

# HPF Coupling with Grey Matter Volume in Cognitive Decline

## Set up

### Load data

```
all_data_complete <- read.csv("datafile/all_data_complete_matched.csv",
                              stringsAsFactors = F)

dim(all_data_complete)

## [1] 369  32

all_data_complete$Research.Group <- factor(all_data_complete$Research.Group,
                                           levels = c("CN","MCI","AD"))

# correlation between left and right HPF
cor.test(all_data_complete$LHPF, all_data_complete$RHPF)

##
## Pearson's product-moment correlation
##
## data: all_data_complete$LHPF and all_data_complete$RHPF
## t = 33.703, df = 367, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8420093 0.8922667
## sample estimates:
##      cor
## 0.8693683

# sample in different groups
table(all_data_complete$Research.Group)

##
##  CN MCI  AD
## 143 101 125

table(all_data_complete$Research.Group, all_data_complete$Gender)

##
##      1  2
##  CN  81 62
##  MCI 47 54
##  AD  54 71
```

```

# code scanning parameters as a dummy variable
# as.factor(all_data_complete$Imaging.Protocol)

parameters <- strsplit(all_data_complete$Imaging.Protocol, ";")

info <- as.data.frame(matrix(0, 369, 6))
for (i in 1:369) {
  k <- grep("Manufacturer", parameters[[i]])
  m <- grep("Mfg Model", parameters[[i]])
  n <- grep("Field Strength", parameters[[i]])
  x <- grep("Slice Thickness", parameters[[i]])
  y <- grep("Matrix Z", parameters[[i]])

  info[i, 1:5] <- parameters[[i]][c(k, m, n, x, y)]
  info[i, 6] <- paste0(parameters[[i]][k], ";", parameters[[i]][m], ";",
    parameters[[i]][n], ";", parameters[[i]][x], ";",
    parameters[[i]][y])
}

for (i in 1:369) {

  j <- which(info$V6[i] == levels(as.factor(info$V6)))

  all_data_complete$scan_code[i] <- j
}

```

## Demographic and clinical information

```

demo_info <- as.data.frame(matrix(0,6,7))
colnames(demo_info) <- c("CN_mean", "CN_sd", "MCI_mean", "MCI_sd", "AD_mean",
  "AD_range", "p value")

rownames(demo_info) <- c("Gender", "Age", "Education", "CDR", "MMSE", "MoCA")

# gender: chi-squared test
gender_gr <- table(all_data_complete$Research.Group, all_data_complete$Gender)

chisq.test(gender_gr)$p.value

```

```
## [1] 0.07226473
```

```

demo_info["Gender", 1] <- paste0(gender_gr["CN",2] ,"/",gender_gr["CN",1])
demo_info["Gender", 3] <- paste0(gender_gr["MCI",2] ,"/",gender_gr["MCI",1])
demo_info["Gender", 5] <- paste0(gender_gr["AD",2] ,"/",gender_gr["AD",1])
demo_info["Gender", 7] <- round(chisq.test(gender_gr)$p.value,3)

# Age, Education MMSE, MoCA: mean, sd, range
mean_all <- describeBy(all_data_complete[, c("Research.Group", "Age", "Education",
  "Global.CDR",

```

```

      "MMSE.Total.Score", "MoCA.Total")],
    group = "Research.Group",
    digits = 3)

demo_info[c("Age", "Education", "CDR", "MMSE", "MoCA"), 1:6] <-
  cbind(paste0(round(mean_all[[1]]$mean[2:6], 2), "±",
    round(mean_all[[1]]$sd[2:6], 2)),
    paste0(round(mean_all[[1]]$min[2:6], 2), "-",
    round(mean_all[[1]]$max[2:6], 2)),
    paste0(round(mean_all[[2]]$mean[2:6], 2), "±",
    round(mean_all[[2]]$sd[2:6], 2)),
    paste0(round(mean_all[[2]]$min[2:6], 2), "-",
    round(mean_all[[2]]$max[2:6], 2)),
    paste0(round(mean_all[[3]]$mean[2:6], 2), "±",
    round(mean_all[[3]]$sd[2:6], 2)),
    paste0(round(mean_all[[3]]$min[2:6], 2), "-",
    round(mean_all[[3]]$max[2:6], 2)))

## one way ANOVA
ano_age <- anova(lm(Age ~ Research.Group, all_data_complete))
ano_education <- anova(lm(Education ~ Research.Group, all_data_complete))
ano_MMSE <- anova(lm(MMSE.Total.Score ~ Research.Group, all_data_complete))
ano_MoCA <- anova(lm(MoCA.Total ~ Research.Group, all_data_complete))

demo_info["Age", 7] <- round(ano_age$`Pr(>F)`[1], 3)
demo_info["Education", 7] <- round(ano_education$`Pr(>F)`[1], 3)
demo_info["MMSE", 7] <- round(ano_MMSE$`Pr(>F)`[1], 4)
demo_info["MoCA", 7] <- round(ano_MoCA$`Pr(>F)`[1], 4)

# Age
t.test(all_data_complete$Age[all_data_complete$Research.Group == "AD"],
  all_data_complete$Age[all_data_complete$Research.Group == "CN"])

