x-mean(x))/sd(x))
Y <- subset_Dfemale$A4

set.seed(101)
foci_Dfemale <- foci(Y, X, numCores = 1, stop = FALSE)

subset_Dmale <- select(train_male, A4, starts_with("D")) %>%
```

```
  preprocessing()
test_Dmale <- select(test_male, A4, starts_with("D")) %>%
  preprocessing()
X <- apply(subset_Dmale[, -1], 2, function(x) (x-mean(x))/sd(x))
Y <- subset_Dmale$A4

set.seed(103)
foci_Dmale <- foci(Y, X, numCores = 1, stop = FALSE)


plot_foci(foci_Dfemale, foci_Dmale, ind = c(10, 11))
```

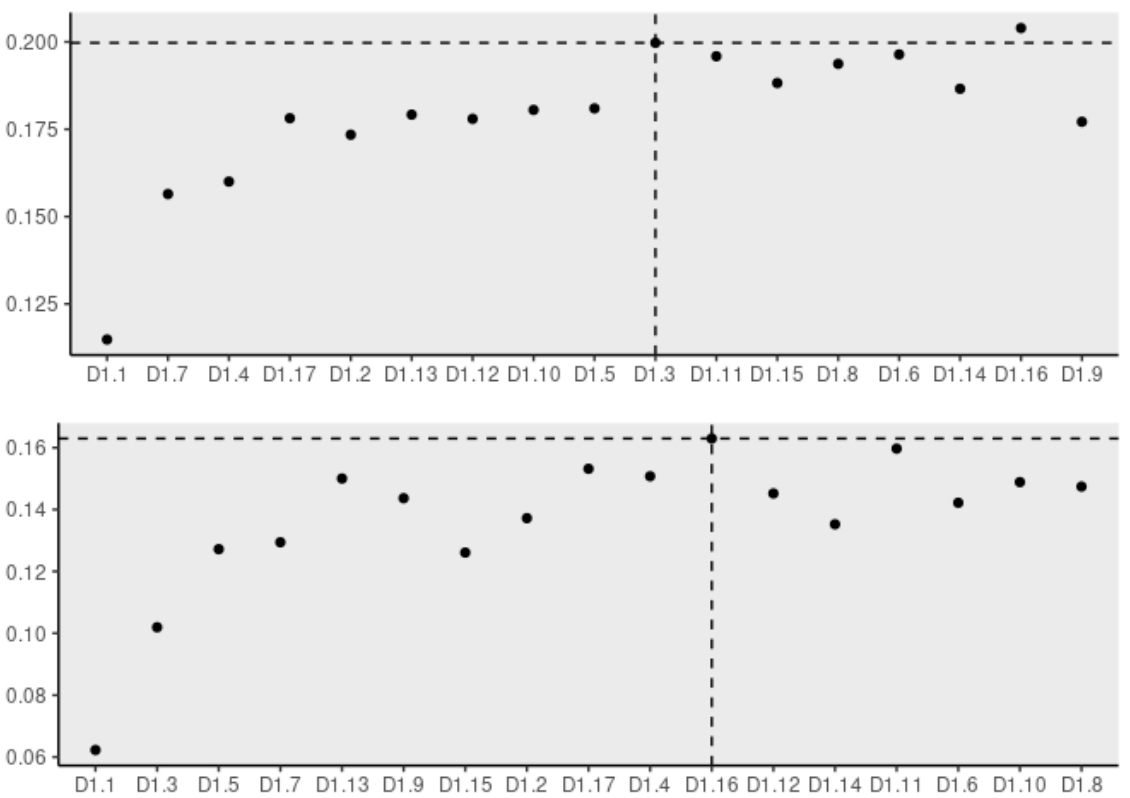

Figure: feature ordering by conditional independence (abbr. FOCI) for disease-related features on female (upper) and male (lower) separately.

```
cbind(colnames(subset_Dfemale[-1]),
      question$colname[match(colnames(subset_Dfemale[-1]), question$question_id)])
```

```
##       [,1]    [,2]
##  [1,] "D1.1"  "高血压"
##  [2,] "D1.10" "颈腰椎疾病"
##  [3,] "D1.2"  "糖尿病"
##  [4,] "D1.3"  "心脏病"
##  [5,] "D1.4"  "中风及脑血管疾病"
##  [6,] "D1.5"  "支气管炎，肺气肿，哮喘或肺炎"
##  [7,] "D1.6"  "肺结核"
##  [8,] "D1.7"  "白内障"
```

```
##  [9,] "D1.8"  "慢性肾炎"
## [10,] "D1.9"  "癌症"
## [11,] "D1.11" "胃肠疾病"
## [12,] "D1.12" "帕金森氏病"
## [13,] "D1.13" "跌倒"
## [14,] "D1.14" "关节炎"
## [15,] "D1.15" "痴呆"
## [16,] "D1.16" "肢体残疾"
## [17,] "D1.17" "代谢紊乱（血糖、血脂、血尿酸水平升高）"
```

- Features selected by both: 高血压、糖尿病、心脏病、中风及脑血管疾病、支气管炎，肺气肿，哮喘或肺炎、白内障、跌倒、代谢紊乱。
- Features selected by neither: 肺结核、慢性肾炎、胃肠疾病、关节炎。

BAs Construction

```
set.seed(102)
select_Dfemale <- foci_Dfemale$selectedVar$names[1:10]
RF_Dfemale <- randomForest(A4~., data=subset_Dfemale[, c("A4", select_Dfemale)],
                     tree=400)


set.seed(104)
select_Dmale <- foci_Dmale$selectedVar$names[1:11]
RF_Dmale <- randomForest(A4~., data=subset_Dmale[, c("A4", select_Dmale)],
                     tree=400)


