R Script

Lifestyle risk factors and socioeconomic inequality in mortality: Decomposing differential exposure and vulnerability in the United States

Puka et al.

14/10/2021

Load libraries and dataset

```
# Load libraries
library(tidyverse)  # data management
library(skimr)      # descriptive statistics
library(tableone)   # create table one
library(survival)   # surivval analyses
library(survminer)  # surivval analyses
library(biostat3)   # survRate command
library(timereg)    # additive survival models
library(VGAM)       # multinomial regression, needed for causal mediation
library(MASS)       # needed for causal mediation functions
library(survey)     # for survey weighted cox model

setwd("~/")

# Import data
nhis        <- readRDS ("nhis.rds")
nhis_female <- filter(nhis, female==1)
nhis_male   <- filter(nhis, female==0)
```

Data Structure

Variable	Label	Values (Value Labels)
allcause_death	Status at last follow-up	0 (Alive); 1 (Deceased)
alcohol5v2	Alcohol use	1 (Never Drinker); 2 (Former Drinker); 3 (Category I; reference); 4 (Category II); 5 (Category III)
smoking4	Smoking status	1 (Never smoker; reference); 2 (Former smoker); 3 (Current some day smoker); 4 (Current everyday smoker)
bmi_cat	Body mass index	1 (Underweight); 2 (Healthy weight; reference); 3 (Overweight); 4 (Obese)
phy_act3	Physical activity	1 (Sedentary); 2 (Somewhat active); 3 (Active; reference)
edu	Educational attainment	1 (Highschool); 2 (Some college); 3 (Bachelors; reference)
female	Sex	0 (Men); 1 (Women)

Variable	Label	Values (Value Labels)
bl_age	Age (years) at baseline	Continuous: 25 to 85
end_age	Age (years) at last follow-up/death	Continuous: 25 to 85
ethnicity	Ethnicity	1 (Non-Hispanic White; referebce); 2 (Non-Hispanic Black); 3 (Hispanic); 4 (Other)
married	Marital status	0 (Not married/living together); 1 (Married/cohabitating)
srvy_yr	Survey year	Integer: 1997 to 2014
hed	Heavy episodic drinking	1 (No HED); 2 (HED <1/month); 3 (HED >1/month, <1/week); 4 (HED >=1/week)

Note: in the R script below variables with the suffix ".factor" use the value label noted above, whereas variables without the suffix use the numeric value noted above

Descriptive Statistics

Table 1

```
# Participant characteristics - Stratified by sex and education
tab1 <-CreateTableOne(vars= c("bl_age", "yrs_followup","allcause_death.factor",
                              "alcohol5v2.factor","smoking4.factor",  "bmi_cat.factor",
                              "phy_act3.factor", "ethnicity.factor", "married.factor"),
                   factorVars = c("allcause_death.factor", "alcohol5v2.factor",
                              "smoking4.factor", "bmi_cat.factor", "phy_act3.factor",
                              "ethnicity.factor", "married.factor"),
                   strata= c("edu.factor", "female.factor"), addOverall = TRUE, data=nhis)
  table1 <- print(tab1, noSpaces = TRUE, catDigits = 0, contDigits = 1,
                  printToggle = FALSE, test=FALSE)
  kableone(table1)


# Person years and death rate
survRate(Surv(yrs_followup, allcause_death)~1, data=nhis)          # overall
survRate(Surv(yrs_followup, allcause_death)~edu, data=nhis)        # for each SES category
survRate(Surv(yrs_followup, allcause_death)~female+edu, data=nhis) # for each SES & sex
```

Figure 2

```
# Survival plot
ggsurvplot_facet(fit = survfit(Surv(bl_age, end_age, allcause_death) ~ edu, data = nhis),
  data=nhis, facet.by="female.factor", censor = FALSE,xlim = c(25, 100),
  conf.int = TRUE,
  legend.labs = c("Low SES", "Medium SES", "High SES"),
  xlab = "Age (years)",
  ylab = "Overall survival probability")
```

```
# Age Medium Survival:
survfit(Surv(bl_age, end_age, allcause_death) ~ edu, data = nhis_female)
survfit(Surv(bl_age, end_age, allcause_death) ~ edu, data = nhis_male)
```

Objective 1: Aalen Additive Hazard Models

Check Assumption

First, check the time-invariant assumption (in our case, referred to as 'age-invariant'); whether the effect of covariates is age-varying or constant with time (similar to proportional hazard assumption in Cox models). The "const()" wrapper is used to make the effect of a variable age-invariant; without this wrapper the effect of the variable will be age-varying. Start by fitting the model where all components of the model have age-varying effects, then iteratively simplify the model by making the variables age-invariant one at a time (based on the plot and the Kolmogorov-Smirnov / Cramer von Mises tests).

