

```
# =====  
# R CODE  
# small scale simulation study to investigate impact of measurement error  
# measurement error on (continuous) exposure and/or (continuous) confounding variable  
# =====  
# libraries:  
library(Hmisc)  
library(mice)  
library(tidyverse)  
#setwd("")  
# =====  
# set working directory:  
# setwd("")  
  
# =====  
# The data can be downloaded in xpt form from https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2015  
# read data:  
d1 <- sasxport.get("DEMO_I.xpt")  
d2 <- sasxport.get("BPX_I.xpt")  
d3 <- sasxport.get("BMX_I.xpt")  
d4 <- sasxport.get("GHB_I.xpt")  
d5 <- sasxport.get("TCHOL_I.xpt")  
  
d1.t <- subset(d1,select=c("seqn","riagendr","ridageyr"))  
d2.t <- subset(d2,select=c("seqn","bpxsy1"))  
d3.t <- subset(d3,select=c("seqn","bmxbmi"))  
d4.t <- subset(d4,select=c("seqn","lbggh"))  
d5.t <- subset(d5,select=c("seqn","lbdtsi"))  
  
d <- merge(d1.t,d2.t)  
d <- merge(d,d3.t)  
d <- merge(d,d4.t)  
d <- merge(d,d5.t)  
  
# =====  
# rename variables:  
# RIAGENDR - Gender  
# RIDAGEYR - Age in years at screening  
# BPXSY1 - Systolic: Blood pres (1st rdg) mm Hg  
# BMXBMI - Body Mass Index (kg/m**2)  
# LBDTCSI - Total Cholesterol (mmol/L)  
# LBXGH - Glycohemoglobin (%)  
  
d$age <- d$ridageyr  
d$sex <- d$riagendr  
d$bp <- d$bpxsy1  
d$bmi <- d$bmxbmi  
d$HbA1C <- d$lbggh  
d$chol <- d$lbdtsi  
d$age[d$age<18] <- NA  
  
# =====  
# select complete cases:  
dc <- cc(subset(d,select=c("age","sex","bmi","HbA1C","bp")))  
  
# analysis:  
summary(lm(bp ~ HbA1C + age + as.factor(sex), data=dc))  
confint(lm(bp ~ HbA1C + age + as.factor(sex), data=dc))  
  
summary(lm(bp ~ HbA1C + bmi + age + as.factor(sex), data=dc))  
confint(lm(bp ~ HbA1C + bmi + age + as.factor(sex), data=dc))
```

```

# =====
# simulation of measurement error:
ref      <- lm(bp ~ HbA1C + bmi + age + as.factor(sex), data=dc)$coef[2]
n.sim    <- 1e3
perc.me.exp <- seq(0,.5,.1)
perc.me.conf<- seq(0,.5,.1)
scenarios <- expand.grid(perc.me.exp,perc.me.conf)
var.exp  <- var(dc$HbA1C)
var.conf <- var(dc$bmi)
n        <- dim(dc)[1]
beta.hat <- matrix(ncol=dim(scenarios)[1], nrow=n.sim)

for (k in 1:n.sim){
  print(k)
  set.seed(k)
  for (i in 1:dim(scenarios)[1]){
    var.me.exp <- var.exp*scenarios[i,1]/(1-scenarios[i,1])
    var.me.conf <- var.conf*scenarios[i,2]/(1-scenarios[i,2])
    dc$HbA1C.me <- dc$HbA1C + rnorm(dim(dc)[1], 0, sqrt(var.me.exp) )
    dc$bmi.me <- dc$bmi + rnorm(dim(dc)[1], 0, sqrt(var.me.conf) )
    beta.hat[k,i] <- lm(bp ~ HbA1C.me + age + bmi.me + as.factor(sex), data=dc)$coef[2]
  }
}

# =====
# create figure:
tot.mat <- cbind(100*scenarios,apply(beta.hat,2,mean))
colnames(tot.mat) <- c("me.exp","me.conf","estimate")

FIGURE <- ggplot(tot.mat, aes(me.exp, me.conf)) +
  geom_tile(color="white",aes(fill = estimate)) +
  geom_text(aes(label = round(estimate, 2))) +
  scale_fill_gradient2(low="#D55E00",mid="white",high = "#56B4E9", midpoint=ref) +
  labs(x=paste("% of total variance of HbA1c due to measurement error"),
       y=paste("% of total variance of BMI due to measurement error")) +
  coord_equal()+
  scale_y_continuous(breaks=unique(tot.mat[,1]))+
  scale_x_continuous(breaks=unique(tot.mat[,1]))+
  theme(panel.background = element_rect(fill='white', colour='grey'),
        plot.title=element_text(hjust=0),
        axis.ticks=element_blank(),
        axis.title=element_text(size=12),
        axis.text=element_text(size=10),
        legend.title=element_text(size=12),
        legend.text=element_text(size=10))

FIGURE
# savePlot("Figure_STRATOS.tif", type="tif")
# =====
# END OF R CODE
# =====

```