plot_age(test_Dfemale$A4, predict(RF_Dfemale, test_Dfemale),
         test_Dmale$A4, predict(RF_Dmale, test_Dmale))
```

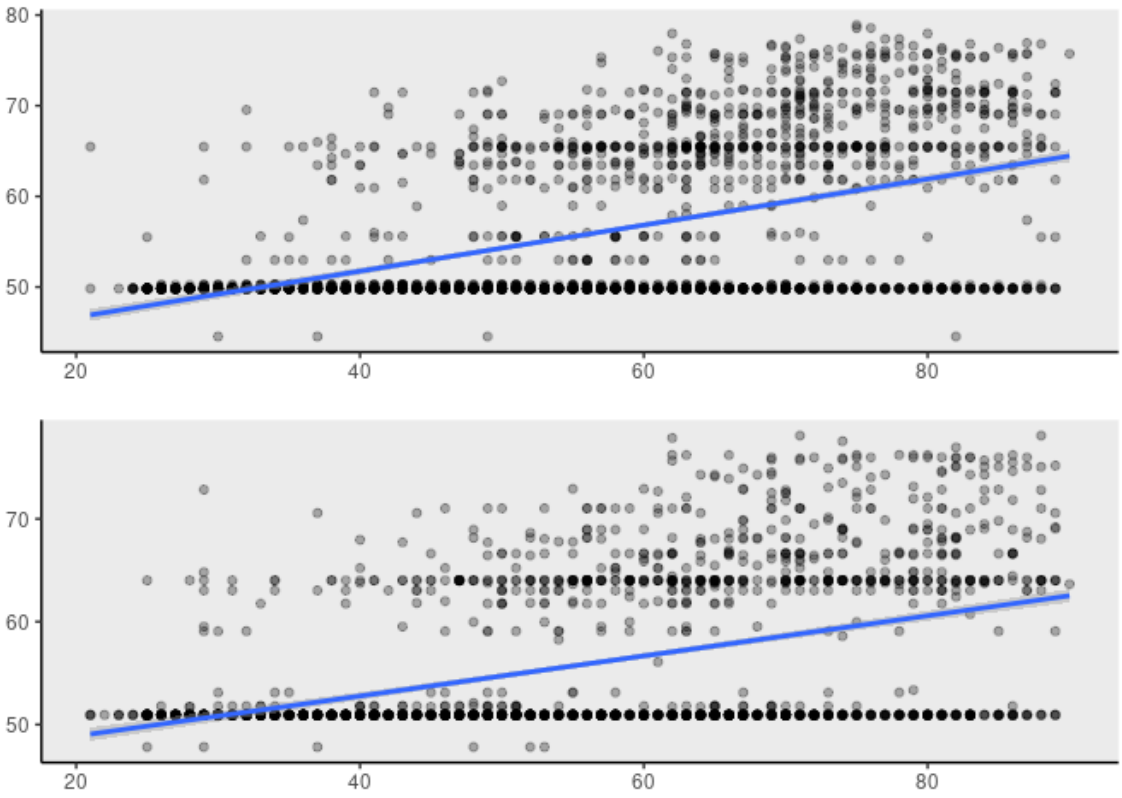

Figure: the disease ages in the vertical axis versus the chronological ages in the horizontal axis, superimposed with the best fitting straight line, on female (upper) and male (lower) separately.

```
cor_Dfemale <- rbind(apply(test_Dfemale[, -1], 2, function(x) cor(test_Dfemale[, 1], x)),
                     apply(test_Dfemale[, -1], 2, function(x)
                       cor(predict(RF_Dfemale, test_Dfemale), x,)))
rownames(cor_Dfemale) <- c("CA", "BA")
cor_Dmale <- rbind(apply(test_Dmale[, -1], 2, function(x) cor(test_Dmale[, 1], x)),
                   apply(test_Dmale[, -1], 2, function(x)
                     cor(predict(RF_Dmale, test_Dmale), x)))
rownames(cor_Dmale) <- c("CA", "BA")
cor_D <- rbind(cor_Dfemale, cor_Dmale)
select_D <- matrix(1, nrow = nrow(cor_D), ncol = ncol(cor_D))
select_D[2, match(select_Dfemale, colnames(cor_D))] <- 0
select_D[4, match(select_Dmale, colnames(cor_D))] <- 0
rownames(select_D) <- rownames(cor_D)
colnames(cor_D) <- question$colname[match(colnames(cor_D),
                                          question$question_id)]
colnames(select_D) <- colnames(cor_D)
corrplot(-cor_D, p.mat = select_D, sig.level = 0.05, insig = "label_sig", pch.cex = 1,
         is.corr = F, tl.col = "black", tl.cex = 0.75, cl.pos = "b", bg = "lightgrey")
```

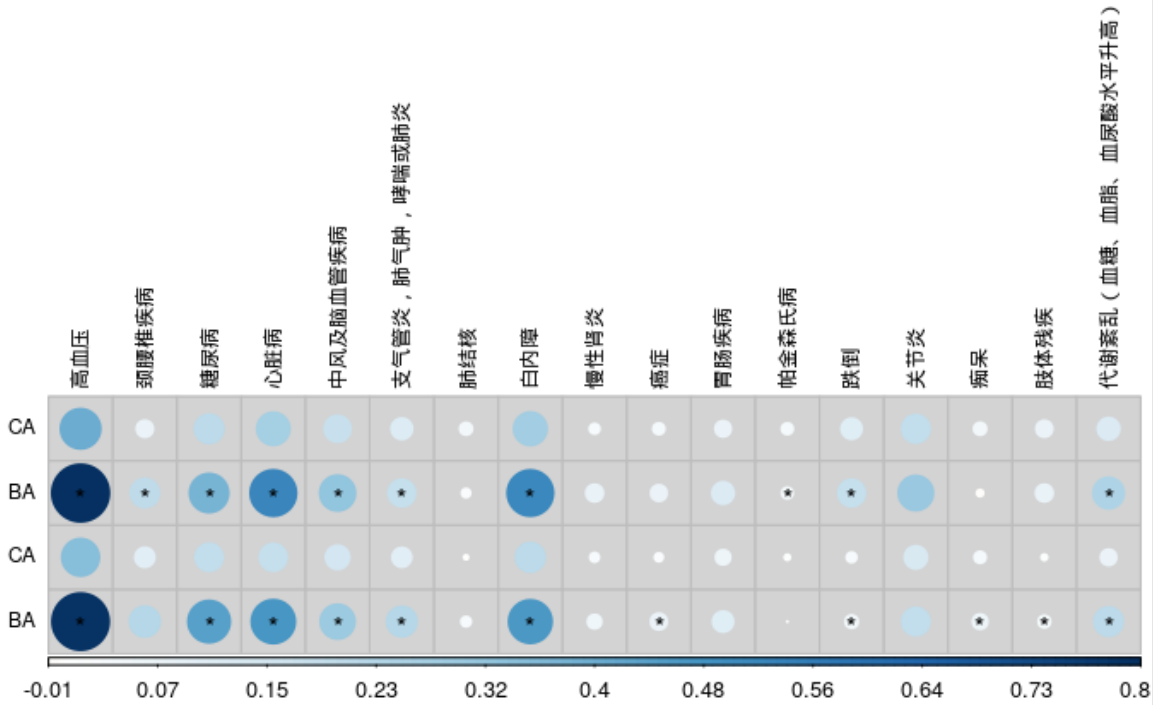

Figure: Pearson correlations between chronological ages or disease ages and disease-related features, where "*" marks the features selected by FOCI, on female (upper) and male (lower) separately.

- We can view disease age as a parsimonious summary of age-related diseases. Due to inter-correlations among variables, some features screened out by FOCI (due to marginally high but conditionally low correlations with age) were still reasonably correlated with disease age.

Other Biological Ages

Organ Ages