```

```

##
## Welch Two Sample t-test
##
## data: all_data_complete$Age[all_data_complete$Research.Group == "AD"] and all_data_complete$Age[all_data_complete$Research.Group == "CN"]
## t = 0.30813, df = 264.06, p-value = 0.7582
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.599405 2.192860
## sample estimates:
## mean of x mean of y
## 75.22400 74.92727

```

```

t.test(all_data_complete$Age[all_data_complete$Research.Group == "CN"],
  all_data_complete$Age[all_data_complete$Research.Group == "MCI"])

```

```

##
## Welch Two Sample t-test
##
## data: all_data_complete$Age[all_data_complete$Research.Group == "CN"] and all_data_complete$Age[all_data_complete$Research.Group == "MCI"]

```

```
## t = 1.6734, df = 198.07, p-value = 0.09582
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.3382409 4.1294200
## sample estimates:
## mean of x mean of y
## 74.92727 73.03168
```

```
t.test(all_data_complete$Age[all_data_complete$Research.Group == "AD"],
       all_data_complete$Age[all_data_complete$Research.Group == "MCI"])
```

```
##
## Welch Two Sample t-test
##
## data: all_data_complete$Age[all_data_complete$Research.Group == "AD"] and all_data_complete$Age[all_data_complete$Research.Group == "MCI"]
## t = 1.9229, df = 195.33, p-value = 0.05594
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.05613204 4.44076571
## sample estimates:
## mean of x mean of y
## 75.22400 73.03168
```

```
## CDR
kruskal.test(Global.CDR ~ Research.Group, data = all_data_complete)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: Global.CDR by Research.Group
## Kruskal-Wallis chi-squared = 306.68, df = 2, p-value < 2.2e-16
```

```
demo_info["CDR", 7] <- "< 0.001"
```

```
knitr::kable(demo_info)
```

	CN_mean	CN_sd	MCI_mean	MCI_sd	AD_mean	AD_range	p value
Gender	62/81	0	54/47	0	71/54	0	0.072
Age	74.93±8.07	58.4-94.7	73.03±9.14	55.8-97.4	75.22±7.68	56-88	0.105
Education	16.89±2.32	12-20	16.21±2.59	8-20	15.47±2.51	8-20	0
CDR	0.01±0.06	0-0.5	0.48±0.15	0-1	0.8±0.33	0.5-2	< 0.001
MMSE	29.01±1.1	25-30	27.66±2.09	19-30	22.94±3.21	5-30	0
MoCA	24.31±1.79	18-28	22.86±3.32	10-29	18.03±5.26	0-27	0

```
#writexl::write_xlsx(demo_info, "results/demo_info.xlsx")
```

```
# subjects with education
```

```
length(which(!is.na(all_data_complete$Education[all_data_complete$Research.Group == "CN"])))
```

```
## [1] 139
```

```
length(which(!is.na(all_data_complete$Education[all_data_complete$Research.Group == "MCI"])))

## [1] 100

length(which(!is.na(all_data_complete$Education[all_data_complete$Research.Group == "AD"])))

## [1] 123

# subjects with MoCA
length(which(!is.na(all_data_complete$MoCA.Total[all_data_complete$Research.Group == "AD"])))

## [1] 74

length(which(!is.na(all_data_complete$MoCA.Total[all_data_complete$Research.Group == "CN"])))

## [1] 133

length(which(!is.na(all_data_complete$MoCA.Total[all_data_complete$Research.Group == "MCI"])))

## [1] 90
```

