Developing an artificial intelligence-guided signal detection in the Food and Drug Administration Adverse Event Reporting System (FAERS): A proof-of-concept study using galcanezumab and simulated data.

Fahed Al-Azzawi[1], Israa Mahmoud[1], François Haguinet[2], Andrew Bate[3,4,5], Maurizio Sessa[1]

[1] Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark

[2] GSK, Wavre, Belgium

[3] GSK, London, UK

[4] London School of Hygiene and Tropical Medicine, University of London, London, UK

[5] New York University, New York, NY, USA

Corresponding author

Fahed Al-Azzawi, MPharm
Department of Drug Design and Pharmacology,
University of Copenhagen,
Jagtvej 160Copenhagen 2100, Denmark
Email: kdz640@sund.ku.dk

Electronic supplementary material 1

Version: 2023-01-23

```r
################################################################################
# Simulated data to validate the framework                                     #
################################################################################
#Libraries

library(MASS)
library(tidyverse)
library(GGally)
library(bindata)
library(partykit)
library(Publish)
library(openxlsx)

####################
########SCENARIO 1
####################

#Set seed for reproducibility
set.seed(5)

# create the variance covariance matrix
sigma<-rbind(c(1,-0.8,0.7), c(-0.8,1, 0.9), c(0.7,-0.9,1))

# create the mean vector
mu<-c(40, 75, 2)

# generate the multivariate normal distribution
df<-as.data.frame(mvrnorm(n=1000, mu=mu, Sigma=sigma))

df<-df%>%mutate(MyBinary = ifelse(V1>median(V1), 0 ,1))

df<-df%>%mutate(MyNoisyBinary = ifelse(V1>median(V1), sample(c(0,1),n(), replace = TRUE, p=c(0.25, 0.75)) ,
                        sample(c(0,1),n(), replace = TRUE, p=c(0.75, 0.25))))

df<-df%>%mutate(MyBinary2 = ifelse(V2>median(V2), 1 ,0))

# Construct a binary correlation matrix
rho <- 0
m <- matrix(c(1,rho,rho,1), ncol=2)

# Simulate x-y pairs, and check that they have the specified correlation structure
x <- rmvbin(1000, margprob = c(0.5, 0.5), bincorr = m)
df2 <- as.data.frame(x)
df$V4 <- df2$V1
colnames(df) <- c("Age", "Weight", "N_co_reported_drugs", "Co_reported_drug_1", "Co_reported_drug_2",
"Outcome", "Co_reported_drug_3")
df$Caseid <- 1:1000

#Correlation matrix
ggpairs(df)
```

```r
#Data management
df$Outcome <- factor(df$Outcome,levels=c("0","1"),labels=c("Control","Galcanezumab"))
df$Co_reported_drug_1 <- factor(df$Co_reported_drug_1,levels=c("0","1"),labels=c("No","Yes"))
df$Co_reported_drug_2 <- factor(df$Co_reported_drug_2,levels=c("0","1"),labels=c("No","Yes"))
df$Co_reported_drug_3 <- factor(df$Co_reported_drug_3,levels=c("0","1"),labels=c("No","Yes"))

colnames(df)

## [1] "Age"                "Weight"             "N_co_reported_drugs"
## [4] "Co_reported_drug_1" "Co_reported_drug_2" "Outcome"
## [7] "Co_reported_drug_3" "Caseid"

df$N_co_reported_drugs <- round(df$N_co_reported_drugs)
df$Age <- round(df$Age)
df$Weight <- round(df$Weight)

#Descriptive table
table1 <- summary(univariateTable(Outcome ~ Age +
Weight+N_co_reported_drugs+Co_reported_drug_1+Co_reported_drug_2+Co_reported_drug_3, data=df))

#Conditional inference tree
tree <- partykit::ctree(Outcome ~ Co_reported_drug_1+Co_reported_drug_2+Co_reported_drug_3, data=df)
plot(tree)

####################
########SCENARIO 2
###################

#Set seed for reproducibility
set.seed(5)

# create the variance covariance matrix
sigma<-rbind(c(1,-0.8,0.7), c(-0.8,1, 0.9), c(0.7,-0.9,1))

# create the mean vector
mu<-c(40, 75, 2)

# generate the multivariate normal distribution
df<-as.data.frame(mvrnorm(n=50, mu=mu, Sigma=sigma))

df<-df%>%mutate(MyBinary = ifelse(V1>median(V1), 0 ,1))