```
subset_Cfemale <- select(train_female, A4, starts_with("C")) %>%
  preprocessing()
test_Cfemale <- select(test_female, A4, starts_with("C")) %>%
  preprocessing()
X <- apply(subset_Cfemale[, -1], 2, function(x) (x-mean(x))/sd(x))
Y <- subset_Cfemale$A4

set.seed(105)
foci_Cfemale <- foci(Y, X, numCores = 1, stop = FALSE)


subset_Cmale <- select(train_male, A4, starts_with("C")) %>%
  preprocessing()
test_Cmale <- select(test_male, A4, starts_with("C")) %>%
  preprocessing()
X <- apply(subset_Cmale[, -1], 2, function(x) (x-mean(x))/sd(x))
Y <- subset_Cmale$A4
```

```
set.seed(107)
foci_Cmale <- foci(Y, X, numCores = 1, stop = FALSE)


plot_foci(foci_Cfemale, foci_Cmale, ind = c(5, 5))
```

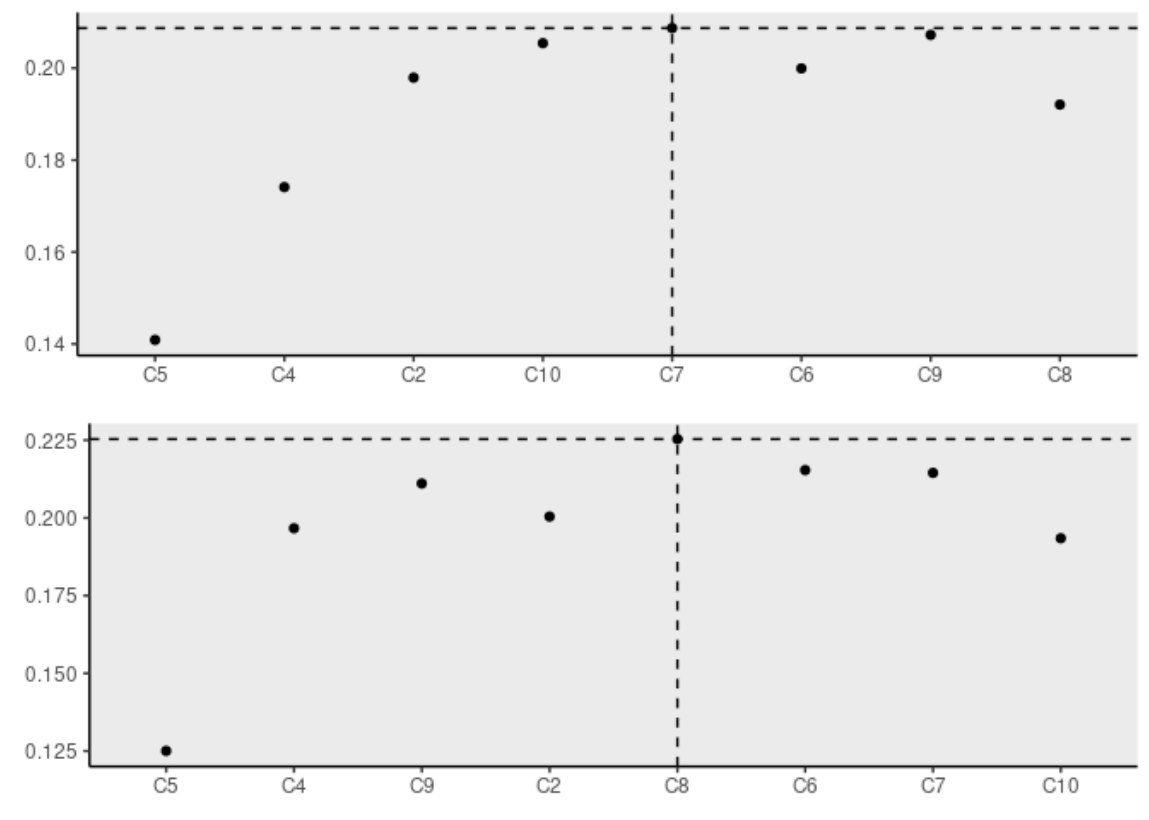

Figure: FOCI for organ-related features on female (upper) and male (lower) separately.

