# GLM\_analysis - backup (original all)

install.packages("gtsummary",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

library(gtsummary)

data\_sg3 <- read.csv("Z:/zio/R/final\_data\_for\_analysis\_sg3.csv")

data\_sg3\_1 <- read.csv("Z:/zio/R/final\_data\_for\_analysis\_sg3\_1.csv")

data\_sg3\_bmi\_imputed <- read.csv('Z:/zio/R/final\_data\_for\_analysis\_sg3\_bmi.csv') # bmi mising values imputed with mean value of each study group

data\_sg3\_bmi\_imputed\_ageRef\_changed <- read.csv('Z:/zio/R/final\_data\_for\_analysis\_sg3\_bmi\_ageRefchanged.csv') # bmi mising values imputed with mean value of each study group and age reference changed to 20-29

# 전체 362만 아기로 재정리한 데이터

data\_whole <- read.csv("Z:/zio/R/prediction\_dta.csv") # 3623766 263

# any\_neruo

tbl\_uvregression(

data=data,

method=glm,

method.args=list(binomial()),

exponentiate=T,

hide\_n=F,

y='any\_neruo',

include=c("study\_group\_1", "study\_group\_2", "study\_group\_3", "study\_group\_3\_1",

"sex\_type", 'smoke', 'mother\_age', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght', 'g1e\_bmi', 'g1e\_bp\_sys', 'g1e\_bp\_dia',

'g1e\_urn\_prot', 'g1e\_tot\_chol'),

)

# any\_cm

tbl\_uvregression(

data=data,

method=glm,

method.args=list(binomial()),

exponentiate=T,

# add\_estimate\_to\_reference\_rows=T,

hide\_n=F,

y='any\_cm',

include=c("study\_group\_1", "study\_group\_2", "study\_group\_3", "study\_group\_3\_1", "sex\_type", 'smoke', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght', 'g1e\_bmi', 'g1e\_bp\_sys', 'g1e\_bp\_dia',

'g1e\_urn\_prot', 'g1e\_tot\_chol'),

)

# test

tbl\_uvregression(

data=trial\_nonna, method=glm,

y=death,

include=c("trt", "marker", "stage")

)

trial\_nonna <- na.omit(trial)

tbl\_uvregression(

data=trial\_nonna, method=glm,

y=death,

)

library(broom)

library(dplyr)

model <- glm(death~trt, family="binomial", data=trial) %>%

tidy(exponentiate=TRUE, conf.int=TRUE)

# %>% mutate(across(where(is.numeric), round, digits=2))

### UNIVARIATE ANALYSIS ###

install.packages("gtsummary",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("car",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("openxlsx",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("r2glmm",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

library(broom)

library(dplyr)

library(stringr)

library(tidyr)

library(purrr)

library(car) # vif

library(openxlsx) # for multiple sheets

library(r2glmm) # r2 of model

# copy data needed

data <- data\_sg3

data <- data\_sg3\_1 # study group == 3 of this data is actually study group 3-1

data <- data\_sg3\_bmi\_imputed # bmi mising values imputed with mean value of each study group

data <- data\_sg3\_bmi\_imputed\_ageRef\_changed

data <- data\_whole

# factor categorical variables

data$study\_group <- factor(data$study\_group)

data$SEX\_TYPE <- factor(data$SEX\_TYPE)

data$SEX\_TYPE <- relevel(data$SEX\_TYPE, ref=2) # fetal sex reference를 female로 변경

data$smoke <- factor(data$smoke)

data$HTN <- factor(data$HTN)

data$DM <- factor(data$DM)

data$DSPDA <- factor(data$DSPDA)

data$MI <- factor(data$MI)

data$CHF <- factor(data$CHF)

data$CDO <- factor(data$CDO)

data$TIA <- factor(data$TIA)

data$CLD <- factor(data$CLD)

data$COPD <- factor(data$COPD)

data$Malignancy <- factor(data$Malignancy)

data$GDM <- factor(data$GDM)

data$PE <- factor(data$PE)

data$CE <- factor(data$CE)

data$PA <- factor(data$PA)

data$PP <- factor(data$PP)

data$NP <- factor(data$NP)

data$LGA <- factor(data$LGA)

data$pb\_real <- factor(data$pb\_real)

data$low\_bth\_wght <- factor(data$low\_bth\_wght)

data$bmi\_category <- factor(data$bmi\_category)

data$sudogwon <- factor(data$sudogwon)

data$income\_level <- factor(data$income\_level)

data$insurance\_type <- factor(data$insurance\_type)

data$any\_cm <- factor(data$any\_cm)

data$any\_neruo <- factor(data$any\_neruo)

data$APO <- factor(data$APO)

## original age group ##

data$age\_group <- factor(data$age\_group)

data$age\_group <- relevel(data$age\_group, ref=1) # age\_group reference를 1로 지정 (이미 자동으로 되어있지만, 확인차..)

## 세분화된 age group ## data\_sg3\_bmi\_imputed\_ageRef\_changed 사용할 때만 실행 !!! ##

data$age\_group <- factor(data$age\_group)

data$age\_group <- relevel(data$age\_group, ref=2) # age\_group reference를 20-29세로 지정

## final age group ## 이걸로 실행 ##

data$age\_group <- factor(data$age\_group)

data$age\_group <- relevel(data$age\_group, ref=2) # age\_group reference를 2로 지정 (25-29세)

## study group 1, 2를 합칠 때만 실행 !!! ##

data[which(data$study\_group == 2), 'study\_group'] <- 1

table(data$study\_group)

colSums(is.na(data))

## study group 3를 삭제할 때만 실행 !!! ##

data <- data[which(data$study\_group != 3),]

table(data$study\_group)

# define variables of interest

explanatory\_vars <- c("study\_group", "SEX\_TYPE",

"age\_group",

'smoke', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level', 'insurance\_type')

explanatory\_vars\_w\_interaction\_term <- c("study\_group",

"SEX\_TYPE", "age\_group",

'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level', 'insurance\_type',

'pb\_real\*low\_bth\_wght',

'pb\_real\*PE')

# APO outcome 용 목록 생성 (GDM, PE, PA, PB, LBW 제외)

explanatory\_vars\_w\_interaction\_term <- c("study\_group",

"SEX\_TYPE", "age\_group",

'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'CE', 'PP',

'NP', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level')

# age group 에 대한 전체 (362만 아기) univariate analysis 를 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("age\_group")

# low\_bth\_wght univariate analysis 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("low\_bth\_wght")

# study group univariate analysis 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("study\_group")

## model 1 ##

explanatory\_vars\_model\_1 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP')

## model 2 ##

# DM

explanatory\_vars\_model\_2 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM', 'bmi\_category', 'sudogwon', 'income\_level')

# smoke

explanatory\_vars\_model\_2 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'smoke', 'bmi\_category', 'sudogwon', 'income\_level')

# smoke \* DM interaction term

explanatory\_vars\_model\_2 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM', 'smoke', 'bmi\_category', 'smoke\*DM', 'sudogwon', 'income\_level')

# w/ DM, HTN, SES - FINAL

explanatory\_vars\_model\_2 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM', 'sudogwon', 'income\_level')

# w/ DM, HTN

explanatory\_vars\_model\_2 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM')

# w/o DM, smoke, with BMI

explanatory\_vars\_model\_2 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'bmi\_category', 'sudogwon', 'income\_level')

## model 2 end. ##

## model 3 ##

# preterm birth

# explanatory\_vars\_model\_3 <- c("study\_group",

# "SEX\_TYPE", "age\_group", 'NP', 'HTN', 'smoke', 'bmi\_category',

# 'pb\_real', 'GDM', 'sudogwon', 'income\_level')

# low\_bth\_wght

# explanatory\_vars\_model\_3 <- c("study\_group",

# "SEX\_TYPE", "age\_group", 'NP', 'HTN', 'smoke', 'bmi\_category',

# 'low\_bth\_wght', 'GDM', 'sudogwon', 'income\_level')

# PE

# explanatory\_vars\_model\_3 <- c("study\_group",

# "SEX\_TYPE", "age\_group", 'NP', 'HTN', 'smoke', 'bmi\_category',

# 'PE', 'GDM', 'sudogwon', 'income\_level')

# pb\_real\*low\_bth\_wght, pb\_real\*PE

explanatory\_vars\_model\_3 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'smoke', 'bmi\_category',

'pb\_real', 'GDM', 'PE', 'low\_bth\_wght', 'pb\_real\*low\_bth\_wght',

'pb\_real\*PE', 'sudogwon', 'income\_level')

# PE, low\_bth\_wght

explanatory\_vars\_model\_3 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'smoke', 'bmi\_category',

'GDM', 'PE', 'low\_bth\_wght', 'sudogwon', 'income\_level')

# PE, low\_bth\_wght without GDM

explanatory\_vars\_model\_3 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'smoke', 'bmi\_category',

'PE', 'low\_bth\_wght', 'sudogwon', 'income\_level')

# low\_bth\_wght without GDM

explanatory\_vars\_model\_3 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'smoke', 'bmi\_category',

'low\_bth\_wght', 'sudogwon', 'income\_level')

# w/o BMI

# PE, PB, LBW

explanatory\_vars\_model\_3 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM',

'pb\_real', 'GDM', 'PE', 'low\_bth\_wght', 'pb\_real\*low\_bth\_wght',

'pb\_real\*PE','sudogwon', 'income\_level')

# all (final for neuro model) - FINAL

explanatory\_vars\_model\_3 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM',

'pb\_real', 'GDM', 'PE', 'low\_bth\_wght', 'pb\_real\*low\_bth\_wght',

'pb\_real\*PE','sudogwon', 'income\_level')

# all (final for neuro model) - w/o interaction - NOT FINAL

explanatory\_vars\_model\_3 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM',

'pb\_real', 'GDM', 'PE', 'low\_bth\_wght', 'sudogwon', 'income\_level')

# for APO outcome - cancel, APO는 model 2까지만!

# explanatory\_vars\_model\_3 <- c("study\_group",

# "SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM', )

## UNIVARIATE ANALYSIS ##

## 1. explanatory\_vars\_w\_interaction\_term ##

models <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("any\_cm ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models)

saveWorkbook(excelworkbook, file="Z:/zio/out/any\_cm\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_2 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("any\_neruo ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_2)

saveWorkbook(excelworkbook, file="Z:/zio/out/any\_neruo\_univariate\_results\_age\_whole.xlsx")

##########################################################

models\_2 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("APO ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_2)

saveWorkbook(excelworkbook, file="Z:/zio/out/APO\_univariate\_results.xlsx")

##########################################################

models\_3 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q00\_07 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_3)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q00\_07\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_4 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q10\_18 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_4)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q10\_18\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_5 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q20\_28 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_5)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q20\_28\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_6 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q30\_34 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_6)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q30\_34\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_7 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q35\_37 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_7)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q35\_37\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_8 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q38\_45 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_8)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q38\_45\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_9 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q50\_56 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_9)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q50\_56\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_10 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q60\_64 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_10)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q60\_64\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_11 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q65\_79 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_11)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q65\_79\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_12 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q80\_89 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_12)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q80\_89\_univariate\_results\_SG12\_only.xlsx")

##########################################################

models\_13 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("MDD ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_13)

saveWorkbook(excelworkbook, file="Z:/zio/out/MDD\_univariate\_results.xlsx")

##########################################################

models\_14 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("CDD ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_14)

saveWorkbook(excelworkbook, file="Z:/zio/out/CDD\_univariate\_results.xlsx")

##########################################################

models\_15 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("ADHD ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_15)

saveWorkbook(excelworkbook, file="Z:/zio/out/ADHD\_univariate\_results.xlsx")

##########################################################

models\_16 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("TICS ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_16)

saveWorkbook(excelworkbook, file="Z:/zio/out/TICS\_univariate\_results.xlsx")