## Group differences in the HPF across diagnostic groups and barplots

```
### regression analysis
# convert variables

all_data_complete$TIV_z <- (all_data_complete$TIV-mean(all_data_complete$TIV))/
  sd(all_data_complete$TIV)

median(all_data_complete$Age)

## [1] 75.2

all_data_complete$Age_centered <- all_data_complete$Age-median(all_data_complete$Age)

all_data_complete$AgeSquared <- all_data_complete$Age*all_data_complete$Age

all_data_complete$AgeSquared_centered <- all_data_complete$AgeSquared -
  median(all_data_complete$Age)*median(all_data_complete$Age)

all_data_complete$Research.Group <- as.character(all_data_complete$Research.Group)

all_data_complete$Subject.ID <- as.factor(all_data_complete$Subject.ID)

# statistical analysis
fit_HPF <- lm(HPF ~ Research.Group + Gender + Age_centered + AgeSquared_centered +
```

```

scan_code + TIV_z, data = all_data_complete)

fit_AI <- lm(HPF_AI ~ Research.Group + Gender + Age_centered + AgeSquared_centered +
scan_code + TIV_z, data = all_data_complete)

anova(fit_HPF)

```

```

## Analysis of Variance Table
##
## Response: HPF
##
##      Df Sum Sq Mean Sq F value    Pr(>F)
## Research.Group      2  1.21374  0.60687  114.3786 < 2.2e-16 ***
## Gender              1  0.17222  0.17222   32.4591 2.529e-08 ***
## Age_centered        1  1.11950  1.11950  210.9956 < 2.2e-16 ***
## AgeSquared_centered  1  0.00335  0.00335    0.6318  0.4272
## scan_code           1  0.00162  0.00162    0.3047  0.5813
## TIV_z               1  0.09283  0.09283   17.4958 3.619e-05 ***
## Residuals          361  1.91539  0.00531
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
anova(fit_AI)
```

```

## Analysis of Variance Table
##
## Response: HPF_AI
##
##      Df Sum Sq Mean Sq F value    Pr(>F)
## Research.Group      2 1563.8   781.88  18.8438 1.644e-08 ***
## Gender              1  134.6   134.62   3.2444 0.072502 .
## Age_centered        1   362.9   362.91   8.7464 0.003306 **
## AgeSquared_centered  1    9.6     9.60   0.2313 0.630869
## scan_code           1    2.8     2.81   0.0676 0.794946
## TIV_z               1   188.1   188.14   4.5344 0.033897 *
## Residuals          361 14978.9   41.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

# group differences
diff_hpf <- as.data.frame(matrix(0, 6,5))

colnames(diff_hpf) <- c("", "", "t value", "p value", "cohen's d")

k <- 1
for (hpf in c("HPF_AI", "HPF")) {

  t1 <- t.test(all_data_complete[all_data_complete$Research.Group == "CN", hpf],
    all_data_complete[all_data_complete$Research.Group == "MCI", hpf])
  t2 <- t.test(all_data_complete[all_data_complete$Research.Group == "AD", hpf],
    all_data_complete[all_data_complete$Research.Group == "MCI", hpf])
  t3 <- t.test(all_data_complete[all_data_complete$Research.Group == "CN", hpf],
    all_data_complete[all_data_complete$Research.Group == "AD", hpf])

```

```

e1 <- effsize::cohen.d(all_data_complete[all_data_complete$Research.Group == "CN",hpf],
  all_data_complete[all_data_complete$Research.Group == "MCI",hpf])
e2 <- effsize::cohen.d(all_data_complete[all_data_complete$Research.Group == "AD",hpf],
  all_data_complete[all_data_complete$Research.Group == "MCI",hpf])
e3 <- effsize::cohen.d(all_data_complete[all_data_complete$Research.Group == "CN",hpf],
  all_data_complete[all_data_complete$Research.Group == "AD",hpf])

diff_hpf[k:(2+k),1] <- hpf
diff_hpf[k:(2+k),2] <- c("CNvsMCI","ADvsMCI","CNvsAD")
diff_hpf[k:(2+k),3] <- rbind(round(t1$statistic,2),round(t2$statistic,2),
  round(t3$statistic,2))
diff_hpf[k:(2+k),4] <- rbind(round(t1$p.value,3),round(t2$p.value,3),
  round(t3$p.value,3))
diff_hpf[k:(2+k),5] <- rbind(round(e1$estimate,2),round(e2$estimate,2),
  round(e3$estimate,2))
k <- k+3
}

knitr::kable(diff_hpf)