```
cbind(colnames(subset_Cfemale[-1]),
      question$colname[match(colnames(subset_Cfemale[-1]), question$question_id)])
```

```
##      [,1]       [,2]
## [1,] "C2_new"   "视力是否减退"
## [2,] "C4_new"   "听力是否减退"
## [3,] "C5_new"   "牙齿是否缺失"
## [4,] "C6_new"   "是否觉得自己记忆力下降"
## [5,] "C7_new"   "食欲如何"
## [6,] "C9_new"   "是否大便失禁"
## [7,] "C10_new" "是否尿失禁"
## [8,] "C8"       "是否排便困难"
```

- 5 out of 8 features were selected for both cases.
- Features selected by both: 视力是否减退、听力是否减退、牙齿是否缺失。
- Features selected by neither: 是否觉得自己记忆力下降。

```r
set.seed(106)
select_Cfemale <- foci_Cfemale$selectedVar$names[1:5]
RF_Cfemale <- randomForest(A4~., data=subset_Cfemale[, c("A4", select_Cfemale)],
                     tree=400)

set.seed(108)
select_Cmale <- foci_Cmale$selectedVar$names[1:5]
RF_Cmale <- randomForest(A4~., data=subset_Cmale[, c("A4", select_Cmale)],
                     tree=400)

plot_age(test_Cfemale$A4, predict(RF_Cfemale, test_Cfemale),
         test_Cmale$A4, predict(RF_Cmale, test_Cmale))
```

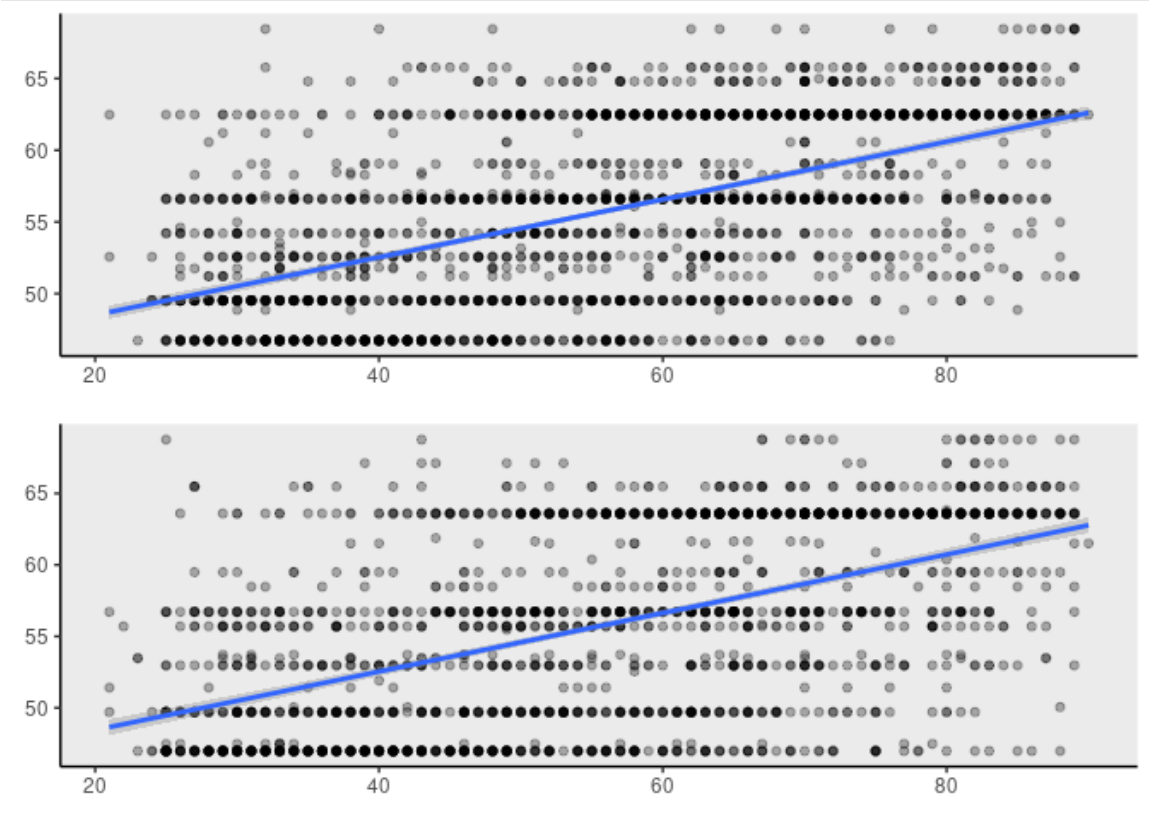

Figure: the organ ages in the vertical axis versus the chronological ages in the horizontal axis, superimposed with the best fitting straight line, on female (upper) and male (lower) separately.

Cognitive and Physical Ages

- Cognitive-related feature: 认知功能总得分。
- Physical-related features:
    - 体重、身高、体质指数；腰围、臀围、腰臀比。
    - 血压（收缩压、舒张压）。
    - 握力（左手、右手）。
    - 日常步速、日常步幅、最快步速、最快步幅。
    - 起立行走、闭眼双脚提踵站立、起坐测试、光反应测试（手）。

```
subset_Kfemale <- select(train_female, A4, J1, starts_with("K")) %>%
  preprocessing()
test_Kfemale <- select(test_female, A4, J1, starts_with("K")) %>%
  preprocessing()
X <- apply(subset_Kfemale[, -1], 2, function(x) (x-mean(x))/sd(x))
Y <- subset_Kfemale$A4

set.seed(109)
foci_Kfemale <- foci(Y, X, numCores = 1, stop = FALSE)


subset_Kmale <- select(train_male, A4, J1, starts_with("K")) %>%
  preprocessing()
test_Kmale <- select(test_male, A4, J1, starts_with("K")) %>%
  preprocessing()
X <- apply(subset_Kmale[, -1], 2, function(x) (x-mean(x))/sd(x))
Y <- subset_Kmale$A4

set.seed(111)
foci_Kmale <- foci(Y, X, numCores = 1, stop = FALSE)