##########################################################

models\_17 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("EFS ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_17)

saveWorkbook(excelworkbook, file="Z:/zio/out/EFS\_univariate\_results.xlsx")

##########################################################

### MULTIVARIATE ANALYSIS ###

# select storage by data

storage <- "Z:/zio/out/study\_group\_1\_2\_3/"

storage <- "Z:/zio/out/study\_group\_1\_2\_3-1/"

storage <- "Z:/zio/out/study\_group\_12\_3/"

storage <- "Z:/zio/out/study\_group\_12/"

## 1. MODEL 1 ##

# any\_cm

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("any\_cm ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

summary(mv\_reg)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_cm\_multi\_model1\_SG1\_2.xlsx", sep=""))

##########################################################

# any\_neruo

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("any\_neruo ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_neruo\_multi\_model1.xlsx", sep=""))

##########################################################

# Q00\_07

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q00\_07 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q00\_07\_multi\_model1.xlsx")

##########################################################

# Q10\_18

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q10\_18 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q10\_18\_multi\_model1.xlsx")

##########################################################

# Q20\_28

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q20\_28 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q20\_28\_multi\_model1.xlsx")

##########################################################

# Q30\_34

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q30\_34 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q30\_34\_multi\_model1.xlsx")

##########################################################

# Q35\_37

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q35\_37 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q35\_37\_multi\_model1.xlsx")

##########################################################

# Q38\_45

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q38\_45 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q38\_45\_multi\_model1.xlsx")

##########################################################

# Q50\_56

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q50\_56 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q50\_56\_multi\_model1.xlsx")

##########################################################

# Q60\_64

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q60\_64 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q60\_64\_multi\_model1.xlsx")

##########################################################

# Q65\_79

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q65\_79 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q65\_79\_multi\_model1.xlsx")

##########################################################

# Q80\_89

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q80\_89 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q80\_89\_multi\_model1.xlsx")

##########################################################

# MDD

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("MDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/MDD\_multi\_model1.xlsx")

##########################################################

# CDD

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("CDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/CDD\_multi\_model1.xlsx")

##########################################################

# ADHD

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("ADHD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/ADHD\_multi\_model1.xlsx")

##########################################################

# TICS

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("TICS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/TICS\_multi\_model1.xlsx")

##########################################################

# EFS

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("EFS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/EFS\_multi\_model1.xlsx")

##########################################################

## MODEL 2 ##

# any\_cm

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("any\_cm ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

table(data$study\_group)

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_cm\_multi\_model2\_SG12\_only.xlsx", sep=""))

##########################################################

# any\_neruo

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("any\_neruo ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_neruo\_multi\_model2\_ageRefchanged.xlsx", sep=""))

##########################################################

# APO

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("APO ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "APO\_multi\_model2.xlsx", sep=""))

##########################################################

# APO 세부 outcome 1 - GDM

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("GDM ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "GDM\_multi\_model2.xlsx", sep=""))

##########################################################

# APO 세부 outcome 2 - PE

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("PE ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "PE\_multi\_model2.xlsx", sep=""))

##########################################################

# APO 세부 outcome 3 - pb\_real

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("pb\_real ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "pb\_real\_multi\_model2.xlsx", sep=""))

##########################################################

# APO 세부 outcome 4 - PA

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("PA ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "PA\_multi\_model2.xlsx", sep=""))

##########################################################

# APO 세부 outcome 5 - low\_bth\_wght

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("low\_bth\_wght ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "low\_bth\_wght\_multi\_model2.xlsx", sep=""))

##########################################################

# Q00\_07

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q00\_07 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q00\_07\_multi\_model2.xlsx")

##########################################################

# Q10\_18

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q10\_18 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q10\_18\_multi\_model2.xlsx")

##########################################################

# Q20\_28

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q20\_28 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q20\_28\_multi\_model2.xlsx")

##########################################################

# Q30\_34

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q30\_34 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q30\_34\_multi\_model2.xlsx")

##########################################################

# Q35\_37

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q35\_37 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q35\_37\_multi\_model2.xlsx")

##########################################################

# Q38\_45

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q38\_45 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q38\_45\_multi\_model2.xlsx")

##########################################################

# Q50\_56

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q50\_56 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q50\_56\_multi\_model2.xlsx")

##########################################################

# Q60\_64

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q60\_64 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q60\_64\_multi\_model2.xlsx")

##########################################################

# Q65\_79

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q65\_79 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q65\_79\_multi\_model2.xlsx")

##########################################################

# Q80\_89

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q80\_89 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q80\_89\_multi\_model2.xlsx")

##########################################################

# MDD

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("MDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/MDD\_multi\_model2.xlsx")

##########################################################

# CDD

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("CDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/CDD\_multi\_model2.xlsx")

##########################################################

# ADHD

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("ADHD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/ADHD\_multi\_model2.xlsx")

##########################################################

# TICS

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("TICS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/TICS\_multi\_model2.xlsx")

##########################################################

# EFS

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("EFS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/EFS\_multi\_model2.xlsx")

##########################################################

## MODEL 3 ##

# any\_cm

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("any\_cm ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_cm\_multi\_model3\_ageRefchanged.xlsx", sep=""))

##########################################################

# any\_neruo

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("any\_neruo ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_neruo\_multi\_model3\_woLBWnull.xlsx", sep=""))

##########################################################

# APO

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("APO ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "APO\_multi\_model3\_noInteraction\_noSES.xlsx", sep=""))

##########################################################

# Q00\_07

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("Q00\_07 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q00\_07\_multi\_model3.xlsx")

##########################################################

# Q10\_18

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("Q10\_18 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q10\_18\_multi\_model3.xlsx")

##########################################################

# Q20\_28

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("Q20\_28 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q20\_28\_multi\_model3.xlsx")

##########################################################

# Q30\_34

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("Q30\_34 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q30\_34\_multi\_model3.xlsx")

##########################################################

# Q35\_37

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("Q35\_37 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q35\_37\_multi\_model3.xlsx")

##########################################################

# Q38\_45

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("Q38\_45 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q38\_45\_multi\_model3.xlsx")

##########################################################

# Q50\_56

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("Q50\_56 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q50\_56\_multi\_model3.xlsx")

##########################################################

# Q60\_64

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("Q60\_64 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q60\_64\_multi\_model3.xlsx")

##########################################################

# Q65\_79

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("Q65\_79 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q65\_79\_multi\_model3.xlsx")

##########################################################

# Q80\_89

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("Q80\_89 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q80\_89\_multi\_model3.xlsx")

##########################################################

# MDD

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("MDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/MDD\_multi\_model3.xlsx")

##########################################################

# CDD

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("CDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/CDD\_multi\_model3.xlsx")

##########################################################

# ADHD

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("ADHD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/ADHD\_multi\_model3.xlsx")

##########################################################

# TICS

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("TICS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/TICS\_multi\_model3.xlsx")

##########################################################

# EFS

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("EFS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file="Z:/zio/out/EFS\_multi\_model3.xlsx")

##########################################################

# GLM\_analysis - for APO outcome

data\_sg3 <- read.csv("Z:/zio/R/final\_data\_for\_analysis\_sg3\_apo.csv")

# 전체 362만 아기로 재정리한 데이터

data\_whole <- read.csv("Z:/zio/R/prediction\_dta.csv") # 3623766 263

### UNIVARIATE ANALYSIS ###

install.packages("gtsummary",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("car",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("openxlsx",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("r2glmm",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

library(broom)

library(dplyr)

library(stringr)

library(tidyr)

library(purrr)

library(car) # vif

library(openxlsx) # for multiple sheets

library(r2glmm) # r2 of model

# copy data needed

data <- data\_sg3

# factor categorical variables

data$study\_group <- factor(data$study\_group)

data$SEX\_TYPE <- factor(data$SEX\_TYPE)

data$SEX\_TYPE <- relevel(data$SEX\_TYPE, ref=2) # fetal sex reference를 female로 변경

data$smoke <- factor(data$smoke)

data$HTN <- factor(data$HTN)

data$DM <- factor(data$DM)

data$DSPDA <- factor(data$DSPDA)

data$MI <- factor(data$MI)

data$CHF <- factor(data$CHF)

data$CDO <- factor(data$CDO)

data$TIA <- factor(data$TIA)

data$CLD <- factor(data$CLD)

data$COPD <- factor(data$COPD)

data$Malignancy <- factor(data$Malignancy)

data$GDM <- factor(data$GDM)

data$PE <- factor(data$PE)

data$CE <- factor(data$CE)

data$PA <- factor(data$PA)

data$PP <- factor(data$PP)

data$NP <- factor(data$NP)

data$LGA <- factor(data$LGA)

data$pb\_real <- factor(data$pb\_real)

data$low\_bth\_wght <- factor(data$low\_bth\_wght)

data$bmi\_category <- factor(data$bmi\_category)

data$sudogwon <- factor(data$sudogwon)

data$income\_level <- factor(data$income\_level)

data$insurance\_type <- factor(data$insurance\_type)

data$any\_cm <- factor(data$any\_cm)

data$any\_neruo <- factor(data$any\_neruo)

data$APO <- factor(data$APO)

## original age group ##

data$age\_group <- factor(data$age\_group)

data$age\_group <- relevel(data$age\_group, ref=1) # age\_group reference를 1로 지정 (이미 자동으로 되어있지만, 확인차..)

## 세분화된 age group ## data\_sg3\_bmi\_imputed\_ageRef\_changed 사용할 때만 실행 !!! ##

data$age\_group <- factor(data$age\_group)

data$age\_group <- relevel(data$age\_group, ref=2) # age\_group reference를 20-29세로 지정

## final age group ## 이걸로 실행 ##

data$age\_group <- factor(data$age\_group)

data$age\_group <- relevel(data$age\_group, ref=2) # age\_group reference를 2로 지정 (25-29세)

## study group 1, 2를 합칠 때만 실행 !!! ##

data[which(data$study\_group == 2), 'study\_group'] <- 1

table(data$study\_group)

colSums(is.na(data))

## study group 3를 삭제할 때만 실행 !!! ##

data <- data[which(data$study\_group != 3),]

table(data$study\_group)

# define variables of interest

explanatory\_vars <- c("study\_group", "SEX\_TYPE",

"age\_group",

'smoke', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level', 'insurance\_type')

explanatory\_vars\_w\_interaction\_term <- c("study\_group",

"SEX\_TYPE", "age\_group",

'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level', 'insurance\_type',

'pb\_real\*low\_bth\_wght',

'pb\_real\*PE')

# APO outcome 용 목록 생성 (GDM, PE, PA, PB, LBW 제외)

explanatory\_vars\_w\_interaction\_term <- c("study\_group",

"SEX\_TYPE", "age\_group",

'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'CE', 'PP',

'NP', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level')

# age group 에 대한 전체 (362만 아기) univariate analysis 를 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("age\_group")

# low\_bth\_wght univariate analysis 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("low\_bth\_wght")

# study group univariate analysis 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("study\_group")

## model 1 ##

explanatory\_vars\_model\_1 <- c("study\_group", "age\_group", 'NP')

## model 2 ##

explanatory\_vars\_model\_2 <- c("study\_group", "age\_group", 'NP', 'bmi\_category', 'smoke', 'HTN', 'DM', 'DSPDA')

## model 3 ##

explanatory\_vars\_model\_3 <- c("study\_group", "age\_group", 'NP', 'bmi\_category', 'smoke', 'HTN', 'DM', 'DSPDA', 'sudogwon', 'income\_level')

## UNIVARIATE ANALYSIS ##

## 1. explanatory\_vars\_w\_interaction\_term ##

models <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("APO ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models)

saveWorkbook(excelworkbook, file="Z:/zio/out/APO\_SG\_univariate\_result.xlsx")