Ultimately, the variables that are part of an interaction have to have a age-invariant effect, and sensitivity analyses (stratifying by age group) were ran to examine the potential impact if the assumption was violated.

```
# WOMEN, Alcohol * Education Model

## Iteration 1 - Start with all variables having age-varying effects

assump_alc_f1 <- aalen(Surv(bl_age, end_age, allcause_death) ~ alcohol5v2.factor +
                    edu.factor + married.factor + ethnicity.factor + factor(srvy_yr),
                    data = nhis_female)
    plot(assump_alc_f1)
    summary(assump_alc_f1)
    # RESULT: SrvyYear should be made age-invariant


## Iteration 2
assump_alc_f2 <- aalen(Surv(bl_age, end_age, allcause_death) ~ alcohol5v2.factor +
                    edu.factor + married.factor + ethnicity.factor +
                    const(factor(srvy_yr)), data = nhis_female)
    plot(assump_alc_f2)
    summary(assump_alc_f2)
    # RESULT: Marital Status should be made age-invariant


## Iteration 3
assump_alc_f3 <- aalen(Surv(bl_age, end_age, allcause_death) ~ alcohol5v2.factor +
                      edu.factor + const(married.factor) + ethnicity.factor +
                      const(factor(srvy_yr)), data = nhis_female)
    plot(assump_alc_f3)
    summary(assump_alc_f3)
    # RESULT: Education should be made age-invariant


## Iteration 4
assump_alc_f4 <- aalen(Surv(bl_age, end_age, allcause_death) ~ alcohol5v2.factor +
                    const(edu.factor) +  const(married.factor) + ethnicity.factor +
                    const(factor(srvy_yr)), data = nhis_female)
```

```
    plot(assump_alc_f4)
    summary(assump_alc_f4)
    # RESULT: alcohol (former, high risk) and ethnicity should be kept age-varying



# The process above was repeated for the other lifestyle factors and for males
```

Run Aalen Models

The effect estimates from the model can be directly interpreted as the number of additional events (deaths) per 1 person-year at risk. Two different versions of the model were ran to identify the interaction effect (the first model, with the interaction term) and the joint effects (the second model, with interacting variable).

```
# WOMEN, Alcohol * Education Model
## 1) Interaction model
aalen1 <- aalen(Surv(bl_age, end_age, allcause_death) ~ const(edu.factor) *
                  const(alcohol5v2.factor) + const(married.factor) +
                  ethnicity.factor + const(factor(srvy_yr)),  data = nhis_female)

## 2) Joint effect model
nhis_female$edu.alc <- interaction(nhis_female$edu.factor, nhis_female$alcohol5v2.factor)
aalen2 <- aalen(Surv(bl_age, end_age, allcause_death) ~ const(edu.alc) +
                  const(married.factor) + ethnicity.factor + const(factor(srvy_yr)),
                  data = nhis_female)


# Create function to extract the coefficients (based on their position) and
# multiple them by 10,000 (to get estimates per 10,000 person years)
result <- function(model, x) {
  mu <- model$gamma[x]
  var <-  model$var.gamma[x,x]
  confint.lower <- round((mu - (1.96 * sqrt(var)))*10000,1)
  confint.upper <- round((mu + (1.96 * sqrt(var)))*10000,1)
  mu <- round(mu*10000,1)
  output<-paste0(mu, " (",confint.lower,", ", confint.upper, ")")
  return(cat(output, "\n"))}


# Get the results
# The displayed results pertain to the effect of:
  # "Low SES, Category I drinking",
  # "High SES, Category III drinking",
  # "Low SES, Category III drinking" , and
  # "Additional deaths due interaction", respectively
result(aalen1, 1); result(aalen1, 6); result(aalen2, 13); result(aalen1, 31)



# The process above was repeated for the other life-style factors and for males
```