```

		t value	p value	cohen's d
HPF_AI	CNvsMCI	-1.32	0.188	-0.17
HPF_AI	ADvsMCI	3.95	0.000	0.51
HPF_AI	CNvsAD	-5.55	0.000	-0.70
HPF	CNvsMCI	2.43	0.016	0.32
HPF	ADvsMCI	-7.64	0.000	-1.01
HPF	CNvsAD	11.01	0.000	1.37

```
#View(diff_hpf)
```

```
# estimated marginal means
```

```
emmHPF <- emmeans(fit_HPF, "Research.Group")
class(emmHPF)
```

```
## [1] "emmGrid"
## attr(,"package")
## [1] "emmeans"
```

```

emmHPF <- as.data.frame(emmHPF)
emmHPF_new <- emmHPF[c(2:3,1),]

row.names(emmHPF_new) <- NULL
emmHPF_new$Research.Group <- factor(emmHPF_new$Research.Group,levels = c("CN", "MCI", "AD"))

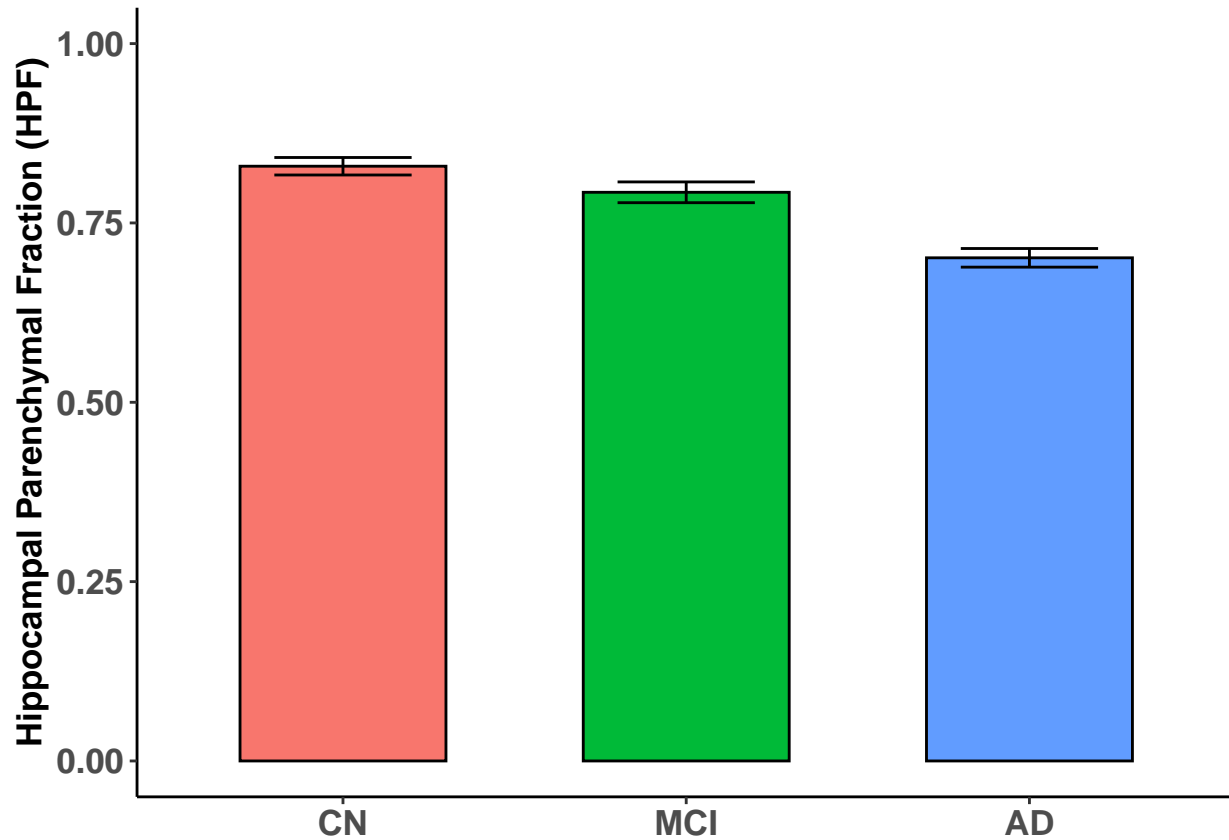
ggplot(emmHPF_new, aes(x = Research.Group, y = emmean, fill = Research.Group, width=.6)) +
  geom_bar(stat="identity", color="black", position=position_dodge()) +
  geom_errorbar(aes(ymin=lower.CL, ymax=upper.CL), width=.4,
    position=position_dodge(.9)) +
  labs(y = "Hippocampal Parenchymal Fraction (HPF)", x = "Group") +
  guides(color = F, fill = F) +
  theme(axis.text = element_text(size = 13, face = "bold"),