##########################################################

models\_2 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("PE ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_2)

saveWorkbook(excelworkbook, file="Z:/zio/out/PE\_SG\_univariate\_result.xlsx")

##########################################################

models\_2 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("GDM ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_2)

saveWorkbook(excelworkbook, file="Z:/zio/out/GDM\_SG\_univariate\_result.xlsx")

##########################################################

models\_3 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("pb\_real ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_3)

saveWorkbook(excelworkbook, file="Z:/zio/out/pb\_real\_SG\_univariate\_result.xlsx")

##########################################################

models\_4 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("PA ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_4)

saveWorkbook(excelworkbook, file="Z:/zio/out/PA\_SG\_univariate\_result.xlsx")

##########################################################

models\_5 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("low\_bth\_wght ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_5)

saveWorkbook(excelworkbook, file="Z:/zio/out/low\_bth\_wght\_SG\_univariate\_result.xlsx")

##########################################################

### MULTIVARIATE ANALYSIS ###

# select storage by data

storage <- "Z:/zio/out/study\_group\_1\_2\_3/"

storage <- "Z:/zio/out/study\_group\_1\_2\_3-1/"

storage <- "Z:/zio/out/study\_group\_12\_3/"

storage <- "Z:/zio/out/study\_group\_12/"

storage <- "Z:/zio/out/APO\_abstract/"

## 1. MODEL 1 ##

# APO

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("APO ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

summary(mv\_reg)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "APO\_multi\_model1.xlsx", sep=""))

##########################################################

# PE

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("PE ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "PE\_multi\_model1.xlsx", sep=""))

##########################################################

# GDM

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("GDM ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "GDM\_multi\_model1.xlsx", sep=""))

##########################################################

# pb\_real

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("pb\_real ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "pb\_real\_multi\_model1.xlsx", sep=""))

##########################################################

# PA

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("PA ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "PA\_multi\_model1.xlsx", sep=""))

##########################################################

# low\_bth\_wght

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("low\_bth\_wght ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "low\_bth\_wght\_multi\_model1.xlsx", sep=""))

##########################################################

## MODEL 2 ##

# APO

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("APO ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "APO\_multi\_model2.xlsx", sep=""))

##########################################################

# APO 세부 outcome 1 - GDM

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("GDM ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "GDM\_multi\_model2.xlsx", sep=""))

##########################################################

# APO 세부 outcome 2 - PE

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("PE ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "PE\_multi\_model2.xlsx", sep=""))

##########################################################

# APO 세부 outcome 3 - pb\_real

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("pb\_real ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "pb\_real\_multi\_model2.xlsx", sep=""))

##########################################################

# APO 세부 outcome 4 - PA

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("PA ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "PA\_multi\_model2.xlsx", sep=""))

##########################################################

# APO 세부 outcome 5 - low\_bth\_wght

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("low\_bth\_wght ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "low\_bth\_wght\_multi\_model2.xlsx", sep=""))

##########################################################

## MODEL 3 ##

# APO

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("APO ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "APO\_multi\_model3.xlsx", sep=""))

##########################################################

# PE

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("PE ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "PE\_multi\_model3.xlsx", sep=""))

##########################################################

# GDM

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("GDM ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "GDM\_multi\_model3.xlsx", sep=""))

##########################################################

# pb\_real

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("pb\_real ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "pb\_real\_multi\_model3.xlsx", sep=""))

##########################################################

# PA

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("PA ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "PA\_multi\_model3.xlsx", sep=""))

##########################################################

# low\_bth\_wght

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("low\_bth\_wght ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "low\_bth\_wght\_multi\_model3.xlsx", sep=""))

##########################################################

# GLM\_analysis - for CM outcome

data\_sg3 <- read.csv("Z:/zio/R/final\_data\_for\_analysis\_sg3.csv")

### UNIVARIATE ANALYSIS ###

install.packages("gtsummary",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("car",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("openxlsx",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("r2glmm",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

library(broom)

library(dplyr)

library(stringr)

library(tidyr)

library(purrr)

library(car) # vif

library(openxlsx) # for multiple sheets

library(r2glmm) # r2 of model

# copy data needed

data <- data\_sg3

data <- data\_sg3\_1 # study group == 3 of this data is actually study group 3-1

data <- data\_sg3\_bmi\_imputed # bmi mising values imputed with mean value of each study group

data <- data\_sg3\_bmi\_imputed\_ageRef\_changed

data <- data\_whole

# factor categorical variables

data$study\_group <- factor(data$study\_group)

data$SEX\_TYPE <- factor(data$SEX\_TYPE)

data$SEX\_TYPE <- relevel(data$SEX\_TYPE, ref=2) # fetal sex reference를 female로 변경

data$smoke <- factor(data$smoke)

data$HTN <- factor(data$HTN)

data$DM <- factor(data$DM)

data$DSPDA <- factor(data$DSPDA)

data$MI <- factor(data$MI)

data$CHF <- factor(data$CHF)

data$CDO <- factor(data$CDO)

data$TIA <- factor(data$TIA)

data$CLD <- factor(data$CLD)

data$COPD <- factor(data$COPD)

data$Malignancy <- factor(data$Malignancy)

data$GDM <- factor(data$GDM)

data$PE <- factor(data$PE)

data$CE <- factor(data$CE)

data$PA <- factor(data$PA)

data$PP <- factor(data$PP)

data$NP <- factor(data$NP)

data$LGA <- factor(data$LGA)

data$pb\_real <- factor(data$pb\_real)

data$low\_bth\_wght <- factor(data$low\_bth\_wght)

data$bmi\_category <- factor(data$bmi\_category)

data$sudogwon <- factor(data$sudogwon)

data$income\_level <- factor(data$income\_level)

data$insurance\_type <- factor(data$insurance\_type)

data$any\_cm <- factor(data$any\_cm)

data$any\_neruo <- factor(data$any\_neruo)

data$APO <- factor(data$APO)

## final age group ## 이걸로 실행 ##

data$age\_group <- factor(data$age\_group)

data$age\_group <- relevel(data$age\_group, ref=2) # age\_group reference를 2로 지정 (25-29세)

## study group 1, 2를 합칠 때만 실행 !!! ## - CM outcome의 경우 실행

data[which(data$study\_group == 2), 'study\_group'] <- 1

table(data$study\_group)

colSums(is.na(data))

## study group 3를 삭제할 때만 실행 !!! ##

data <- data[which(data$study\_group != 3),]

table(data$study\_group)

# define variables of interest

explanatory\_vars <- c("study\_group", "SEX\_TYPE",

"age\_group",

'smoke', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level', 'insurance\_type')

explanatory\_vars\_w\_interaction\_term <- c("study\_group",

"SEX\_TYPE", "age\_group",

'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level', 'insurance\_type',

'pb\_real\*low\_bth\_wght',

'pb\_real\*PE')

# APO outcome 용 목록 생성 (GDM, PE, PA, PB, LBW 제외)

explanatory\_vars\_w\_interaction\_term <- c("study\_group",

"SEX\_TYPE", "age\_group",

'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'CE', 'PP',

'NP', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level')

# age group 에 대한 전체 (362만 아기) univariate analysis 를 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("age\_group")

# low\_bth\_wght univariate analysis 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("low\_bth\_wght")

# study group univariate analysis 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("study\_group")

## model 1 ##

explanatory\_vars\_model\_1 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP')

## model 2 ##

# w/ DM, HTN, SES - FINAL

explanatory\_vars\_model\_2 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM', 'sudogwon', 'income\_level')

## model 2 end. ##

## model 3 ##

# all (final for neuro model) - FINAL

explanatory\_vars\_model\_3 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM',

'pb\_real', 'GDM', 'PE', 'low\_bth\_wght', 'pb\_real\*low\_bth\_wght',

'pb\_real\*PE','sudogwon', 'income\_level')

## UNIVARIATE ANALYSIS ##

## 1. explanatory\_vars\_w\_interaction\_term ##

models <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("any\_cm ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models)

saveWorkbook(excelworkbook, file="Z:/zio/out/any\_cm\_univariate\_results.xlsx")

##########################################################

models\_3 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q00\_07 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_3)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q00\_07\_univariate\_results.xlsx")

##########################################################

models\_4 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q10\_18 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_4)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q10\_18\_univariate\_results.xlsx")

##########################################################

models\_5 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q20\_28 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_5)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q20\_28\_univariate\_results.xlsx")

##########################################################

models\_6 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q30\_34 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_6)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q30\_34\_univariate\_results.xlsx")

##########################################################

models\_7 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q35\_37 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_7)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q35\_37\_univariate\_results.xlsx")

##########################################################

models\_8 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q38\_45 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_8)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q38\_45\_univariate\_results.xlsx")

##########################################################

models\_9 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q50\_56 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_9)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q50\_56\_univariate\_results.xlsx")

##########################################################

models\_10 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q60\_64 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_10)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q60\_64\_univariate\_results.xlsx")

##########################################################

models\_11 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q65\_79 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_11)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q65\_79\_univariate\_results.xlsx")

##########################################################

models\_12 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("Q80\_89 ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_12)

saveWorkbook(excelworkbook, file="Z:/zio/out/Q80\_89\_univariate\_results.xlsx")

##########################################################

### MULTIVARIATE ANALYSIS ###

# select storage by data

storage <- "Z:/zio/out/study\_group\_1\_2\_3/"

storage <- "Z:/zio/out/study\_group\_1\_2\_3-1/"

storage <- "Z:/zio/out/study\_group\_12\_3/"

storage <- "Z:/zio/out/study\_group\_12/"

## 1. MODEL 1 ##

# any\_cm

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("any\_cm ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

summary(mv\_reg)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_cm\_multi\_model1.xlsx", sep=""))

##########################################################

# Q00\_07

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q00\_07 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q00\_07\_multi\_model1.xlsx", sep=""))

##########################################################

# Q10\_18

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q10\_18 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q10\_18\_multi\_model1.xlsx", sep=""))

##########################################################

# Q20\_28

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q20\_28 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q20\_28\_multi\_model1.xlsx", sep=""))

##########################################################

# Q30\_34

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q30\_34 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q30\_34\_multi\_model1.xlsx", sep=""))

##########################################################

# Q35\_37

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q35\_37 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q35\_37\_multi\_model1.xlsx", sep=""))

##########################################################

# Q38\_45

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q38\_45 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q38\_45\_multi\_model1.xlsx", sep=""))

##########################################################

# Q50\_56

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q50\_56 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q50\_56\_multi\_model1.xlsx", sep=""))

##########################################################

# Q60\_64

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q60\_64 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q60\_64\_multi\_model1.xlsx", sep=""))

##########################################################

# Q65\_79

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q65\_79 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q65\_79\_multi\_model1.xlsx", sep=""))

##########################################################

# Q80\_89

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("Q80\_89 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q80\_89\_multi\_model1.xlsx", sep=""))

##########################################################

## MODEL 2 ##

# any\_cm

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("any\_cm ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

table(data$study\_group)

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_cm\_multi\_model2.xlsx", sep=""))

##########################################################

# Q00\_07

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q00\_07 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q00\_07\_multi\_model2.xlsx", sep=""))

##########################################################

# Q10\_18

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q10\_18 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q10\_18\_multi\_model2.xlsx", sep=""))

##########################################################

# Q20\_28

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q20\_28 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q20\_28\_multi\_model2.xlsx", sep=""))

##########################################################

# Q30\_34

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q30\_34 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q30\_34\_multi\_model2.xlsx", sep=""))

##########################################################

# Q35\_37

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q35\_37 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q35\_37\_multi\_model2.xlsx", sep=""))

##########################################################

# Q38\_45

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q38\_45 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q38\_45\_multi\_model2.xlsx", sep=""))

##########################################################

# Q50\_56

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q50\_56 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q50\_56\_multi\_model2.xlsx", sep=""))