plot_foci(foci_Kfemale, foci_Kmale, ind = c(11, 9))
```

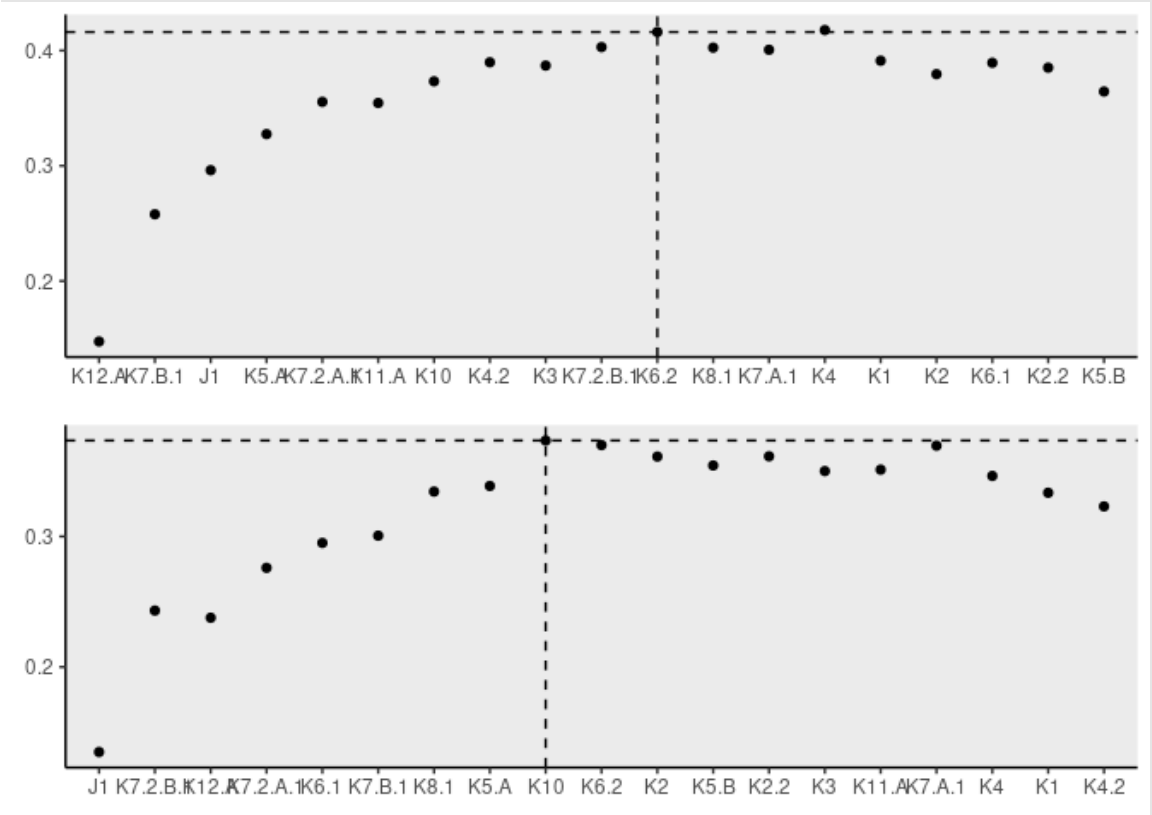

Figure: FOCI for cognitive-and-physical-related features on female (upper) and male (lower) separately.

- Features selected (11 for female, 9 for male):

- In common: 认知功能总得分；收缩压；日常步幅、最快步速、最快步幅；起立行走、闭眼双脚提踵站立、光反应测试（手）。
- Female selected 右手握力，while male selected 左手握力。
- Two more features selected by female: 腰围、腰臀比。

```
set.seed(110)
select_Kfemale <- foci_Kfemale$selectedVar$names[1:11]
RF_Kfemale <- randomForest(A4~., data=subset_Kfemale[, c("A4", select_Kfemale)],
                     tree=400)


set.seed(112)
select_Kmale <- foci_Kmale$selectedVar$names[1:9]
RF_Kmale <- randomForest(A4~., data=subset_Kmale[, c("A4", select_Kmale)],
                     tree=400)


plot_age(test_Kfemale$A4, predict(RF_Kfemale, test_Kfemale),
         test_Kmale$A4, predict(RF_Kmale, test_Kmale))
```

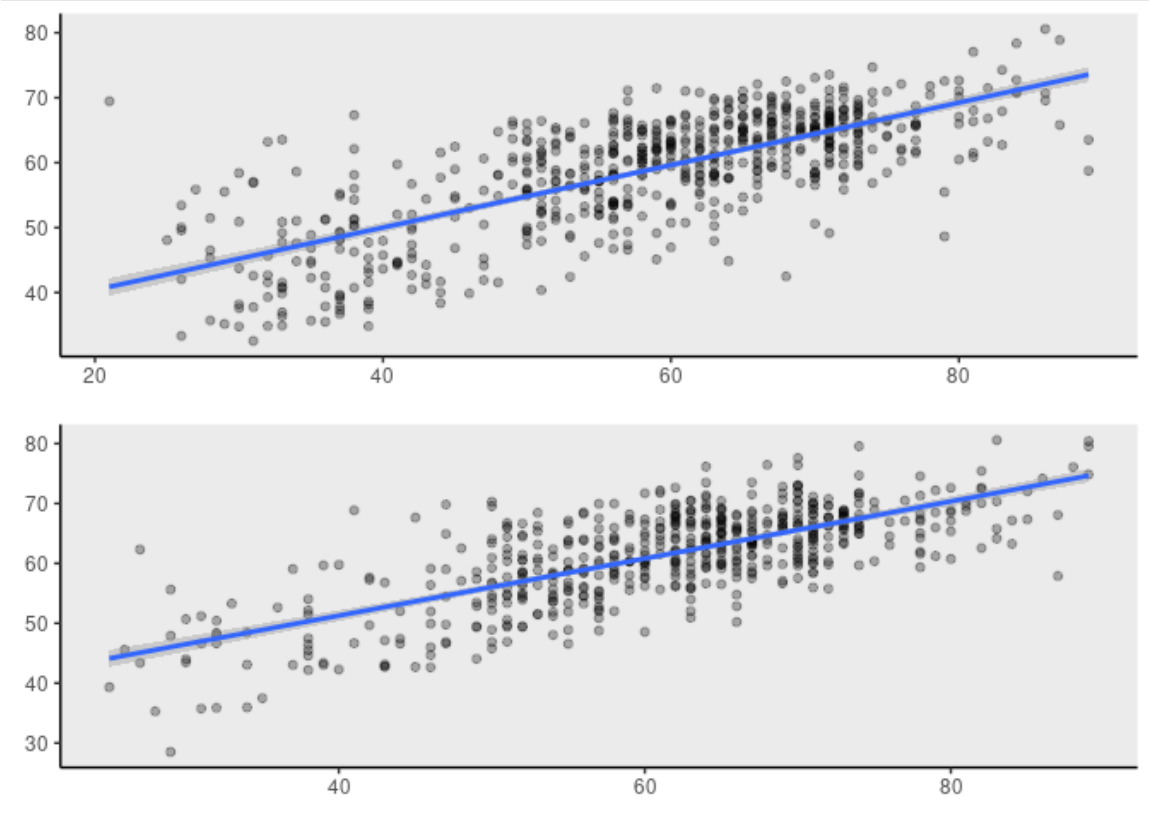

Figure: the cognitive and physical ages in the vertical axis versus the chronological ages in the horizontal axis, superimposed with the best fitting straight line, on female (upper) and male (lower) separately.

Habit Ages