```

```

axis.title.x = element_blank(),
axis.title.y = element_text(size = 13, face = "bold"))+
theme(panel.border = element_blank(),
panel.background = element_blank(),
panel.grid = element_blank(),
axis.line = element_line(colour = "black")) +
ylim(0,1)

```



```

ggsave("results/HPF_barplot.png", width = 5, height = 4, units = c("in"), dpi = 200)

```

```

## HPF AI
fit_HPF_AI <- lm(HPF_AI ~ Research.Group + Sex + Age_centered +
  AgeSquared_centered + TIV_z, data = all_data_complete)

anova(fit_HPF_AI)

```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: HPF_AI
```

```
##
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## Research.Group	2	1563.8	781.88	18.8193	1.685e-08 ***
## Sex	3	166.4	55.48	1.3353	0.262633
## Age_centered	1	364.1	364.07	8.7628	0.003278 **
## AgeSquared_centered	1	10.3	10.29	0.2476	0.619101



```
## TIV_z          1   179.3   179.34   4.3166   0.038451 *
## Residuals      360 14956.8    41.55
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#plot_model(fit_HPF_AI, type = "pred", terms = c("Age_centered"))
```

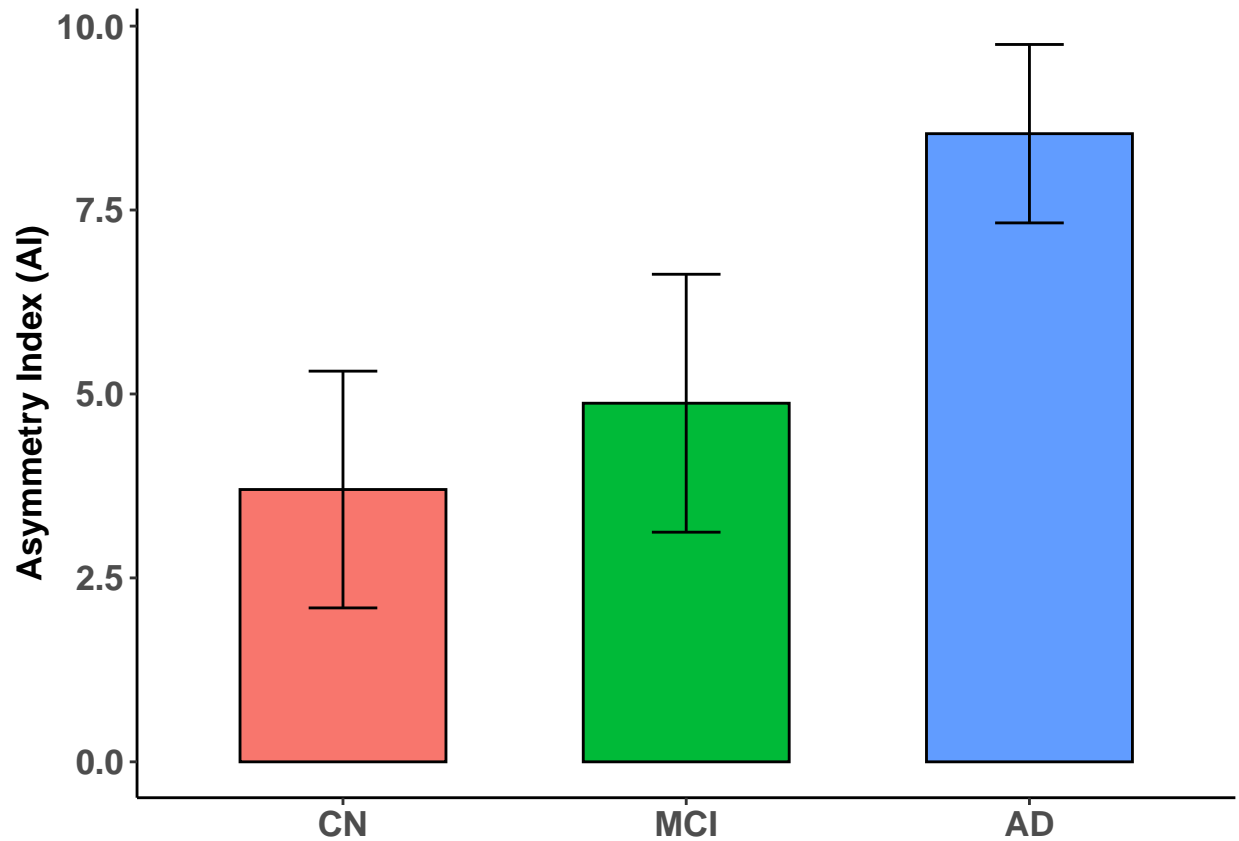
```
#plot_model(fit_HPF_AI, type = "pred", terms = c("TIV_z"))
```

```
# plot with predicted marginal means
emmHPF_AI <- emmeans(fit_HPF_AI, "Research.Group")
emmHPF_AI <- as.data.frame(emmHPF_AI)
emmHPF_AI$Research.Group
```

```
## [1] AD CN MCI
## Levels: AD CN MCI
```

```
emmHPF_AI$Research.Group <- factor(emmHPF_AI$Research.Group,
                                   levels = c("CN", "MCI", "AD"))

ggplot(emmHPF_AI, aes(x = Research.Group, y = emmean, fill = Research.Group, width=.6)) +
  geom_bar(stat="identity", color="black", position=position_dodge()) +
  geom_errorbar(aes(ymin=lower.CL, ymax=upper.CL), width=.2,
               position=position_dodge(.9)) +
  labs(y = "Asymmetry Index (AI)", x = "Group") +
  guides(color = F, fill = F) +
  theme(axis.text = element_text(size = 13, face = "bold"),
        axis.title.x = element_blank(),
        axis.title.y = element_text(size = 13, face = "bold"))+
  theme(panel.border = element_blank(),
        panel.background = element_blank(),
        panel.grid = element_blank(),
        axis.line = element_line(colour = "black"))
```



```
ggsave("results/HPF_AI_barplot.png", width = 5, height = 4, units = c("in"), dpi = 200)
```

```
# variance in HPF, RHPF - LHPF
```

```
AD_subj <- all_data_complete$Subject.ID[all_data_complete$Research.Group == "AD"]