Objective 2: Causal Mediation Analyses

The causal mediation analyses involves four steps: 1) Fit separate multinomial logistic regressions with each mediator (M1, M2, M3, and M4) as the outcome. 2) Create copies of the dataset to account for all possible combinations of the exposure and mediators (3^4*= 81); the dataset was expanded from 229,994 to 18,629,514 (women) and 185,770 to 15,047,370 (men) pseudoobservations. 3) Using the expanded dataset, calculate weights for each mediator using the predicted probabilities from Step 1. 4) Fit a marginal structural model using Aalen additive hazards with the weight as weight and the id as a cluster level; this ensures thatrobust standard errors are calculated. The model with robust variance and resampling (robust=TRUE) was not used because of computation limiations.

```
# Step 0: Select data to use *********************************************************
mydata <- nhis_female %>%
  mutate(A.edu = edu,
    M1.alc = alcohol5v2,
    M2.smk = smoking4,
    M3.bmi = bmi_cat,
    M4.phy = phy_act3) %>%
  dplyr::select(A.edu, M1.alc, M2.smk, M3.bmi, M4.phy, allcause_death, bl_age, end_age,
                married, ethnicity, srvy_yr)

    # specifies the reference category
    mydata$A.edu <- factor(mydata$A.edu, levels=c(1,2,3), labels = c("Low", "Med", "High"))
    mydata$A.edu <- relevel(mydata$A.edu, ref = "High")

    # For technical reasons, the mediators should be coded as integers starting with 1


# Step 1: Fit a model for each mediator *********************************************
# First, create & use a copy of the exposure variable (for technical reasons related to R)
mydata$ATemp <- mydata$A.edu

# Fit model for each mediator, conditioning on exposure (education) and all covariates
fitM1 <- vglm(M1.alc ~ ATemp + bl_age + married + factor(ethnicity) + factor(srvy_yr),
              data = mydata, family=multinomial(refLevel = 3))

fitM2 <- vglm(M2.smk ~ ATemp + bl_age + married + factor(ethnicity) + factor(srvy_yr),
              data = mydata, family=multinomial(refLevel = 1))

fitM3 <- vglm(M3.bmi ~ ATemp + bl_age + married + factor(ethnicity) + factor(srvy_yr),
              data = mydata, family=multinomial(refLevel = 2))

fitM4 <- vglm(M4.phy ~ ATemp + bl_age + married + factor(ethnicity) + factor(srvy_yr),
              data = mydata, family=multinomial(refLevel = 3))



# Step 2: Construct copies of ID and exposure *****************************************

#Create an ID Variable
mydata$ID <- 1:nrow(mydata)

# Create counterfactual version of exposure; repeated 4 times because there are 4 mediators
levelsOfEDU <- unique(mydata$A.edu)
```

```
myData1 <- mydata
myData2 <- mydata
myData3 <- mydata
myData1$edu_M1.alc <- levelsOfEDU[1]
myData2$edu_M1.alc <- levelsOfEDU[2]
myData3$edu_M1.alc <- levelsOfEDU[3]
tempMyData <- rbind(myData1, myData2, myData3)

myData1 <- tempMyData
myData2 <- tempMyData
myData3 <- tempMyData
myData1$edu_M2.smk <- levelsOfEDU[1]
myData2$edu_M2.smk <- levelsOfEDU[2]
myData3$edu_M2.smk <- levelsOfEDU[3]
tempMyData <- rbind(myData1, myData2, myData3)

myData1 <- tempMyData
myData2 <- tempMyData
myData3 <- tempMyData
myData1$edu_M3.bmi <- levelsOfEDU[1]
myData2$edu_M3.bmi <- levelsOfEDU[2]
myData3$edu_M3.bmi <- levelsOfEDU[3]
tempMyData <- rbind(myData1, myData2, myData3)

myData1 <- tempMyData
myData2 <- tempMyData
myData3 <- tempMyData
myData1$edu_M4.phy <- levelsOfEDU[1]
myData2$edu_M4.phy <- levelsOfEDU[2]
myData3$edu_M4.phy <- levelsOfEDU[3]
newMyData <- rbind(myData1, myData2, myData3)




# Step 3: Construct weights *******************************************************

# M1: alcohol
newMyData$ATemp <- newMyData$A.edu
tempDir1 <- as.matrix(predict(fitM1,type = "response",
                       newdata=newMyData))[cbind(1:nrow(newMyData),newMyData$M1.alc)]

newMyData$ATemp <- newMyData$edu_M1.alc
tempIndir1 <- as.matrix(predict(fitM1,type = "response",
                         newdata=newMyData))[cbind(1:nrow(newMyData),newMyData$M1.alc)]

newMyData$weight1 <- tempIndir1/tempDir1


#M2: Smoking
newMyData$ATemp <- newMyData$A.edu
tempDir2 <- as.matrix(predict(fitM2,type = "response",
                       newdata=newMyData))[cbind(1:nrow(newMyData),newMyData$M2.smk]
```

```r
newMyData$ATemp <- newMyData$edu_M2.smk