```r
subset_Bfemale <- select(train_female, A4, starts_with("B")) %>%
  preprocessing()
test_Bfemale <- select(test_female, A4, starts_with("B")) %>%
  preprocessing()
X <- apply(subset_Bfemale[, -1], 2, function(x) (x-mean(x))/sd(x))
Y <- subset_Bfemale$A4


set.seed(113)
foci_Bfemale <- foci(Y, X, numCores = 1, stop = FALSE)


subset_Bmale <- select(train_male, A4, starts_with("B")) %>%
  preprocessing()
test_Bmale <- select(test_male, A4, starts_with("B")) %>%
  preprocessing()
X <- apply(subset_Bmale[, -1], 2, function(x) (x-mean(x))/sd(x))
Y <- subset_Bmale$A4

set.seed(115)
foci_Bmale <- foci(Y, X, numCores = 1, stop = FALSE)


plot_foci(foci_Bfemale, foci_Bmale, ind = c(11, 12))
```

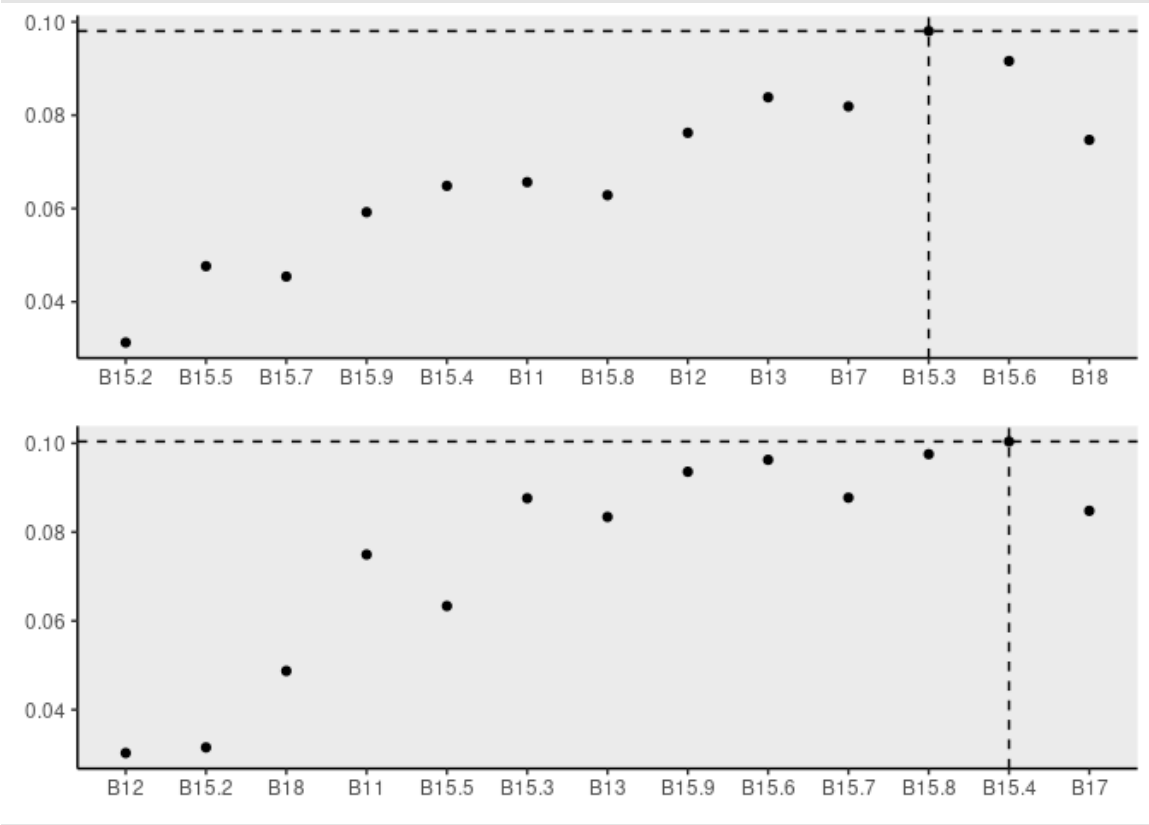

Figure: FOCI for habit-related features on female (upper) and male (lower) separately.

```
set.seed(114)
select_Bfemale <- foci_Bfemale$selectedVar$names[1:11]
RF_Bfemale <- randomForest(A4~., data=subset_Bfemale[, c("A4", select_Bfemale)],
                      tree=400)


set.seed(116)
select_Bmale <- foci_Bmale$selectedVar$names[1:12]
RF_Bmale <- randomForest(A4~., data=subset_Bmale[, c("A4", select_Bmale)],
                      tree=400)