##########################################################

# Q60\_64

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q60\_64 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q60\_64\_multi\_model2.xlsx", sep=""))

##########################################################

# Q65\_79

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q65\_79 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q65\_79\_multi\_model2.xlsx", sep=""))

##########################################################

# Q80\_89

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("Q80\_89 ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "Q80\_89\_multi\_model2.xlsx", sep=""))

##########################################################

# ## MODEL 3 ##

#

# # any\_cm

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("any\_cm ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

# with(summary(mv\_reg), 1 - deviance/null.deviance)

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "any\_cm\_multi\_model3.xlsx", sep=""))

# ##########################################################

#

#

#

# # Q00\_07

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("Q00\_07 ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "Q00\_07\_multi\_model3.xlsx", sep=""))

# ##########################################################

#

#

# # Q10\_18

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("Q10\_18 ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "Q10\_18\_multi\_model3.xlsx", sep=""))

# ##########################################################

#

#

# # Q20\_28

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("Q20\_28 ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "Q20\_28\_multi\_model3.xlsx", sep=""))

# ##########################################################

#

#

# # Q30\_34

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("Q30\_34 ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "Q30\_34\_multi\_model3.xlsx", sep=""))

# ##########################################################

#

#

# # Q35\_37

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("Q35\_37 ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "Q35\_37\_multi\_model3.xlsx", sep=""))

# ##########################################################

#

#

# # Q38\_45

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("Q38\_45 ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "Q38\_45\_multi\_model3.xlsx", sep=""))

# ##########################################################

#

#

# # Q50\_56

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("Q50\_56 ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "Q50\_56\_multi\_model3.xlsx", sep=""))

# ##########################################################

#

#

# # Q60\_64

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("Q60\_64 ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "Q60\_64\_multi\_model3.xlsx", sep=""))

# ##########################################################

#

#

#

# # Q65\_79

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("Q65\_79 ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "Q65\_79\_multi\_model3.xlsx", sep=""))

# ##########################################################

#

#

# # Q80\_89

# mv\_reg <- explanatory\_vars\_model\_3 %>%

# str\_c(collapse = "+") %>%

# str\_c("Q80\_89 ~ ", .) %>%

# glm(family="binomial", data=data)

#

# mv\_tab\_base <- mv\_reg %>%

# tidy(exponentiate = TRUE, conf.int=TRUE)

#

# vif <- as.data.frame(vif(mv\_reg))

#

# excelworkbook <- createWorkbook("excelworkbook")

# addWorksheet(excelworkbook, "result")

# addWorksheet(excelworkbook, "VIF")

#

# writeDataTable(excelworkbook, "result", mv\_tab\_base)

# writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

#

# saveWorkbook(excelworkbook, file=paste(storage, "Q80\_89\_multi\_model3.xlsx", sep=""))

# ##########################################################

# GLM\_analysis - for NEURO outcome

data\_sg3 <- read.csv("Z:/zio/R/final\_data\_for\_analysis\_sg3.csv")

### UNIVARIATE ANALYSIS ###

install.packages("gtsummary",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("car",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("openxlsx",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("r2glmm",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

library(broom)

library(dplyr)

library(stringr)

library(tidyr)

library(purrr)

library(car) # vif

library(openxlsx) # for multiple sheets

library(r2glmm) # r2 of model

# copy data needed

data <- data\_sg3

data <- data\_sg3\_1 # study group == 3 of this data is actually study group 3-1

data <- data\_sg3\_bmi\_imputed # bmi mising values imputed with mean value of each study group

data <- data\_sg3\_bmi\_imputed\_ageRef\_changed

data <- data\_whole

# factor categorical variables

data$study\_group <- factor(data$study\_group)

data$SEX\_TYPE <- factor(data$SEX\_TYPE)

data$SEX\_TYPE <- relevel(data$SEX\_TYPE, ref=2) # fetal sex reference를 female로 변경

data$smoke <- factor(data$smoke)

data$HTN <- factor(data$HTN)

data$DM <- factor(data$DM)

data$DSPDA <- factor(data$DSPDA)

data$MI <- factor(data$MI)

data$CHF <- factor(data$CHF)

data$CDO <- factor(data$CDO)

data$TIA <- factor(data$TIA)

data$CLD <- factor(data$CLD)

data$COPD <- factor(data$COPD)

data$Malignancy <- factor(data$Malignancy)

data$GDM <- factor(data$GDM)

data$PE <- factor(data$PE)

data$CE <- factor(data$CE)

data$PA <- factor(data$PA)

data$PP <- factor(data$PP)

data$NP <- factor(data$NP)

data$LGA <- factor(data$LGA)

data$pb\_real <- factor(data$pb\_real)

data$low\_bth\_wght <- factor(data$low\_bth\_wght)

data$bmi\_category <- factor(data$bmi\_category)

data$sudogwon <- factor(data$sudogwon)

data$income\_level <- factor(data$income\_level)

data$insurance\_type <- factor(data$insurance\_type)

data$any\_cm <- factor(data$any\_cm)

data$any\_neruo <- factor(data$any\_neruo)

data$APO <- factor(data$APO)

## final age group ## 이걸로 실행 ##

data$age\_group <- factor(data$age\_group)

data$age\_group <- relevel(data$age\_group, ref=2) # age\_group reference를 2로 지정 (25-29세)

colSums(is.na(data))

## study group 3를 삭제할 때만 실행 !!! ##

data <- data[which(data$study\_group != 3),]

table(data$study\_group)

# define variables of interest

explanatory\_vars <- c("study\_group", "SEX\_TYPE",

"age\_group",

'smoke', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level', 'insurance\_type')

explanatory\_vars\_w\_interaction\_term <- c("study\_group",

"SEX\_TYPE", "age\_group",

'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level', 'insurance\_type',

'pb\_real\*low\_bth\_wght',

'pb\_real\*PE')

# APO outcome 용 목록 생성 (GDM, PE, PA, PB, LBW 제외)

explanatory\_vars\_w\_interaction\_term <- c("study\_group",

"SEX\_TYPE", "age\_group",

'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'CE', 'PP',

'NP', 'bmi\_category', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA',

'G1E\_URN\_PROT', 'G1E\_TOT\_CHOL',

'sudogwon', 'income\_level')

# age group 에 대한 전체 (362만 아기) univariate analysis 를 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("age\_group")

# low\_bth\_wght univariate analysis 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("low\_bth\_wght")

# study group univariate analysis 위해 임시로 생성

explanatory\_vars\_w\_interaction\_term <- c("study\_group")

## model 1 ##

explanatory\_vars\_model\_1 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP')

## model 2 ##

# w/ DM, HTN, SES - FINAL

explanatory\_vars\_model\_2 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM', 'sudogwon', 'income\_level')

## model 2 end. ##

## model 3 ##

# all (final for neuro model) - FINAL

explanatory\_vars\_model\_3 <- c("study\_group",

"SEX\_TYPE", "age\_group", 'NP', 'HTN', 'DM',

'pb\_real', 'GDM', 'PE', 'low\_bth\_wght', 'pb\_real\*low\_bth\_wght',

'pb\_real\*PE','sudogwon', 'income\_level')

## UNIVARIATE ANALYSIS ##

## 1. explanatory\_vars\_w\_interaction\_term ##

models\_2 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("any\_neruo ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_2)

saveWorkbook(excelworkbook, file="Z:/zio/out/any\_neruo\_univariate\_results.xlsx")

##########################################################

models\_13 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("MDD ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_13)

saveWorkbook(excelworkbook, file="Z:/zio/out/MDD\_univariate\_results.xlsx")

##########################################################

models\_14 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("CDD ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_14)

saveWorkbook(excelworkbook, file="Z:/zio/out/CDD\_univariate\_results.xlsx")

##########################################################

models\_15 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("ADHD ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_15)

saveWorkbook(excelworkbook, file="Z:/zio/out/ADHD\_univariate\_results.xlsx")

##########################################################

models\_16 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("TICS ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_16)

saveWorkbook(excelworkbook, file="Z:/zio/out/TICS\_univariate\_results.xlsx")

##########################################################

models\_17 <- explanatory\_vars\_w\_interaction\_term %>%

str\_c("EFS ~ ", .) %>%

# iterate through each univariate formula

map(

.f = ~glm(

formula=as.formula(.x),

family="binomial",

data=data)) %>%

# tidy up each of the glm regression

map(

.f = ~tidy(

.x,

exponentiate=TRUE,

conf.int=TRUE)) %>%

bind\_rows()

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", models\_17)

saveWorkbook(excelworkbook, file="Z:/zio/out/EFS\_univariate\_results.xlsx")

##########################################################

### MULTIVARIATE ANALYSIS ###

# select storage by data

storage <- "Z:/zio/out/study\_group\_1\_2\_3/"

storage <- "Z:/zio/out/study\_group\_1\_2\_3-1/"

storage <- "Z:/zio/out/study\_group\_12\_3/"

storage <- "Z:/zio/out/study\_group\_12/"

## 1. MODEL 1 ##

# any\_neruo

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("any\_neruo ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_neruo\_multi\_model1.xlsx", sep=""))

##########################################################

# MDD

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("MDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "MDD\_multi\_model1.xlsx", sep=""))

##########################################################

# CDD

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("CDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "CDD\_multi\_model1.xlsx", sep=""))

##########################################################

# ADHD

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("ADHD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "ADHD\_multi\_model1.xlsx", sep=""))

##########################################################

# TICS

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("TICS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "TICS\_multi\_model1.xlsx", sep=""))

##########################################################

# EFS

mv\_reg <- explanatory\_vars\_model\_1 %>%

str\_c(collapse = "+") %>%

str\_c("EFS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "EFS\_multi\_model1.xlsx", sep=""))

##########################################################

## MODEL 2 ##

# any\_neruo

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("any\_neruo ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_neruo\_multi\_model2.xlsx", sep=""))

##########################################################

# MDD

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("MDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "MDD\_multi\_model2.xlsx", sep=""))

##########################################################

# CDD

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("CDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "CDD\_multi\_model2.xlsx", sep=""))

##########################################################

# ADHD

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("ADHD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "ADHD\_multi\_model2.xlsx", sep=""))

##########################################################

# TICS

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("TICS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "TICS\_multi\_model2.xlsx", sep=""))

##########################################################

# EFS

mv\_reg <- explanatory\_vars\_model\_2 %>%

str\_c(collapse = "+") %>%

str\_c("EFS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "EFS\_multi\_model2.xlsx", sep=""))

##########################################################

## MODEL 3 ##

# any\_neruo

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("any\_neruo ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

with(summary(mv\_reg), 1 - deviance/null.deviance)

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "any\_neruo\_multi\_model3.xlsx", sep=""))

##########################################################

# MDD

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("MDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "MDD\_multi\_model3.xlsx", sep=""))

##########################################################

# CDD

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("CDD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "CDD\_multi\_model3.xlsx", sep=""))

##########################################################

# ADHD

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("ADHD ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "ADHD\_multi\_model3.xlsx", sep=""))

##########################################################

# TICS

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("TICS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "TICS\_multi\_model3.xlsx", sep=""))

##########################################################

# EFS

mv\_reg <- explanatory\_vars\_model\_3 %>%

str\_c(collapse = "+") %>%

str\_c("EFS ~ ", .) %>%

glm(family="binomial", data=data)

mv\_tab\_base <- mv\_reg %>%

tidy(exponentiate = TRUE, conf.int=TRUE)

vif <- as.data.frame(vif(mv\_reg))

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

addWorksheet(excelworkbook, "VIF")

writeDataTable(excelworkbook, "result", mv\_tab\_base)

writeDataTable(excelworkbook, "VIF", vif, rowNames=TRUE)

saveWorkbook(excelworkbook, file=paste(storage, "EFS\_multi\_model3.xlsx", sep=""))

##########################################################

# Figures and Plots

install.packages("Publish",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("data.table",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