tempIndir2 <- as.matrix(predict(fitM2,type = "response",
                        newdata=newMyData))[cbind(1:nrow(newMyData),newMyData$M2.smk)]

newMyData$weight2 <- tempIndir2/tempDir2


#M3: BMI
newMyData$ATemp <- newMyData$A.edu
tempDir3 <- as.matrix(predict(fitM3,type = "response",
                        newdata=newMyData))[cbind(1:nrow(newMyData),newMyData$M3.bmi)]

newMyData$ATemp <- newMyData$edu_M3.bmi
tempIndir3 <- as.matrix(predict(fitM3,type = "response",
                        newdata=newMyData))[cbind(1:nrow(newMyData),newMyData$M3.bmi)]

newMyData$weight3 <- tempIndir3/tempDir3


#M4: Physical activity
newMyData$ATemp <- newMyData$A.edu
tempDir4 <- as.matrix(predict(fitM4,type = "response",
                        newdata=newMyData))[cbind(1:nrow(newMyData),newMyData$M4.phy)]

newMyData$ATemp <- newMyData$edu_M4.phy
tempIndir4 <- as.matrix(predict(fitM4,type = "response",
                        newdata=newMyData))[cbind(1:nrow(newMyData),newMyData$M4.phy)]

newMyData$weight4 <- tempIndir4/tempDir4


# Final weight
newMyData$weightM <- newMyData$weight1 * newMyData$weight2 *
                        newMyData$weight3 * newMyData$weight4
hist(newMyData$weightM)


# Step 4: Run the Causal Mediation Model ************************************************

CMed_model <- aalen(Surv(bl_age, end_age, allcause_death) ~
                        const(A.edu) * const(edu_M1.alc) +
                        const(A.edu) * const(edu_M2.smk) +
                        const(A.edu) * const(edu_M3.bmi) +
                        const(A.edu) * const(edu_M4.phy) +
                        const(married) + factor(ethnicity) + const(factor(srvy_yr)),
                     data=newMyData, weights=newMyData$weightM,
                     clusters=newMyData$ID, robust=0)


# Step 5: Extract causal mediation results ***********************************************
```

```r
# The 'summary(model)' command will produce the direct effect, indirect effects, and
# mediated interaction effects. Functions were used to extract the other details.

# Function to get the total effect and proportion mediated
getTE <- function(CMed_model, v){
  TE <- sum(CMed_model$gamma[v])
  mu <- CMed_model$gamma[v]
  Omega <- CMed_model$var.gamma[v,v]      # To obtain non-robust estimates
  # Omega <- CMed_model$robvar.gamma[v,v] # To obtain robust estimates
  temp <- mvrnorm(n=10^4, mu=mu, Sigma=Omega)
  temp_TE <- apply(temp,1,sum)
  med_prop <- c(mu/TE,1)
  med_prop_CI <- rbind(t(apply(temp/temp_TE, 2, quantile, c(0.025, 0.975))), c(1,1))
  output <- cbind(c(mu,TE), c(apply(temp,2,sd),sd(temp_TE)), med_prop, med_prop_CI)
  colnames(output) <- c("Est.", "SE", "med_prop", "lowerCI", "UpperCI")
  rownames(output) <- c(rownames(CMed_model$gamma)[v],"TE")
  return(output)}


# Function to get the total combined indirect effect
getIE <- function(CMed_model, v){
  IE <- sum(CMed_model$gamma[v])
  mu <- CMed_model$gamma[v]
  Omega <- CMed_model$var.gamma[v,v] # To obtain non-robust estimates
  # Omega <- CMed_model$robvar.gamma[v,v] # To obtain robust estimates
  require(MASS)
  temp <- mvrnorm(n=10^4, mu=mu, Sigma=Omega)
  temp_IE <- apply(temp,1,sum)
  med_prop <- c(mu/IE,1)
  med_prop_CI <- rbind(t(apply(temp/temp_IE, 2, quantile, c(0.025, 0.975))), c(1,1))
  output <- cbind(c(mu,IE), c(apply(temp,2,sd),sd(temp_IE)), med_prop, med_prop_CI)
  colnames(output) <- c("Est.", "SE", "med_prop", "lowerCI", "UpperCI")
  rownames(output) <- c(rownames(CMed_model$gamma)[v],"IE")
  return(output)}

# Function to get the proportion mediated of the combined indirect effect
getTE_IE <- function(CMed_model, v, z){
  #total effect
  TE <- sum(CMed_model$gamma[v])
  mu <- CMed_model$gamma[v]
  Omega <- CMed_model$var.gamma[v,v]      # To obtain non-robust estimates
  # Omega <- CMed_model$robvar.gamma[v,v]  # To obtain robust estimates
  require(MASS)
  temp <- mvrnorm(n=10^4, mu=mu, Sigma=Omega)
  temp_TE <- apply(temp,1,sum)
  IE <- sum(CMed_model$gamma[z])
  muIE <- CMed_model$gamma[z]
  OmegaIE <- CMed_model$var.gamma[z,z]      # To obtain non-robust estimates
  #OmegaIE <- CMed_model$robvar.gamma[z,z] # To obtain robust estimates
  require(MASS)
  tempIE <- mvrnorm(n=10^4, mu=muIE, Sigma=OmegaIE)
  temp_IE <- apply(tempIE,1,sum)
  med_prop <- c(IE/TE,1)
```