```

```
CN_subj <- all_data_complete$Subject.ID[all_data_complete$Research.Group == "CN"]
```

```
MCI_subj <- all_data_complete$Subject.ID[all_data_complete$Research.Group == "MCI"]
```

```
all_HPF <- read.csv("datafile/all_HPF_mean_AI_LR.csv", stringsAsFactors = F)
```

```
all_HPF <- as.data.frame(all_HPF)
```

```
# all_HPF$HPF_RL
```

```
# AI using raw values (instead of absolute values)
```

```
all_HPF$rawAI <- (all_HPF$HPF_RL/all_HPF$HPF)*100
```

```
mean(all_HPF$rawAI[all_HPF$subjid %in% AD_subj])
```

```
## [1] 3.195638
```

```
mean(all_HPF$rawAI[all_HPF$subjid %in% CN_subj])
```

```
## [1] 1.147131
```

```
mean(all_HPF$rawAI[all_HPF$subjid %in% MCI_subj])
```

```
## [1] 1.728207
```

```
# mean and sd
```

```
length(all_HPF$HPF_RL[all_HPF$subjid %in% AD_subj])
```

```
## [1] 125
```

```
sd(all_HPF$HPF_RL[all_HPF$subjid %in% AD_subj])
```

```
## [1] 0.07243424
```

```
sd(all_HPF$HPF_RL[all_HPF$subjid %in% CN_subj])
```

```
## [1] 0.04817799
```

```
sd(all_HPF$HPF_RL[all_HPF$subjid %in% MCI_subj])
```

```
## [1] 0.05276699
```

```
sd(all_HPF$HPF[all_HPF$subjid %in% CN_subj])
```

```
## [1] 0.08447472
```

```
sd(all_HPF$HPF[all_HPF$subjid %in% AD_subj])
```

```
## [1] 0.1064342
```

```
sd(all_HPF$HPF[all_HPF$subjid %in% MCI_subj])
```

```
## [1] 0.09417543
```

```
sd(all_HPF$HPF_AI[all_HPF$subjid %in% CN_subj])
```

```
## [1] 5.440186
```

```
sd(all_HPF$HPF_AI[all_HPF$subjid %in% AD_subj])
```

```
## [1] 7.927813
```

```
sd(all_HPF$HPF_AI[all_HPF$subjid %in% MCI_subj])
```

```
## [1] 6.06708
```

```
t.test(all_HPF$HPF_RL[all_HPF$subjid %in% CN_subj],all_HPF$HPF_RL[all_HPF$subjid %in% MCI_subj])

##
## Welch Two Sample t-test
##
## data: all_HPF$HPF_RL[all_HPF$subjid %in% CN_subj] and all_HPF$HPF_RL[all_HPF$subjid %in% MCI_subj]
## t = -0.5679, df = 202.89, p-value = 0.5707
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.01680753 0.00929069
## sample estimates:
## mean of x mean of y
## 0.008683364 0.012441782

t.test(all_HPF$HPF_RL[all_HPF$subjid %in% AD_subj],all_HPF$HPF_RL[all_HPF$subjid %in% CN_subj])

##
## Welch Two Sample t-test
##
## data: all_HPF$HPF_RL[all_HPF$subjid %in% AD_subj] and all_HPF$HPF_RL[all_HPF$subjid %in% CN_subj]
## t = 1.8589, df = 210.91, p-value = 0.06444
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.0008575306 0.0292211553
## sample estimates:
## mean of x mean of y
## 0.022865176 0.008683364
```