plot_age(test_Bfemale$A4, predict(RF_Bfemale, test_Bfemale),
         test_Bmale$A4, predict(RF_Bmale, test_Bmale))
```

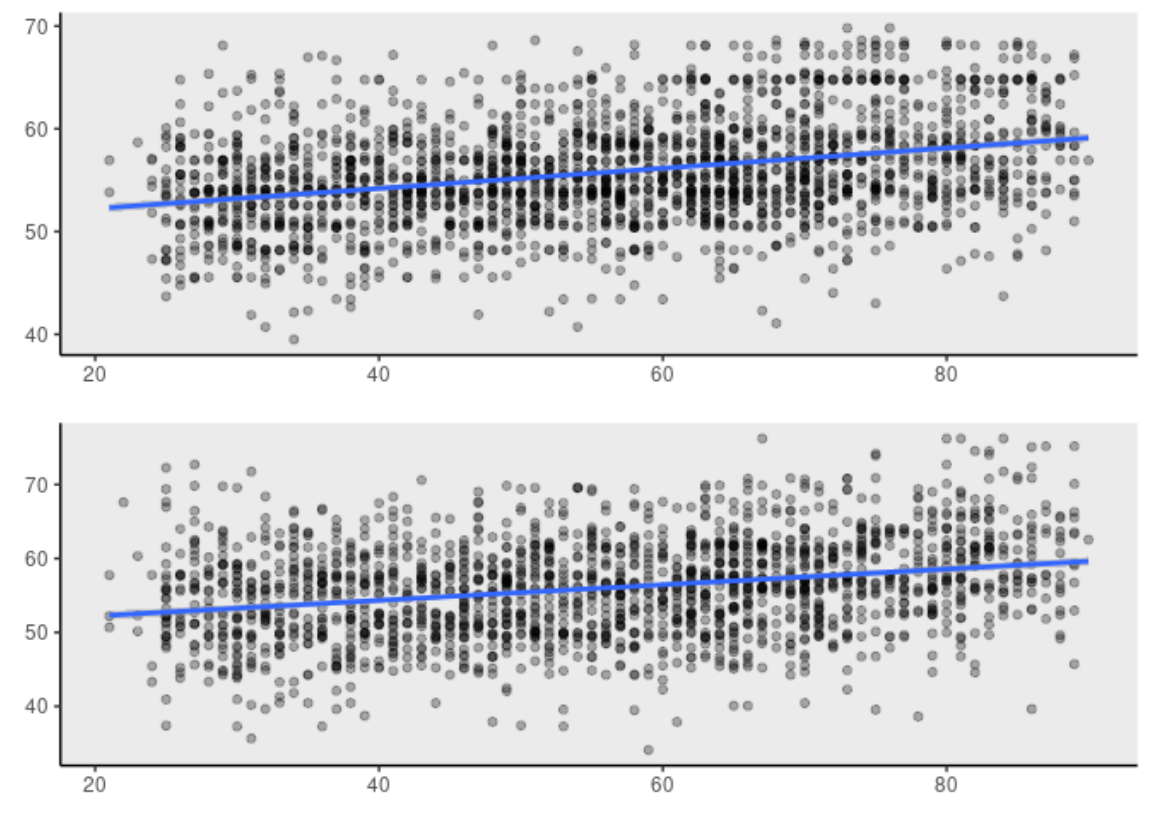

Figure: the habit ages in the vertical axis versus the chronological ages in the horizontal axis, superimposed with the best fitting straight line, on female (upper) and male (lower) separately.

- Among the four BAs, the cognitive and physical age has the least variance among subjects while the variation in habit ages is quite large, indicating potential differences of aging effects (e.g., large aging effects on cognitive and physical abilities, while small effects on living habits).

Applications of Biological Ages

Aging Index (AI)

- Definition:

$$\mathrm{AI} := \frac{\mathrm{rank(BA)} - \mathrm{rank(CA)}}{n},$$

which belongs to $[-1, 1]$. This aging index represents the relative aging acceleration or deceleration.

```
subset_allfemale <- train_female[, c("A4", select_Dfemale, select_Cfemale,
                                     select_Kfemale, select_Bfemale)] %>%
  preprocessing()

test_allfemale <- test_female[, c("A4", select_Dfemale, select_Cfemale,
                                  select_Kfemale, select_Bfemale)] %>%
  preprocessing()
Age_Dfemale <- predict(RF_Dfemale, test_allfemale)
Age_Cfemale <- predict(RF_Cfemale, test_allfemale)
Age_Kfemale <- predict(RF_Kfemale, test_allfemale)
Age_Bfemale <- predict(RF_Bfemale, test_allfemale)

Ind_Dfemale <- (rank(Age_Dfemale)-rank(test_allfemale$A4))/nrow(test_allfemale)
Ind_Cfemale <- (rank(Age_Cfemale)-rank(test_allfemale$A4))/nrow(test_allfemale)
Ind_Kfemale <- (rank(Age_Kfemale)-rank(test_allfemale$A4))/nrow(test_allfemale)
Ind_Bfemale <- (rank(Age_Bfemale)-rank(test_allfemale$A4))/nrow(test_allfemale)

Indfemale <- cbind(Ind_Dfemale, Ind_Cfemale,
                   Ind_Bfemale, Ind_Kfemale)
colnames(Indfemale) <- c("D", "C", "B", "K")


subset_allmale <- train_male[, c("A4", select_Dmale, select_Cmale,
                                 select_Kmale, select_Bmale)] %>%
  preprocessing()

test_allmale <- test_male[, c("A4", select_Dmale, select_Cmale,
                              select_Kmale, select_Bmale)] %>%
  preprocessing()
Age_Dmale <- predict(RF_Dmale, test_allmale)
Age_Cmale <- predict(RF_Cmale, test_allmale)
Age_Kmale <- predict(RF_Kmale, test_allmale)
Age_Bmale <- predict(RF_Bmale, test_allmale)

Ind_Dmale <- (rank(Age_Dmale)-rank(test_allmale$A4))/nrow(test_allmale)
Ind_Cmale <- (rank(Age_Cmale)-rank(test_allmale$A4))/nrow(test_allmale)
Ind_Kmale <- (rank(Age_Kmale)-rank(test_allmale$A4))/nrow(test_allmale)
Ind_Bmale <- (rank(Age_Bmale)-rank(test_allmale$A4))/nrow(test_allmale)

Indmale <- cbind(Ind_Dmale, Ind_Cmale, Ind_Bmale, Ind_Kmale)
colnames(Indmale) <- c("D", "C", "B", "K")

rbind(cor(Indfemale[, 1], Indfemale[, -1]),
      cor(Indmale[, 1], Indmale[, -1]))
```

```
##              C         B         K
## [1,] 0.3556388 0.3417384 0.3224937
## [2,] 0.3840946 0.3971869 0.3123600
```

- Pearson correlations between the disease aging index (“D”) and the other three aging indexes (“C” for organ, “B” for habit and “K” for cognitive and physical ability) were then assessed. It turns out that these three aging indexes had generally non-ignorable positive correlations with the disease one, ranging from 0.312 to 0.397. This indicates the disease aging index, which is our main interest, is to some extent predictable from the other three aging indexes.

A Cluster Analysis

- For each subject, the three aging indexes were used for clustering analysis, superimposed with the disease aging index, with the aim of identifying different aging patterns within the study population.