8

```
  med_prop_CI <- (temp_IE/temp_TE)
  output <- cbind(IE, med_prop, quantile)
  quantile <- quantile(med_prop_CI, c(0.025, 0.975))
  output <- cbind(IE, med_prop, quantile)
  return(output)}
```

```
# Extract results
summary(CMed_model)
getTE(CMed_model, c(1,3,5,7,9,29,33,37,41))
getIE(CMed_model, c(3,5,7,9,29,33,37,41))
getTE_IE(CMed_model, c(1,3,5,7,9,29,33,37,41), c(3,5,7,9,29,33,37,41))
```

```
# The process above was repeated for males
```

Sensitivity Analyses

1) Alcohol use indexed by heavy episodic drinking (HED)

The analyses outlined above were repeated, using HED as the measure of alcohol use

2) Analyses stratified by age

Objective 1

```
# Women, Alcohol * Education Model

## Ages 25 - 60
    # Interaction model
    aalen1_a <- aalen(Surv(bl_age, end_age, allcause_death) ~ const(edu.factor) *
                         const(alcohol5v2.factor) + const(married.factor) +
                         ethnicity.factor + const(factor(srvy_yr)),
                         data = nhis_female, start.time=25, max.time=59.999)

    # Joint effect model
     aalen1_b <- aalen(Surv(bl_age, end_age, allcause_death) ~ const(edu.alc) +
                            const(married.factor) + ethnicity.factor +
                            const(factor(srvy_yr)),
                            data = nhis_female, start.time=25, max.time=59.999)


    # Get results (using the 'result' function created earlier)
    result(aalen1_a, 1); result(aalen1_a, 6); result(aalen1_b, 13); result(aalen1_a, 31)



## Ages 60 - 70
    # Interaction model
    aalen2_a <- aalen(Surv(bl_age, end_age, allcause_death) ~ const(edu.factor) *
```

```r
                              const(alcohol5v2.factor) + const(married.factor) +
                              ethnicity.factor + const(factor(srvy_yr)),
                              data = nhis_female, start.time=60, max.time=69.999)

    # Joint effect model
     aalen2_b <- aalen(Surv(bl_age, end_age, allcause_death) ~ const(edu.alc) +
                              const(married.factor) + ethnicity.factor +
                              const(factor(srvy_yr)),
                              data = nhis_female, start.time=60, max.time=69.999)

    # Get results (using the 'result' function created earlier)
    result(aalen2_a, 1); result(aalen2_a, 6); result(aalen2_b, 13); result(aalen2_a, 31)



## Ages 70 - 85
    # Interaction model
    aalen3_a <- aalen(Surv(bl_age, end_age, allcause_death) ~ const(edu.factor) *
                              const(alcohol5v2.factor) + const(married.factor) +
                              ethnicity.factor + const(factor(srvy_yr)),
                              data = nhis_female, start.time=70, max.time=84.999)

    # Joint effect model
     aalen3_b <- aalen(Surv(bl_age, end_age, allcause_death) ~ const(edu.alc) +
                              const(married.factor) + ethnicity.factor +
                              const(factor(srvy_yr)),
                              data = nhis_female, start.time=70, max.time=84.999)


    # Get results (using the 'result' function created earlier)
    result(aalen3_a, 1); result(aalen3_a, 6); result(aalen3_b, 13); result(aalen3_a, 31)



# The process above was repeated for the other life-style factors and for males
```