# install.packages("gt",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("openxlsx",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("cmprsk",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("survminer",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("ggsurvfit",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

library(Publish)

library(dplyr)

library(data.table)

library(openxlsx)

library(cmprsk)

library(survminer)

library(ggsurvfit)

data\_sg3 <- read.csv("Z:/zio/R/data/final\_data\_for\_analysis\_sg3.csv")

# copy data needed

data <- data\_sg3

# factor categorical variables

data$study\_group <- factor(data$study\_group)

data$SEX\_TYPE <- factor(data$SEX\_TYPE)

data$SEX\_TYPE <- relevel(data$SEX\_TYPE, ref=2) # fetal sex reference를 female로 변경

data$smoke <- factor(data$smoke)

data$HTN <- factor(data$HTN)

data$DM <- factor(data$DM)

data$DSPDA <- factor(data$DSPDA)

data$MI <- factor(data$MI)

data$CHF <- factor(data$CHF)

data$CDO <- factor(data$CDO)

data$TIA <- factor(data$TIA)

data$CLD <- factor(data$CLD)

data$COPD <- factor(data$COPD)

data$Malignancy <- factor(data$Malignancy)

data$GDM <- factor(data$GDM)

data$PE <- factor(data$PE)

data$CE <- factor(data$CE)

data$PA <- factor(data$PA)

data$PP <- factor(data$PP)

data$NP <- factor(data$NP)

data$LGA <- factor(data$LGA)

data$pb\_real <- factor(data$pb\_real)

data$low\_bth\_wght <- factor(data$low\_bth\_wght)

data$bmi\_category <- factor(data$bmi\_category)

data$sudogwon <- factor(data$sudogwon)

data$income\_level <- factor(data$income\_level)

data$insurance\_type <- factor(data$insurance\_type)

data$any\_cm <- factor(data$any\_cm)

data$any\_neruo <- factor(data$any\_neruo)

data$APO <- factor(data$APO)

## final age group ## 이걸로 실행 ##

data$age\_group <- factor(data$age\_group)

data$age\_group <- relevel(data$age\_group, ref=2) # age\_group reference를 2로 지정 (25-29세)

##### CM paper #####

## study group 1, 2를 합칠 때만 실행 !!! ## - CM outcome의 경우 실행

data[which(data$study\_group == 2), 'study\_group'] <- 1

table(data$study\_group)

### Figure 2. Forest plots for subgroup analysis of congenital malformation outcomes by maternal kidney disease.

## (a) CKD group

# study group 3 <-> control group

dataSG3 <- data[which(data$study\_group != 1),]

table(dataSG3$study\_group)

# create new study group variable

# control group should be only control group (KT+ESKD group should not be included)

dataSG3$study\_group\_ckd <- ifelse(dataSG3$study\_group == 3, 1, 0)

dataSG3$study\_group\_ckd <- factor(dataSG3$study\_group\_ckd)

table(dataSG3$study\_group\_ckd) # 0: 713776, 1: 69648

# create new factor variables (not used orginally, created only for subgroup analysis)

dataSG3$age\_group\_30\_over <- ifelse(dataSG3$mother\_age >= 30, 1, 0)

dataSG3$age\_group\_30\_over <- factor(dataSG3$age\_group\_30\_over)

table(dataSG3$age\_group\_30\_over)

dataSG3$eGFR\_group <- case\_when(

is.na(dataSG3$G1E\_GFR) ~ 6,

dataSG3$G1E\_GFR > 90 ~ 5,

dataSG3$G1E\_GFR >= 60 & dataSG3$G1E\_GFR <= 90 ~ 4,

dataSG3$G1E\_GFR >= 45 & dataSG3$G1E\_GFR < 60 ~ 3,

dataSG3$G1E\_GFR >= 30 & dataSG3$G1E\_GFR < 45 ~ 2,

dataSG3$G1E\_GFR < 30 ~ 1

)

dataSG3$eGFR\_group <- factor(dataSG3$eGFR\_group)

table(dataSG3$eGFR\_group)

# fit and subgroup analysis

subgroups <- c("age\_group\_30\_over", "HTN", "DM", "income\_level", "eGFR\_group", "sudogwon", "NP")

fit <- glm(any\_cm ~ study\_group\_ckd, data = dataSG3, family=binomial)

summary(fit)

sub\_fit <- subgroupAnalysis(fit,

data = dataSG3,

treatment = "study\_group\_ckd",

subgroups = subgroups) # 변수가 factor화 되어 있으면 에러남

sub\_fit\_tbl <- copy(sub\_fit) %>% as.data.table()

# 보기 좋게 정리

sub\_fit\_tbl[, `:=`(

or\_ci = paste0(format(round(OddsRatio, 2), nsmall=2),

' (', format(round(Lower, 2),nsmall=2), '-',

format(round(Upper,2), nsmall=2),

')'),

control=paste0(event\_0, '/', sample\_0),

case=paste0(event\_1, '/', sample\_1))]

sub\_fit\_tbl[, .(subgroups, level, case, control, or\_ci, pinteraction)] %>%

group\_by(subgroups)

# %>%

# gt() %>%

# cols\_label(

# level=md("\*\*Subgroup\*\*"),

# case=md("\*\*Case\*\*"),

# control=md("\*\*Control\*\*"),

# or\_ci=md("\*\*OR (95% CI)\*\*"),

# pinteraction=md("\*\*\*P\* for interaction\*\*")

# ) %>%

# fmt\_number(

# column=pinteraction,

# decimal=4

# ) %>%

# cols\_align(

# align="center",

# columns=-level

# ) %>%

# cols\_align(

# align="left",

# columns=level

# ) %>%

# tab\_style(

# style=list(cell\_text(weight="bold")),

# locations=cells\_row\_groups()

# )

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", sub\_fit\_tbl)

saveWorkbook(excelworkbook, file="Z:/zio/out/cm\_figure2\_ckd.xlsx")

## (b) KT+ESKD group

# study group 1+2 <-> control group

dataSG3 <- data[which(data$study\_group != 3),]

table(dataSG3$study\_group)

# create new study group variable

# control group should be only control group (ckd group should not be included)

dataSG3$study\_group\_kteskd <- ifelse(dataSG3$study\_group != 0, 1, 0)

dataSG3$study\_group\_kteskd <- factor(dataSG3$study\_group\_kteskd)

table(dataSG3$study\_group\_kteskd) # 0: 713776, 1: 311

# create new factor variables (not used orginally, created only for subgroup analysis)

dataSG3$age\_group\_30\_over <- ifelse(dataSG3$mother\_age >= 30, 1, 0)

dataSG3$age\_group\_30\_over <- factor(dataSG3$age\_group\_30\_over)

table(dataSG3$age\_group\_30\_over)

dataSG3$eGFR\_group <- case\_when(

is.na(dataSG3$G1E\_GFR) ~ 6,

dataSG3$G1E\_GFR > 90 ~ 5,

dataSG3$G1E\_GFR >= 60 & dataSG3$G1E\_GFR <= 90 ~ 4,

dataSG3$G1E\_GFR >= 45 & dataSG3$G1E\_GFR < 60 ~ 3,

dataSG3$G1E\_GFR >= 30 & dataSG3$G1E\_GFR < 45 ~ 2,

dataSG3$G1E\_GFR < 30 ~ 1

)

dataSG3$eGFR\_group <- factor(dataSG3$eGFR\_group)

table(dataSG3$eGFR\_group)

# fit and subgroup analysis

subgroups <- c("age\_group\_30\_over", "HTN", "DM", "income\_level", "eGFR\_group", "sudogwon", "NP")

fit <- glm(any\_cm ~ study\_group\_kteskd, data = dataSG3, family=binomial)

summary(fit)

sub\_fit <- subgroupAnalysis(fit,

data = dataSG3,

treatment = "study\_group\_kteskd",

subgroups = subgroups) # 변수가 factor화 되어 있으면 에러남 - 이건 몇개만 factor화 했더니 에러 안남 ..? 이상하네

sub\_fit\_tbl <- copy(sub\_fit) %>% as.data.table()

# 보기 좋게 정리

sub\_fit\_tbl[, `:=`(

or\_ci = paste0(format(round(OddsRatio, 2), nsmall=2),

' (', format(round(Lower, 2),nsmall=2), '-',

format(round(Upper,2), nsmall=2),

')'),

control=paste0(event\_0, '/', sample\_0),

case=paste0(event\_1, '/', sample\_1))]

sub\_fit\_tbl[, .(subgroups, level, case, control, or\_ci, pinteraction)] %>%

group\_by(subgroups)

# %>%

# gt() %>%

# cols\_label(

# level=md("\*\*Subgroup\*\*"),

# case=md("\*\*Case\*\*"),

# control=md("\*\*Control\*\*"),

# or\_ci=md("\*\*OR (95% CI)\*\*"),

# pinteraction=md("\*\*\*P\* for interaction\*\*")

# ) %>%

# fmt\_number(

# column=pinteraction,

# decimal=4

# ) %>%

# cols\_align(

# align="center",

# columns=-level

# ) %>%

# cols\_align(

# align="left",

# columns=level

# ) %>%

# tab\_style(

# style=list(cell\_text(weight="bold")),

# locations=cells\_row\_groups()

# )

excelworkbook <- createWorkbook("excelworkbook")

addWorksheet(excelworkbook, "result")

writeDataTable(excelworkbook, "result", sub\_fit\_tbl)

saveWorkbook(excelworkbook, file="Z:/zio/out/cm\_figure2\_kteskd.xlsx")

##### NEURO paper #####

##Figure 3. Cumulative incidence of all adverse neurodevelopmental outcomes

##(x axis: Follow-up time (years) / y axis: cumulative incidence)

# 누적 발생률 plot 그리기 - 기본

ci\_fit <- cuminc(ftime=data$mdd\_time, fstatus=data$MDD, group=data$study\_group)

getwd()

png('cumulative\_incidence\_plot\_MDD.png')

plot(ci\_fit, xlab="Time (days)", ylab="Cumulative Incidence", main="Cumulative Incidence Function")

dev.off()

# 누적발생률 plot 그리기 - 심화

# time 단위 변경 (days -> years)

data$mdd\_time <- ifelse(data$mdd\_time < 0, 0, data$mdd\_time)

data$mdd\_time\_years <- data$mdd\_time / 365.25

data$cdd\_time <- ifelse(data$cdd\_time < 0, 0, data$cdd\_time)

data$cdd\_time\_years <- data$cdd\_time / 365.25

data$adhd\_time <- ifelse(data$adhd\_time < 0, 0, data$adhd\_time)

data$adhd\_time\_years <- data$adhd\_time / 365.25

data$tics\_time <- ifelse(data$tics\_time < 0, 0, data$tics\_time)

data$tics\_time\_years <- data$tics\_time / 365.25

data$efs\_time <- ifelse(data$efs\_time < 0, 0, data$efs\_time)

data$efs\_time\_years <- data$efs\_time / 365.25

# study group label 변경

data$study\_group <- factor(data$study\_group, levels=c(0,1,2,3), labels=c("Control", "KT", "ESKD", "CKD"))