df<-df%>%mutate(MyNoisyBinary = ifelse(V1>median(V1), sample(c(0,1),n(), replace = TRUE, p=c(0.25, 0.75)) ,
                 sample(c(0,1),n(), replace = TRUE, p=c(0.75, 0.25))))

df<-df%>%mutate(MyBinary2 = ifelse(V2>median(V2), 1 ,0))


# Construct a binary correlation matrix
rho <- 0
m <- matrix(c(1,rho,rho,1), ncol=2)

# Simulate x-y pairs, and check that they have the specified correlation structure
x <- rmvbin(50, margprob = c(0.5, 0.5), bincorr = m)
df2 <- as.data.frame(x)
df$V4 <- df2$V1
```

```r
colnames(df) <- c("Age", "Weight", "N_co_reported_drugs", "Co_reported_drug_1", "Co_reported_drug_2",
"Outcome", "Co_reported_drug_3")
df$Caseid <- 1:50

#Correlation plot
ggpairs(df)

df$Outcome <- factor(df$Outcome,levels=c("0","1"),labels=c("Control","Galcanezumab"))
df$Co_reported_drug_1 <- factor(df$Co_reported_drug_1,levels=c("0","1"),labels=c("No","Yes"))
df$Co_reported_drug_2 <- factor(df$Co_reported_drug_2,levels=c("0","1"),labels=c("No","Yes"))
df$Co_reported_drug_3 <- factor(df$Co_reported_drug_3,levels=c("0","1"),labels=c("No","Yes"))

colnames(df)

## [1] "Age"              "Weight"            "N_co_reported_drugs"
## [4] "Co_reported_drug_1"  "Co_reported_drug_2"  "Outcome"
## [7] "Co_reported_drug_3"  "Caseid"

df$N_co_reported_drugs <- round(df$N_co_reported_drugs)
df$Age <- round(df$Age)
df$Weight <- round(df$Weight)

#Descriptive table
table1 <- summary(univariateTable(Outcome ~ Age +
Weight+N_co_reported_drugs+Co_reported_drug_1+Co_reported_drug_2+Co_reported_drug_3, data=df))

#Conditional inference tree
tree <- partykit::ctree(Outcome ~ Co_reported_drug_1+Co_reported_drug_2+Co_reported_drug_3, data=df)
plot(tree)

######################
#########SCENARIO 3
###################

#Set seed for reproducibility
set.seed(5)

# create the variance covariance matrix
sigma<-rbind(c(1,-0.8,0.7), c(-0.8,1, 0.9), c(0.7,-0.9,1))

# create the mean vector
mu<-c(40, 75, 2)

# generate the multivariate normal distribution
df<-as.data.frame(mvrnorm(n=1000, mu=mu, Sigma=sigma))

df<-df%>%mutate(MyBinary = ifelse(V1>median(V1), 0 ,1))

df<-df%>%mutate(MyNoisyBinary = ifelse(V1>median(V1), sample(c(0,1),n(), replace = TRUE, p=c(0.25, 0.75)) ,
                      sample(c(0,1),n(), replace = TRUE, p=c(0.55, 0.45))))

df<-df%>%mutate(MyBinary2 = ifelse(V2>median(V2), 1 ,0))


# Construct a binary correlation matrix
rho <- 0
```

```r
m <- matrix(c(1,rho,rho,1), ncol=2)

# Simulate x-y pairs, and check that they have the specified correlation structure
x <- rmvbin(1000, margprob = c(0.5, 0.5), bincorr = m)
df2 <- as.data.frame(x)

df$V4 <- df2$V1
colnames(df) <- c("Age", "Weight", "N_co_reported_drugs", "Co_reported_drug_1", "Co_reported_drug_2",
"Outcome", "Co_reported_drug_3")
df$Caseid <- 1:1000

ggpairs(df)

df$Outcome <- factor(df$Outcome,levels=c("0","1"),labels=c("Control","Galcanezumab"))
df$Co_reported_drug_1 <- factor(df$Co_reported_drug_1,levels=c("0","1"),labels=c("No","Yes"))
df$Co_reported_drug_2 <- factor(df$Co_reported_drug_2,levels=c("0","1"),labels=c("No","Yes"))
df$Co_reported_drug_3 <- factor(df$Co_reported_drug_3,levels=c("0","1"),labels=c("No","Yes"))

colnames(df)

## [1] "Age"                "Weight"             "N_co_reported_drugs"
## [4] "Co_reported_drug_1" "Co_reported_drug_2" "Outcome"
## [7] "Co_reported_drug_3" "Caseid"

df$N_co_reported_drugs <- round(df$N_co_reported_drugs)
df$Age <- round(df$Age)
df$Weight <- round(df$Weight)

df <- df[,-4]

#Descriptive table
table1 <- summary(univariateTable(Outcome ~ Age +
Weight+N_co_reported_drugs+Co_reported_drug_2+Co_reported_drug_3, data=df))

#Conditional inference tree
tree <- partykit::ctree(Outcome ~ Co_reported_drug_2+Co_reported_drug_3, data=df)
plot(tree)

#################
#    END        #
#################
```