```
set.seed(117)
grid.newpage()
pushViewport(viewport(x = 0, width = 0.5, just = "left"))
grid.rect(gp = gpar(fill = "#FF000020"))
ht_listfemale = Heatmap(Indfemale[, -1], name = "Aging Index",
                        width = unit(20, "mm"), k = 8, show_row_names = FALSE) +
  Heatmap(Indfemale[, 1], name = "D",
          width = unit(5, "mm"), show_row_names = FALSE, show_heatmap_legend = FALSE)
draw(ht_listfemale, newpage = FALSE)
popViewport()

pushViewport(viewport(x = 0.5, width = 0.5, just = c("left")))
grid.rect(gp = gpar(fill = "#0000FF20"))
ht_listmale = Heatmap(Indmale[, -1], name = "Aging Index",
                      width = unit(20, "mm"), k = 8, show_row_names = FALSE) +
  Heatmap(Indmale[, 1], name = "D",
          width = unit(5, "mm"), show_row_names = FALSE, show_heatmap_legend = FALSE)
draw(ht_listmale, newpage = FALSE)
popViewport()
```

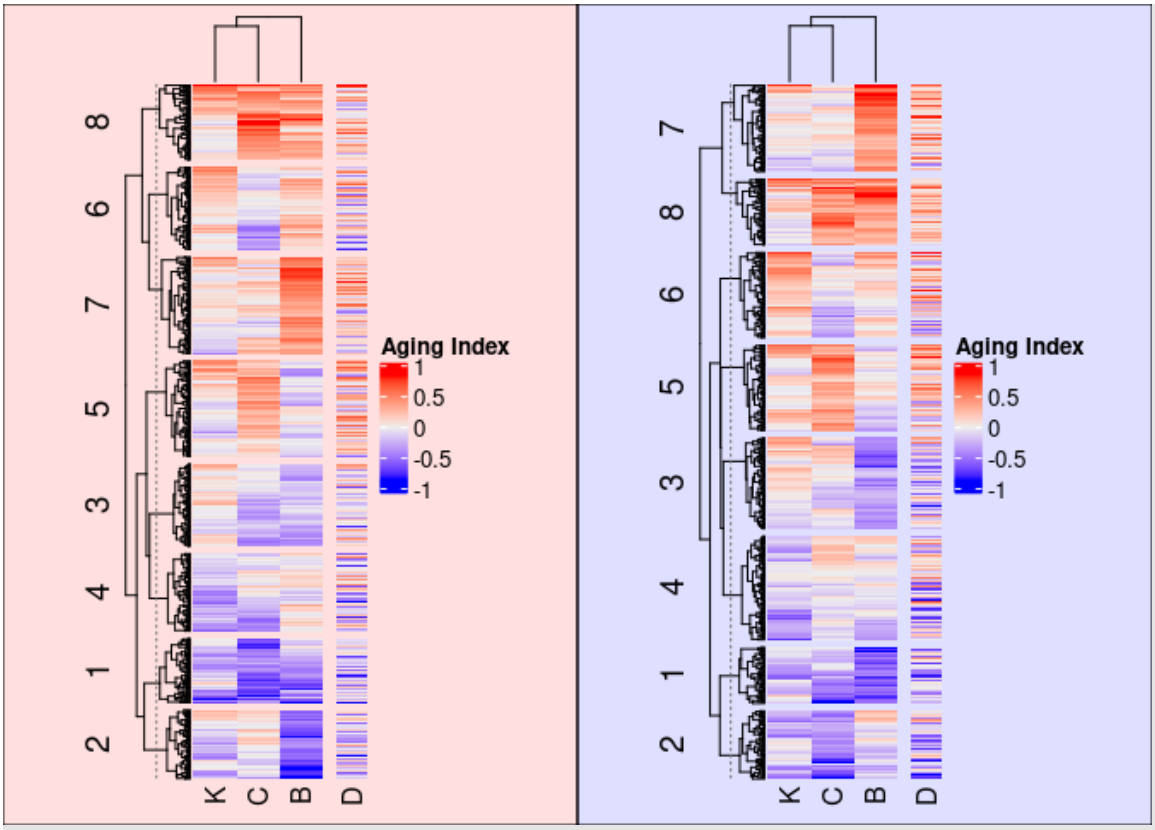

Figure: clustering heatmaps according to different biological ages. The color gradient represents the value of aging indexes. Red indicates faster and blue indicates slower aging rates.

- Individuals could be initially separated into two clusters, that is, one cluster with relatively more redder colors and the other with more bluer ones. It's also observed that if the three aging indexes are all redder (or bluer), the corresponding disease index will also be redder (or bluer).
- Within each cluster, subclusters were formed by the different patterns of aging effects. For instance, cluster 7 of female is redder in habit index, while cluster 5 is redder in organ index.

Conclusions and Discussions

- On one hand, determining BAs of different aspects itself may permit detailed evaluations of the source of accelerated aging components for each individual, which could potentially identify intervention targets for improving health status as well as slowing down the aging process. On the other hand, because of their inter-correlations, BAs that are easily accessible may sound an alarm for possible aging acceleration of other BAs that are harder to obtain.
- Although we focus on questionnaire data in this study, it's also possible to construct BAs utilizing multi-omics data, so that one is able to
  - study the aging effects of different organs and systems, and
  - relate them to specific genetic architectures.
- Another important application of BAs is forecasting an individual's longevity, or, mortality.

References

Azadkia, Mona, and Sourav Chatterjee. 2021. “A Simple Measure of Conditional Dependence.” *The Annals of Statistics* 49 (6): 3070–3102.

Chatterjee, Sourav. 2021. “A New Coefficient of Correlation.” *Journal of the American Statistical Association* 116 (536): 2009–22.

Pearson, Karl. 1895. “VII. Note on Regression and Inheritance in the Case of Two Parents.” *Proceedings of the Royal Society of London* 58 (347-352): 240–42.

Spearman, C. 1904. “The Proof and Measurement of Association Between Two Things.” *The American Journal of Psychology* 15 (1): 72–101.