# # 누적 발생률 계산

# ci\_fit <- cuminc(ftime=data$mdd\_time\_years, fstatus=data$MDD, group=data$study\_group)

# # plot

# p <- ggcompetingrisks(fit=ci\_fit,

# multiple\_panels = F,

# xlab="Time (years)",

# ylab="Cumulative Incidence",

# title="Motor Developmental Delay",

# risk.table=TRUE)

# #+

# # scale\_color\_discrete(labels=c(Control="Control", KT="KT", ESKD="ESKD", CKD="CKD"))

# #+

# # theme(legend.title=element\_blank())

# p$mapping <- aes(x=time, y=est, colour=group, linetype=event)

# p+labs(linetype="event", colour="")

fit <- survfit(Surv(mdd\_time\_years, MDD) ~ study\_group, data=data)

res <- ggsurvplot(fit, data=data, legend.title = "Groups", legend.labs=c("Control", "KT", "ESKD", "CKD"), risk.table=TRUE, risk.table.y.text.col=TRUE, fun='event')

res$plot <- res$plot + labs(title="Motor Developmental Delay")

res # export size width 1100

fit <- survfit(Surv(cdd\_time\_years, CDD) ~ study\_group, data=data)

res <- ggsurvplot(fit, data=data, legend.title = "Groups", legend.labs=c("Control", "KT", "ESKD", "CKD"), risk.table=TRUE, risk.table.y.text.col=TRUE, fun='event')

res$plot <- res$plot + labs(title="Cognitive developmental delay")

res

fit <- survfit(Surv(adhd\_time\_years, ADHD) ~ study\_group, data=data)

res <- ggsurvplot(fit, data=data, legend.title = "Groups", legend.labs=c("Control", "KT", "ESKD", "CKD"), risk.table=TRUE, risk.table.y.text.col=TRUE, fun='event')

res$plot <- res$plot + labs(title="Autism spectrum disorders / ADHD")

res

fit <- survfit(Surv(tics\_time\_years, TICS) ~ study\_group, data=data)

res <- ggsurvplot(fit, data=data, legend.title = "Groups", legend.labs=c("Control", "KT", "ESKD", "CKD"), risk.table=TRUE, risk.table.y.text.col=TRUE, fun='event')

res$plot <- res$plot + labs(title="Tics / stereotypic behavior")

res

fit <- survfit(Surv(efs\_time\_years, EFS) ~ study\_group, data=data)

res <- ggsurvplot(fit, data=data, legend.title = "Groups", legend.labs=c("Control", "KT", "ESKD", "CKD"), risk.table=TRUE, risk.table.y.text.col=TRUE, fun='event')

res$plot <- res$plot + labs(title="Epileptic / febrile seizures")

res

# Causal Inference

## alread downloaded ##

install.packages("Z:/zio/R/downloaded\_packages/grf\_2.3.2.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/Rcpp\_1.0.13.zip",repos=NULL,type="source")

## download ##

install.packages("dplyr",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

#install.packages("openxlsx",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("Rcpp",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

## load ##

library(grf)

library(dplyr) # cran에서 받지 않고 받아놓은 거 쓰면 에러남

#library(openxlsx)

library(Rcpp)

packageVersion("grf")

packageVersion("dplyr")

packageVersion("Rcpp") # 1.0.13 # ㅠㅠㅠㅠ 돌아간다!!!

memory.size(max=T) # memory size of system

memory.limit() # 최대 메모리 사용 한계

memory.size(max=F) # memory usage

rm(list=ls()) # to remove all variable

rm(data\_sg3) # to remove only one variable

gc() # memory clear

memory.limit(size=50000) # memory limit을 늘리기 (if 10000 -> 10gb)

# read data

data\_sg3 <- read.csv("Z:/zio/R/data/final\_data\_for\_analysis\_sg3.csv")

# read rdata

load(file="Z:/zio/R/rdata/causal\_forest\_data.RData")

load(file="Z:/zio/R/rdata/causal\_forest\_data\_anyCM\_modeling.RData")

load(file="Z:/zio/R/rdata/causal\_forest\_data\_anyNEURO\_modeling.RData")

# copy data needed

data <- data\_sg3

data <- data\_sg3\_1 # study group == 3 of this data is actually study group 3-1

data <- data\_sg3\_bmi\_imputed # bmi mising values imputed with mean value of each study group

data <- data\_sg3\_bmi\_imputed\_ageRef\_changed

data <- data\_whole

# select variables needed

data <- data[, c("any\_cm", "any\_neruo", "APO", "study\_group", "SEX\_TYPE", "mother\_age", 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght',

'sudogwon', 'insurance\_type',

'HTN\_med', 'DM\_med', 'HLD\_med', 'Diuretic\_med', 'Antidepressant\_med',

'Antithrombotic\_med', 'Dialysis\_related\_med', 'Steroid\_med')]

# drop all NAs

data <- na.omit(data) # 754031 35

table(data$study\_group)

# create binary treatment variable

data$renal\_disease <- ifelse(data$study\_group != 0, 1, 0)

table(data$renal\_disease)

# cut control group to 20% (memory problem)

set.seed(123)

control\_sample <- data %>% filter(renal\_disease == 0) %>% sample\_frac(0.2)

case\_sample <- data %>% filter(renal\_disease != 0)

# cut control group to n% / case group to n% (memory problem) - 돌아가는지 테스트용

set.seed(123)

control\_sample <- data %>% filter(renal\_disease == 0) %>% sample\_frac(0.0001)

case\_sample <- data %>% filter(renal\_disease != 0) %>% sample\_frac(0.001)

# concat control group (downsized) and case group

balanced\_data <- bind\_rows(control\_sample, case\_sample)

# split train/test

fractionTraining <- 0.7

fractionTest <- 0.3

# compute sample sizes

sampleSizeTraining <- floor(fractionTraining \* nrow(balanced\_data))

sampleSizeTest <- floor(fractionTest \* nrow(balanced\_data))

# create the randomly-sampled indices for the dataframe. Use setdiff() to avoid overlapping subsets of indices

indicesTraining <- sort(sample(seq\_len(nrow(balanced\_data)), size=sampleSizeTraining))

indicesNotTraining <- setdiff(seq\_len(nrow(balanced\_data)), indicesTraining)

indicesTest <- sort(sample(indicesNotTraining, size=sampleSizeTest))

# output the two dataframe for train/test

dfTraining <- data[indicesTraining,]

dfTest <- data[indicesTest, ]

# save.image(file="Z:/zio/R/causal\_forest\_data.RData")

rm(case\_sample) # to remove only one variable

rm(control\_sample)

rm(indicesNotTraining) # to remove only one variable

rm(balanced\_data)

rm(indicesTest)

rm(indicesTraining)

rm(sampleSizeTest)

rm(sampleSizeTraining)

rm(data\_sg3)

memory.size(max=F) # memory usage

gc()

## train a causal forest - sample test##

# test with example first #

n <- 500 # number of rows

p <- 10 # number of covariates

X <- matrix(rnorm(n\*p), n, p) # the covariates used in the causal regression

W <- rbinom(n, 1, 0.5) # the treatment assignment (must be a binary or real numeric vector with no NAs)

Y <- pmax(X[,1], 0) \* W + X[,2] + pmin(X[,3], 0) + rnorm(n) # outcome variable (must be a numeric vector with no NAs)

c.forest <- causal\_forest(X,Y,W) # 이거부터 얼마나 걸리는지 확인

## train a causal forest ##

# actual data - 1. any\_cm #

X <- dfTraining[, c("SEX\_TYPE", "mother\_age", 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght',

'sudogwon', 'insurance\_type',

'HTN\_med', 'DM\_med', 'HLD\_med', 'Diuretic\_med', 'Antidepressant\_med',

'Antithrombotic\_med', 'Dialysis\_related\_med', 'Steroid\_med')]

sapply(X, class)

X <- data.matrix(X)

class(X)

W <- dfTraining[, c("renal\_disease")]

W <- data.matrix(W)

Y <- dfTraining[, c("any\_cm")]

Y <- data.matrix(Y)

rm(dfTraining)

rm(dfTest)

rm(data)

memory.size(max=F) # memory usage

gc()

c.forest <- causal\_forest(X,Y,W)

saveRDS(c.forest, file="Z:/zio/R/causal\_forest\_anyCM.rds")

rm(c.forest) # to remove only one variable

# actual data - 2. any\_neruo #

X <- dfTraining[, c("SEX\_TYPE", "mother\_age", 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght',

'sudogwon', 'insurance\_type',

'HTN\_med', 'DM\_med', 'HLD\_med', 'Diuretic\_med', 'Antidepressant\_med',

'Antithrombotic\_med', 'Dialysis\_related\_med', 'Steroid\_med')]

sapply(X, class)

X <- data.matrix(X)

class(X)

W <- dfTraining[, c("renal\_disease")]

W <- data.matrix(W)

Y <- dfTraining[, c("any\_neruo")]

Y <- data.matrix(Y)

rm(dfTraining)

rm(dfTest)

rm(data)

memory.size(max=F) # memory usage

gc()

c.forest <- causal\_forest(X,Y,W)

saveRDS(c.forest, file="Z:/zio/R/causal\_forest\_anyNEURO\_control20per.rds")

# read trained model

model <- readRDS("Z:/zio/R/models/causal\_forest\_anyCM\_control20per.rds")

model <- readRDS("Z:/zio/R/models/causal\_forest\_anyNEURO\_control20per.rds")

model <- readRDS("Z:/zio/R/causal\_forest\_anyCM\_testreduced.rds")

# predict using the model - 1. CM

X.test <- dfTest[, c("SEX\_TYPE", "mother\_age", 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght',

'sudogwon', 'insurance\_type',

'HTN\_med', 'DM\_med', 'HLD\_med', 'Diuretic\_med', 'Antidepressant\_med',

'Antithrombotic\_med', 'Dialysis\_related\_med', 'Steroid\_med')]

X.test <- data.matrix(X.test)

W.test <- dfTest[, c("renal\_disease")]

W.test <- data.matrix(W.test)

Y.test <- dfTest[, c("any\_cm")]

Y.test <- data.matrix(Y.test)

# predict using the model - 2. NEURO

X.test <- dfTest[, c("SEX\_TYPE", "mother\_age", 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght',

'sudogwon', 'insurance\_type',

'HTN\_med', 'DM\_med', 'HLD\_med', 'Diuretic\_med', 'Antidepressant\_med',

'Antithrombotic\_med', 'Dialysis\_related\_med', 'Steroid\_med')]

X.test <- data.matrix(X.test)

W.test <- dfTest[, c("renal\_disease")]

W.test <- data.matrix(W.test)

Y.test <- dfTest[, c("any\_neruo")]

Y.test <- data.matrix(Y.test)

# predict on out-of-bag training samples

c.pred <- predict(model)

# compute doubly robust ATE estimate

ate <- average\_treatment\_effect(model) # average\_treatment\_effect function is only used for binary treatment

ate

# check variable importance

varimp <- variable\_importance(model)

ranked.vars <- order(varimp, decreasing = TRUE)

colnames(X)[ranked.vars[1:5]]

# check linear association measure for the CATEs

best\_linear\_projection(model, X[,ranked.vars[1:5]])

# evaluation with RATEs and area under TOC curve

tau.hat.eval <- predict(model, X.test)$predictions

eval.forest <- causal\_forest(X.test, Y.test, W.test) # takes long

rate.cate <- rank\_average\_treatment\_effect(eval.forest, tau.hat.eval) # 이제 잡힘.

rate.cate

plot(rate.cate, main="TOC: By decreasing estimated CATE")

save.image(file="Z:/zio/R/causal\_forest\_data\_anyCM\_modeling.RData")

save.image(file="Z:/zio/R/causal\_forest\_data\_anyNEURO\_modeling.RData")

# XAI

## already downloaded ##

install.packages("Z:/zio/R/downloaded\_packages/DALEX\_2.4.3.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/iml\_0.11.3.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/MetricsWeighted\_1.0.3.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/ranger\_0.16.0.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/flashlight\_0.9.0.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/xgboost\_0.90.0.2.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/data.table\_1.12.8.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/magrittr\_1.5.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/stringi\_1.4.3.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/Matrix\_1.2-18.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/MLmetrics\_1.1.1.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/bitops\_1.0-6.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/gtools\_3.8.1.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/gdata\_2.18.0.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/caTools\_1.17.1.3.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/gplots\_3.0.1.1.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/ROCR\_1.0-7.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/pROC\_1.15.3.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/plyr\_1.8.5.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/Rcpp\_1.0.3.zip",repos=NULL,type="source")

install.packages("Z:/zio/R/downloaded\_packages/gbm\_2.1.5.zip",repos=NULL,type="source")

## install in NHIS CRAN ##

# install.packages("xgboost",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

# install.packages("Matrix",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

# install.packages("MLmetrics",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

# install.packages("pROC",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

install.packages("DiagrammeR",repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary") # dependency가 너무 많음. 따로 설치.

install.packages("caret",dependencies = TRUE, repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary") # dependency가 너무 많음. 따로 설치.