## Associations between HPF and MoCA

```
# correlations between age and HPF
cor.test(all_data_complete$MoCA.Total[all_data_complete$Research.Group == "AD"],
         all_data_complete$Age[all_data_complete$Research.Group == "AD"])

##
## Pearson's product-moment correlation
##
## data: all_data_complete$MoCA.Total[all_data_complete$Research.Group == "AD"] and all_data_complete$Age[all_data_complete$Research.Group == "AD"]
## t = 0.88105, df = 72, p-value = 0.3812
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.1282479 0.3241271
## sample estimates:
## cor
## 0.1032774

# regression analysis
all_data_complete_tmp <- all_data_complete[!is.na(all_data_complete$MoCA.Total),]
which(all_data_complete_tmp$MoCA.Total == 0)
```

```
## [1] 242
```

```
all_data_complete_tmp2 <- all_data_complete_tmp[all_data_complete_tmp$MoCA.Total != 0, ]  
all_data_complete_tmp2$HPF_rescale <- all_data_complete_tmp2$HPF*100  
dim(all_data_complete_tmp2)
```

```
## [1] 296 38
```

```
# View(all_data_complete_tmp2)  
fit <- lm(MoCA.Total ~ HPF_rescale + Age_centered + Gender + Education +  
          TIV_z, data = all_data_complete_tmp2)  
  
#anova(fit)  
summary(fit)
```

```
##  
## Call:  
## lm(formula = MoCA.Total ~ HPF_rescale + Age_centered + Gender +  
##      Education + TIV_z, data = all_data_complete_tmp2)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -11.9126  -1.5659   0.4358   2.0405   7.8028   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  -2.70606    2.38046  -1.137  0.25659      
## HPF_rescale   0.23281    0.02326  10.009 < 2e-16 ***  
## Age_centered  0.16289    0.02753   5.917 9.38e-09 ***  
## Gender       -0.20318    0.50638  -0.401  0.68854      
## Education     0.43852    0.08351   5.251 2.97e-07 ***  
## TIV_z         0.79264    0.24661   3.214 0.00146 **    
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 3.307 on 285 degrees of freedom  
## (5 observations deleted due to missingness)  
## Multiple R-squared:  0.3517, Adjusted R-squared:  0.3403   
## F-statistic: 30.93 on 5 and 285 DF, p-value: < 2.2e-16
```

```
# confident interval for estimated coefficient  
confint(fit, 'HPF_rescale', level=0.95)
```

```
##              2.5 %    97.5 %  
## HPF_rescale 0.1870263 0.2785942
```

```
#plot_model(fit, type = "pred", terms = c("HPF_rescale"))
```

```
# post-hoc regression analysis
```

```

gr <- "MCI"

all_data_complete_tmp3 <-
  all_data_complete_tmp2[all_data_complete_tmp2$Research.Group == gr,]

fit <- lm(MoCA.Total ~ HPF_rescale + Age_centered + Gender + Education + TIV_z,
  data = all_data_complete_tmp3)

summary(fit)

##
## Call:
## lm(formula = MoCA.Total ~ HPF_rescale + Age_centered + Gender +
##     Education + TIV_z, data = all_data_complete_tmp3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6719  -1.3787   0.2546   1.5552   6.5908
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.34095     4.35774   0.308 0.759069
## HPF_rescale    0.16284     0.04492   3.625 0.000496 ***
## Age_centered   0.10543     0.04359   2.418 0.017772 *
## Gender         0.52027     0.84207   0.618 0.538366
## Education      0.49136     0.12355   3.977 0.000149 ***
## TIV_z          0.34400     0.41610   0.827 0.410760
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.907 on 83 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.2789, Adjusted R-squared:  0.2355
## F-statistic: 6.421 on 5 and 83 DF,  p-value: 4.325e-05

for (gr in c("AD", "MCI", "CN")) {

  all_data_complete_tmp3 <-
    all_data_complete_tmp2[all_data_complete_tmp2$Research.Group == gr,]

  fit <- lm(MoCA.Total ~ HPF_rescale + Age_centered + Gender + Education + TIV_z,
    data = all_data_complete_tmp3)

  summary(fit)

  plot_model(fit, type = "pred", terms = c("HPF_rescale"))

  ## plot Linear Model Fitted to Data

  # original plot
  ggplot(all_data_complete_tmp3, aes(x=HPF, y=MoCA.Total)) +
    geom_point() +
    geom_smooth(method = "lm") +

