**Appendix B Code used in R Studio to develop and validate the metamodels and create the Bland-Altman plots.**

install.packages("readxl")

install.packages("neuralnet")

install.packages("BlandAltmanLeh")

install.packages("mgcv")

library(readxl)

library(MASS)

library(neuralnet)

library(BlandAltmanLeh)

library(mgcv)

# Define file with the training dataset and the file with the validation dataset

data.train <- read\_excel("Dropbox/MGH/CEPAC metamodeling/data\_tool/Total\_TRAIN.xls")

data.validate <- read\_excel("Dropbox/MGH/CEPAC metamodeling/data\_tool/Total\_VALIDATE.xls")

# OLS metamodel for the EID strategy and the outcome of life expectancy

# Developing the metamodel (the variable numbers refer to the variables listed in Table 1)

EidModelLms <- lm(formula = Lms\_EID ~ (var1 + var2 + var3 + var4 + var5 + var6 + var8 + var9 + var10)^2, data = data.train)

summary(EidModelLms)

# Running the data not used in model development in the newly developed metamodel

EidLmsPredOLS <- predict.lm(EidModelLms, newdata = data.validate)

plot(EidLmsPredOLS, data.validate$Lms\_EID)

summary(lm(EidLmsPredOLS ~ data.validate$Lms\_EID))

# Creating a Bland-Altman plot of the differences

options(max.print = 1000)

bland.altman.plot(EidLmsPredOLS, data.validate$Lms\_EID, main = "Life expectancy for EID", xlab = "Mean life-months ((CEPAC + metamodel) / 2)", ylab = "Difference (CEPAC - metamodel) in life-months")

baplot(EidLmsPredOLS, data.validate$Lms\_EID)

bland.altman.stats(EidLmsPredOLS, data.validate$Lms\_EID)

# OLS metamodel for the EID strategy and the outcome of HIV-related costs

# Developing the metamodel (the variable numbers refer to the variables listed in Table 1)

EidModelCost <- lm(formula = Costs\_EID ~ (var1 + var2 + var3 + var4 + var5 + var6 + var8 + var9 + var10 + var15 + var17 + var18)^2, data = data.train)

summary(EidModelCost)

# Running the data not used in model development in the newly developed metamodel

EidCostPredOLS <- predict.lm(EidModelCost, newdata = data.validate)

plot(EidCostPredOLS, data.validate$Costs\_EID)

summary(lm(EidCostPredOLS ~ data.validate$Costs\_EID))

# Creating a Bland-Altman plot of the differences

options(max.print = 1000)

bland.altman.plot(EidCostPredOLS, data.validate$Costs\_EID, main = "Costs for EID", xlab = "Mean costs in $ ((CEPAC + metamodel) / 2)", ylab = "Difference (CEPAC - metamodel) in $")

baplot(EidCostPredOLS, data.validate$Costs\_EID)

bland.altman.stats(EidCostPredOLS, data.validate$Costs\_EID)

# OLS metamodel for the screen and test strategy and the outcome of life expectancy

# Developing the metamodel (the variable numbers refer to the variables listed in Table 1)

SctModelLms <- lm(formula = Lms\_Sct ~ (var1 + var2 + var3 + var4 + var5 + var6 + var7 + var8 + var9 + var10 + var11 + var12 + var13 + var14)^2, data = data.train)

summary(SctModelLms)

# Running the data not used in model development in the newly developed metamodel

SctPredLms <- predict.lm(SctModelLms, newdata = data.validate)

plot(SctPredLms, data.validate$Lms\_Sct)

summary(lm(SctPredLms ~ data.validate$Lms\_Sct))

# Creating a Bland-Altman plot of the differences

options(max.print = 1000)

bland.altman.plot(SctPredLms, data.validate$Lms\_Sct, main="Life expectancy for screen and test", xlab = "Mean life-months ((CEPAC + metamodel) / 2)", ylab = "Difference (CEPAC - metamodel) in life-months")

baplot(SctPredLms, data.validate$Lms\_Sct)

bland.altman.stats(SctPredLms, data.validate$Lms\_Sct)

# OLS metamodel for the screen and test strategy and the outcome of HIV-related costs

# Developing the metamodel (the variable numbers refer to the variables listed in Table 1)

SctModelCost <- lm(formula = Costs\_Sct ~ (var1 + var2 + var3 + var4 + var5 + var6 + var7 + var8 + var9 + var10 + var11 + var12 + var13 + var14 + var15 + var16 + var17 + var18)^2, data = data.train)

summary(SctModelCost)

# Running the data not used in model development in the newly developed metamodel

SctCostsPred <- predict.lm(SctModelCost, newdata = data.validate)

plot(SctCostsPred, data.validate$Costs\_Sct)

summary(lm(SctCostsPred ~ data.validate$Costs\_Sct))

# Creating a Bland-Altman plot of the differences

options(max.print = 1000)

bland.altman.plot(SctCostsPred, data.validate$Costs\_Sct, main="Costs for screen and test", xlab = "Mean costs in $ ((CEPAC + metamodel) / 2)", ylab = "Difference (CEPAC - metamodel) in $")

baplot(SctCostsPred, data.validate$Costs\_Sct)

bland.altman.stats(SctCostsPred, data.validate$Costs\_Sct)