# install.packages("gbm", repos="http://rcran.nhisin.or.kr:8080/cran",type="win.binary")

# 통계 분석용 데이터

data <- read.csv("Z:/zio/R/final\_data\_for\_analysis\_sg3\_bmi\_ageRefchanged.csv") # 784547 268

# 전체 362만 아기로 재정리한 데이터 (prediction 용)

data <- read.csv("Z:/zio/R/prediction\_dta.csv") # 3623766 263

memory.size(max=T) # memory size of system

memory.size(max=F) # memory usage

rm(list=ls()) # to remove all variable

rm(test\_cm) # to remove only one variable

gc() # memory clear

library(DALEX)

library(ranger)

library(flashlight)

library(xgboost)

library(Matrix)

library(MLmetrics)

library(pROC)

library(DiagrammeR)

library(caret)

library(gbm)

## xgboost model ##

dim(data)

sum(is.na(data$any\_cm))

table(data$any\_cm)

table(data$any\_neruo)

head(data)

# check NAs

colSums(is.na(data))

# factor categorical variables - 약제 변수들 추가 필요

data$study\_group <- factor(data$study\_group)

data$SEX\_TYPE <- factor(data$SEX\_TYPE)

data$SEX\_TYPE <- relevel(data$SEX\_TYPE, ref=2) # fetal sex reference를 female로 변경

data$smoke <- factor(data$smoke)

data$HTN <- factor(data$HTN)

data$DM <- factor(data$DM)

data$DSPDA <- factor(data$DSPDA)

data$MI <- factor(data$MI)

data$CHF <- factor(data$CHF)

data$CDO <- factor(data$CDO)

data$TIA <- factor(data$TIA)

data$CLD <- factor(data$CLD)

data$COPD <- factor(data$COPD)

data$Malignancy <- factor(data$Malignancy)

data$GDM <- factor(data$GDM)

data$PE <- factor(data$PE)

data$CE <- factor(data$CE)

data$PA <- factor(data$PA)

data$PP <- factor(data$PP)

data$NP <- factor(data$NP)

data$LGA <- factor(data$LGA)

data$pb\_real <- factor(data$pb\_real)

data$low\_bth\_wght <- factor(data$low\_bth\_wght)

data$bmi\_category <- factor(data$bmi\_category)

data$sudogwon <- factor(data$sudogwon)

data$income\_level <- factor(data$income\_level)

data$insurance\_type <- factor(data$insurance\_type)

data$any\_cm <- factor(data$any\_cm)

data$any\_neruo <- factor(data$any\_neruo)

data$APO <- factor(data$APO)

## final age group ## 이걸로 실행 ##

data$age\_group <- factor(data$age\_group)

data$age\_group <- relevel(data$age\_group, ref=2) # age\_group reference를 2로 지정 (25-29세)

# copy original data

dt\_xgboost <- data

# memory save

rm(data)

# study group null 값을 0으로

dt\_xgboost[which(is.na(dt\_xgboost$study\_group) == TRUE), 'study\_group'] <- 0

table(dt\_xgboost$study\_group)

# (신장질환군으로만 작업할때) select only patients (study group 1, 2, 3)

dt\_xgboost <- subset(dt\_xgboost, study\_group == 1 | study\_group == 2 | study\_group == 3)

table(dt\_xgboost$study\_group)

# 약제 변수들 추가 필요

# select variables to use only - 검진변수 미포함

# by 2 targets (any\_cm, any\_neruo)

# nicu\_1y = 1: 분만 이후 1년 이내 nicu admission 인 아기 -> 제외!

dt\_xgboost\_cm <- dt\_xgboost[, c("any\_cm", "SEX\_TYPE", "study\_group",

"age\_group",

'smoke', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght',

'sudogwon',

'HTN\_med', 'DM\_med', 'HLD\_med', 'Diuretic\_med', 'Antidepressant\_med',

'Antithrombotic\_med', 'Dialysis\_related\_med', 'Steroid\_med')]

dt\_xgboost\_neruo <- dt\_xgboost[, c("any\_neruo", "SEX\_TYPE", "study\_group",

"age\_group",

'smoke', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght',

'sudogwon',

'HTN\_med', 'DM\_med', 'HLD\_med', 'Diuretic\_med', 'Antidepressant\_med',

'Antithrombotic\_med', 'Dialysis\_related\_med', 'Steroid\_med')]

dt\_xgboost\_apo <- dt\_xgboost[, c("APO", "SEX\_TYPE", "study\_group",

"age\_group",

'smoke', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'CE', 'PP',

'NP', 'LGA',

'sudogwon',

'HTN\_med', 'DM\_med', 'HLD\_med', 'Diuretic\_med', 'Antidepressant\_med',

'Antithrombotic\_med', 'Dialysis\_related\_med', 'Steroid\_med')]

# memory save

rm(dt\_xgboost)

# select variables to use only - 검진변수 포함 (측정값), N수 많이 적어짐

# by 2 targets (any\_cm, any\_neruo)

# nicu\_1y = 1: 분만 이후 1년 이내 nicu admission 인 아기 -> 제외!

dt\_xgboost\_cm <- dt\_xgboost[, c("any\_cm", "SEX\_TYPE", "study\_group",

"age\_group",

'smoke', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght',

'sudogwon',

'HTN\_med', 'DM\_med', 'HLD\_med', 'Diuretic\_med', 'Antidepressant\_med',

'Antithrombotic\_med', 'Dialysis\_related\_med', 'Steroid\_med',

'income\_level', 'insurance\_type', 'G1E\_BMI', "G1E\_VA\_LT", "G1E\_VA\_RT",

'G1E\_HA\_LT', 'G1E\_HA\_RT', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA', 'G1E\_URN\_PROT', 'G1E\_HGB',

'G1E\_FBS', 'G1E\_TOT\_CHOL', 'G1E\_SGOT', 'G1E\_SGPT', 'G1E\_GGT', 'G1E\_HB\_DRK',

'G1E\_HB\_PA', 'G1E\_TRM\_SEQ', 'G1E\_WSTC', 'G1E\_TG', 'G1E\_HDL', 'G1E\_LDL', 'G1E\_CRTN',

'Q\_HBV\_AG', 'Q\_PA\_VD', 'Q\_PA\_MD', 'Q\_PA\_WALK')]

dt\_xgboost\_neruo <- dt\_xgboost[, c("any\_neruo", "SEX\_TYPE", "study\_group",

"age\_group",

'smoke', 'HTN', 'DM', 'DSPDA', 'MI', 'CHF', 'CDO',

'TIA', 'CLD', 'COPD', 'Malignancy', 'GDM', 'PE', 'CE', 'PA', 'PP',

'NP', 'LGA', 'pb\_real', 'low\_bth\_wght',

'sudogwon',

'HTN\_med', 'DM\_med', 'HLD\_med', 'Diuretic\_med', 'Antidepressant\_med',

'Antithrombotic\_med', 'Dialysis\_related\_med', 'Steroid\_med',

'income\_level', 'insurance\_type', 'G1E\_BMI', "G1E\_VA\_LT", "G1E\_VA\_RT",

'G1E\_HA\_LT', 'G1E\_HA\_RT', 'G1E\_BP\_SYS', 'G1E\_BP\_DIA', 'G1E\_URN\_PROT', 'G1E\_HGB',

'G1E\_FBS', 'G1E\_TOT\_CHOL', 'G1E\_SGOT', 'G1E\_SGPT', 'G1E\_GGT', 'G1E\_HB\_DRK',

'G1E\_HB\_PA', 'G1E\_TRM\_SEQ', 'G1E\_WSTC', 'G1E\_TG', 'G1E\_HDL', 'G1E\_LDL', 'G1E\_CRTN',

'Q\_HBV\_AG', 'Q\_PA\_VD', 'Q\_PA\_MD', 'Q\_PA\_WALK')]

colSums(is.na(dt\_xgboost\_cm))

# drop all NAs

dt\_xgboost\_cm <- na.omit(dt\_xgboost\_cm) # 653331 55

dt\_xgboost\_neruo <- na.omit(dt\_xgboost\_neruo) # 653331 55

dt\_xgboost\_apo <- na.omit(dt\_xgboost\_apo) # 3623766 20

## 중간 explainable models test ##

sessionInfo(package='DALEX')

data("titanic\_imputed", package="DALEX")

ranger\_model <- ranger::ranger(any\_neruo ~., data=dt\_xgboost\_neruo,

classification=TRUE, probability=TRUE)

flashlight\_predict <- function(X.model, new\_data)

predict(X.model, new\_data)$predictions[,2]

iml\_predict <- function(X.model, newdata)

predict(X.model, newdata)$predictions[,2]

exp\_dalex <- explain(ranger\_model,

data=dt\_xgboost\_neruo,

y=dt\_xgboost\_neruo$any\_neruo,

label="Ranger Model")

fi\_dalex <- model\_parts(exp\_dalex, B=10, loss\_function=loss\_one\_minus\_auc)

plot(fi\_dalex)

#### test end. ####

## xgboost 다시 ##

# create sparse matrix

dt\_xgboost\_cm\_spmatrix <- sparse.model.matrix(any\_cm~.-1, data=dt\_xgboost\_cm)

dt\_xgboost\_neruo\_spmatrix <- sparse.model.matrix(any\_neruo~.-1, data=dt\_xgboost\_neruo)

dt\_xgboost\_apo\_spmatrix <- sparse.model.matrix(APO~.-1, data=dt\_xgboost\_apo)

dim(dt\_xgboost\_cm\_spmatrix)

# define train dataset sampling index (70%)

train\_index\_cm <- sample(1:nrow(dt\_xgboost\_cm\_spmatrix), nrow(dt\_xgboost\_cm\_spmatrix)\*0.7)

train\_index\_neruo <- sample(1:nrow(dt\_xgboost\_neruo\_spmatrix), nrow(dt\_xgboost\_neruo\_spmatrix)\*0.7)

train\_index\_apo <- sample(1:nrow(dt\_xgboost\_apo\_spmatrix), nrow(dt\_xgboost\_apo\_spmatrix)\*0.7)

# create train/test dataset, label data

#cm

train\_cm\_x <- dt\_xgboost\_cm\_spmatrix[train\_index\_cm, ]

test\_cm\_x <- dt\_xgboost\_cm\_spmatrix[-train\_index\_cm, ]

train\_cm\_y <- dt\_xgboost\_cm[train\_index\_cm, 'any\_cm']

test\_cm\_y <- dt\_xgboost\_cm[-train\_index\_cm, 'any\_cm']

#neruo

train\_neruo\_x <- dt\_xgboost\_neruo\_spmatrix[train\_index\_neruo, ]

test\_neruo\_x <- dt\_xgboost\_neruo\_spmatrix[-train\_index\_neruo, ]

train\_neruo\_y <- dt\_xgboost\_neruo[train\_index\_neruo, 'any\_neruo']

test\_neruo\_y <- dt\_xgboost\_neruo[-train\_index\_neruo, 'any\_neruo']

#APO

train\_apo\_x <- dt\_xgboost\_apo\_spmatrix[train\_index\_apo, ]

test\_apo\_x <- dt\_xgboost\_apo\_spmatrix[-train\_index\_apo, ]

train\_apo\_y <- dt\_xgboost\_apo[train\_index\_apo, 'APO']

test\_apo\_y <- dt\_xgboost\_apo[-train\_index\_apo, 'APO']