```

```

    ggtitle("outcome ~ pred") +
    theme_bw()

# declare new model here
plm <- lm(formula = MoCA.Total ~ HPF + Sex + Age_centered +
          AgeSquared_centered + TIV_z, data=all_data_complete_tmp3)

# plot with lm for outcome ~ pred + factor
plot2 <- ggplot(all_data_complete_tmp3, aes(x=HPF, y=MoCA.Total)) +
  geom_point(size = 5, col = 'purple', alpha = 0.6) +
  geom_smooth(method = "lm", mapping=aes(y=predict(plm,all_data_complete_tmp3)),
             col = 'black', size = 2, se = FALSE) +
#   ggtitle("outcome ~ pred + factor")+
#   ylim(0,30) +
  labs(x = "Hippocampal Parenchymal Fraction", y = "MoCA Scores") +
  theme(legend.title = element_text(colour="black", size=14, face="bold"),
        legend.text = element_text(colour="black", size=14, face="bold")) +
  theme(plot.title = element_text(hjust = 0.5))+
  theme(axis.text = element_text(size = 16, face = "bold"),
        axis.title = element_text(size = 16, face = "bold")) +
  theme(panel.background = element_blank(),
        panel.border = element_blank(),
        panel.grid = element_blank(),
        axis.line = element_line(colour = "black")) +
  coord_cartesian(ylim=c(5,30), xlim = c(0.5, 1)) +
  scale_y_continuous(breaks = seq(5,30,5))

#ggsave(paste0("results/",gr,"_MoCA.png"), width = 5, height = 4,
#       units = c("in"), dpi = 200)
}

```

## Filter variables for brain-HPF correlation analysis

```

#colnames(all_data_complete)

#View(all_data_complete)

# save variables
#colnames(all_data_complete)
all_data_complete_variables <- all_data_complete[, c(1,7, 32, 11, 35, 21, 37,26)]

#View(all_data_complete_variables)

#write.csv(all_data_complete_variables, "datafile/all_data_complete_variables_new.csv",
#         row.names = F, quote = FALSE)

```