# check train&test

dim(train\_cm\_x)

dim(test\_cm\_x)

dim(test\_cm\_y)

# change data type to use xgboost algorithm

dtrain\_cm <- xgb.DMatrix(data=train\_cm\_x, label=as.matrix(train\_cm\_y))

dtest\_cm <- xgb.DMatrix(data=test\_cm\_x, label=as.matrix(test\_cm\_y))

dtrain\_neruo <- xgb.DMatrix(data=train\_neruo\_x, label=as.matrix(train\_neruo\_y))

dtest\_neruo <- xgb.DMatrix(data=test\_neruo\_x, label=as.matrix(test\_neruo\_y))

dtrain\_apo <- xgb.DMatrix(data=train\_apo\_x, label=as.matrix(train\_apo\_y))

dtest\_apo <- xgb.DMatrix(data=test\_apo\_x, label=as.matrix(test\_apo\_y))

## tune and run the model

# xgboost

# parameter setting

param <- list(max\_depth = 6, eta = 0.1, verbose = 0, nthread = 2, objective = "binary:logistic", eval\_metric = "auc")

watchlist\_cm <- list(train=dtrain\_cm, test=dtest\_cm)

watchlist\_neruo <- list(train=dtrain\_neruo, test=dtest\_neruo)

watchlist\_apo <- list(train=dtrain\_apo, test=dtest\_apo)

## xgboost 5 fold cross validation ##

set.seed(200)

xgb\_cv <- xgb.cv(data=dtrain\_neruo, label=train\_neruo\_y, nfold=5, nrounds=50, early\_stopping\_rounds = 10,

objective='binary:logistic', metrics=list('rmse', 'auc'), num\_class=2, prediction=T, print\_every\_n=10,

params=list(eta=0.05, max\_depth=8, subsample=0.8, colsample\_bytree=0.8, stratified=T))

# xgboost algorithm

xgb\_cm <- xgb.train(params = param,

data=dtrain\_cm,

nrounds = 1000, # 수정 필요

subsample = 0.5,

colsample\_bytree = 0.5,

num\_class = 1,

watchlist=watchlist\_cm)

xgb\_neruo <- xgb.train(params = param,

data=dtrain\_neruo,

nrounds = 1000, # 수정 필요

subsample = 0.5,

colsample\_bytree = 0.5,

num\_class = 1,

watchlist = watchlist\_neruo)

xgb\_apo <- xgb.train(params = param,

data=dtrain\_apo,

nrounds = 1000, # 수정 필요

subsample = 0.5,

colsample\_bytree = 0.5,

num\_class = 1,

watchlist = watchlist\_apo)

# predict

# cm

train\_y\_pred\_cm <- predict(xgb\_cm, dtrain\_cm)

test\_y\_pred\_cm <- predict(xgb\_cm, dtest\_cm)

# neruo

train\_y\_pred\_neruo <- predict(xgb\_neruo, dtrain\_neruo)

test\_y\_pred\_neruo <- predict(xgb\_neruo, dtest\_neruo)

# apo

train\_y\_pred\_apo <- predict(xgb\_apo, dtrain\_apo)

test\_y\_pred\_apo <- predict(xgb\_apo, dtest\_apo)

# transform prediction in a binary classifiction

# threshold를 0.5 -> 0.3으로 바꿔보자

train\_y\_pred\_cm <- as.numeric(train\_y\_pred\_cm > 0.5)

test\_y\_pred\_cm <- as.numeric(test\_y\_pred\_cm > 0.5)

train\_y\_pred\_neruo <- as.numeric(train\_y\_pred\_neruo > 0.3)

test\_y\_pred\_neruo <- as.numeric(test\_y\_pred\_neruo > 0.3)

train\_y\_pred\_apo <- as.numeric(train\_y\_pred\_apo > 0.5)

test\_y\_pred\_apo <- as.numeric(test\_y\_pred\_apo > 0.5)

table(train\_y\_pred\_cm)

table(train\_y\_pred\_neruo)

table(train\_y\_pred\_apo)

# confusion matrix

# cm train

train\_y\_pred\_cm <- as.factor(train\_y\_pred\_cm)

train\_cm\_y <- as.factor(train\_cm\_y)

confusionMatrix(train\_y\_pred\_cm, train\_cm\_y)

# cm test

test\_y\_pred\_cm <- as.factor(test\_y\_pred\_cm)

test\_cm\_y <- as.factor(test\_cm\_y)

confusionMatrix(test\_y\_pred\_cm, test\_cm\_y)

# neuro train

train\_y\_pred\_neruo <- as.factor(train\_y\_pred\_neruo)

train\_neruo\_y <- as.factor(train\_neruo\_y)

confusionMatrix(train\_y\_pred\_neruo, train\_neruo\_y)

# neuro test

test\_y\_pred\_neruo <- as.factor(test\_y\_pred\_neruo)

test\_neruo\_y <- as.factor(test\_neruo\_y)

confusionMatrix(test\_y\_pred\_neruo, test\_neruo\_y)

# APO train

train\_y\_pred\_apo <- as.factor(train\_y\_pred\_apo)

train\_apo\_y <- as.factor(train\_apo\_y)

confusionMatrix(train\_y\_pred\_apo, train\_apo\_y)

# APO test

test\_y\_pred\_apo <- as.factor(test\_y\_pred\_apo)

test\_apo\_y <- as.factor(test\_apo\_y)

confusionMatrix(test\_y\_pred\_apo, test\_apo\_y)

# roc plot

library(ROCR)

pred <- prediction(c(train\_y\_pred\_cm), c(train\_cm\_y))

plot(performance(pred, "tpr", "fpr"))

# importance matrix

importance\_matrix\_cm <- xgb.importance(model=xgb\_cm)

xgb.plot.importance(importance\_matrix = importance\_matrix\_cm)

importance\_matrix\_neruo <- xgb.importance(model=xgb\_neruo)

xgb.plot.importance(importance\_matrix = importance\_matrix\_neruo)

importance\_matrix\_apo <- xgb.importance(model=xgb\_apo)

xgb.plot.importance(importance\_matrix = importance\_matrix\_apo)

# # dump tree and visualize

# xgb.dump(xgb\_cm, with\_stats = TRUE)

# xgb.plot.tree(model=xgb\_cm)

####### GBM ############

# define train dataset sampling index (70%)

train\_index\_cm <- sample(1:nrow(dt\_xgboost\_cm), nrow(dt\_xgboost\_cm)\*0.7)

train\_index\_neruo <- sample(1:nrow(dt\_xgboost\_neruo), nrow(dt\_xgboost\_neruo)\*0.7)

train\_index\_apo <- sample(1:nrow(dt\_xgboost\_apo), nrow(dt\_xgboost\_apo)\*0.7)

# create train/test dataset, label data

#cm

train\_cm <- dt\_xgboost\_cm[train\_index\_cm, ]

test\_cm <- dt\_xgboost\_cm[-train\_index\_cm, ]

#neruo

train\_neruo <- dt\_xgboost\_neruo[train\_index\_neruo, ]

test\_neruo <- dt\_xgboost\_neruo[-train\_index\_neruo, ]

#apo

train\_apo <- dt\_xgboost\_apo[train\_index\_apo, ]

test\_apo <- dt\_xgboost\_apo[-train\_index\_apo, ]

rm(dt\_xgboost\_cm\_spmatrix)

rm(test\_cm\_x)

rm(train\_cm\_x)

rm(train\_cm\_y)

rm(test\_cm\_y)

rm(dtest\_cm)

rm(dtrain\_cm)

rm(gbm\_cm)

set.seed(123)

# cm

gbm\_cm <- gbm(

formula = any\_cm ~ .,

distribution = 'bernoulli',

data = train\_cm,

var.monotone = NULL,

n.trees = 100,

interaction.depth = 3,

n.minobsinnode = 10,

shrinkage = 0.1,

bag.fraction = 0.5,

train.fraction = 0.7,

cv.folds = 2,

keep.data = TRUE,

verbose = TRUE,

n.cores = NULL

)

summary(

gbm\_cm,

# cbars = 10,

method = relative.influence,

las = 2

)

best.iter <- gbm.perf(gbm\_cm, plot.it = T, oobag.curve = T, method = 'OOB')

print(best.iter)

# find index for n trees with minimum CV error

min\_MSE <- which.min(gbm\_cm$cv.error)

# get MSE and compute RMSE

sqrt(gbm\_cm$cv.error[min\_MSE])

# construct univariate partial dependence plots

plot(gbm\_cm, i.var = 1, n.trees=best.iter)

# construct bivariate partial dependence plots

plot(gbm\_cm, i.var = 1:2, n.trees=best.iter)

# predict with test data

yhat <- predict(gbm\_cm, newdata=test\_cm, n.trees=best.iter, type='link')

# lest squares error

print(sum((test\_cm$any\_cm - yhat)^2))

# roc

gbm.roc.area(test\_cm$any\_cm, yhat)

# neruo

gbm\_neruo <- gbm(

formula = any\_neruo ~ .,

distribution = 'bernoulli',

data = train\_neruo,

var.monotone = NULL,

n.trees = 100,

interaction.depth = 3,

n.minobsinnode = 10,

shrinkage = 0.1,

bag.fraction = 0.5,

train.fraction = 0.7,

cv.folds = 2,

keep.data = TRUE,

verbose = TRUE,

n.cores = NULL

)

summary(

gbm\_neruo,

cbars = 10,

method = relative.influence,

las = 2

)

best.iter <- gbm.perf(gbm\_neruo, plot.it = T, oobag.curve = T, method = 'OOB')

print(best.iter)

# find index for n trees with minimum CV error

min\_MSE <- which.min(gbm\_neruo$cv.error)

# get MSE and compute RMSE

sqrt(gbm\_neruo$cv.error[min\_MSE])

# construct univariate partial dependence plots

plot(gbm\_neruo, i.var = 1, n.trees=best.iter)

# construct bivariate partial dependence plots

plot(gbm\_neruo, i.var = 1:2, n.trees=best.iter)

# predict with test data

yhat <- predict(gbm\_neruo, newdata=test\_neruo, n.trees=best.iter, type='link')

# lest squares error

print(sum((test\_neruo$any\_neruo - yhat)^2))

# roc

gbm.roc.area(test\_neruo$any\_neruo, yhat)

# apo

gbm\_apo <- gbm(

formula = APO ~ .,

distribution = 'bernoulli',

data = train\_apo,

var.monotone = NULL,

n.trees = 100,

interaction.depth = 3,

n.minobsinnode = 10,

shrinkage = 0.1,

bag.fraction = 0.5,

train.fraction = 0.7,

cv.folds = 2,

keep.data = TRUE,

verbose = TRUE,

n.cores = NULL

)

summary(

gbm\_apo,

# cbars = 10,

method = relative.influence,

las = 2

)

best.iter <- gbm.perf(gbm\_apo, plot.it = T, oobag.curve = T, method = 'OOB')

print(best.iter)

# find index for n trees with minimum CV error

min\_MSE <- which.min(gbm\_apo$cv.error)

# get MSE and compute RMSE

sqrt(gbm\_apo$cv.error[min\_MSE])

# construct univariate partial dependence plots

plot(gbm\_apo, i.var = 1, n.trees=best.iter)

# construct bivariate partial dependence plots

plot(gbm\_apo, i.var = 1:2, n.trees=best.iter)

# predict with test data

yhat <- predict(gbm\_apo, newdata=test\_apo, n.trees=best.iter, type='link')

# lest squares error

print(sum((test\_apo$APO - yhat)^2))

# roc

gbm.roc.area(test\_apo$APO, yhat)