Objective 2

```r
# The expanded dataframe created earlier was used

# Ages 25 - 60
CMed_model_25.60 <- aalen(Surv(bl_age, end_age, allcause_death) ~
                              const(A.edu) * const(edu_M1.alc) +
                              const(A.edu) * const(edu_M2.smk) +
                              const(A.edu) * const(edu_M3.bmi) +
                              const(A.edu) * const(edu_M4.phy) +
                              const(married) + factor(ethnicity) +
                              const(factor(srvy_yr)),
                        start.time=25, max.time=59.999,
                        data=newMyData, weights=newMyData$weightM,
                        clusters=newMyData$ID, robust=0)
```

```r
    # Extract results
    summary(CMed_model_25.60)
    getTE(CMed_model_25.60, c(1,3,5,7,9,29,33,37,41))
    getIE(CMed_model_25.60, c(3,5,7,9,29,33,37,41))
    getTE_IE(CMed_model_25.60, c(1,3,5,7,9,29,33,37,41), c(3,5,7,9,29,33,37,41))



# Ages 60 - 70
CMed_model_60.70 <- aalen(Surv(bl_age, end_age, allcause_death) ~
                          const(A.edu) * const(edu_M1.alc) +
                          const(A.edu) * const(edu_M2.smk) +
                          const(A.edu) * const(edu_M3.bmi) +
                          const(A.edu) * const(edu_M4.phy) +
                          const(married) + factor(ethnicity) +
                          const(factor(srvy_yr)),
                     start.time=60, max.time=69.999,
                     data=newMyData, weights=newMyData$weightM,
                     clusters=newMyData$ID, robust=0)

    # Extract results
    summary(CMed_model_60.70)
    getTE(CMed_model_60.70, c(1,3,5,7,9,29,33,37,41))
    getIE(CMed_model_60.70, c(3,5,7,9,29,33,37,41))
    getTE_IE(CMed_model_60.70, c(1,3,5,7,9,29,33,37,41), c(3,5,7,9,29,33,37,41))



# Ages 70 - 85
CMed_model_70.85 <- aalen(Surv(bl_age, end_age, allcause_death) ~
                          const(A.edu) * const(edu_M1.alc) +
                          const(A.edu) * const(edu_M2.smk) +
                          const(A.edu) * const(edu_M3.bmi) +
                          const(A.edu) * const(edu_M4.phy) +
                          const(married) + factor(ethnicity) +
                          const(factor(srvy_yr)),
                     start.time=70, max.time=84.999,
                     data=newMyData, weights=newMyData$weightM,
                     clusters=newMyData$ID, robust=0)

    # Extract results
    summary(CMed_model_70.85)
    getTE(CMed_model_70.85, c(1,3,5,7,9,29,33,37,41))
    getIE(CMed_model_70.85, c(3,5,7,9,29,33,37,41))
    getTE_IE(CMed_model_70.85, c(1,3,5,7,9,29,33,37,41), c(3,5,7,9,29,33,37,41))
```

3) Analyses among the entire sample

The analyses outlined above were repeated, using the entire sample

4) (Multiplicative) Hazard models adjusted for the sampling design

```r
# Step 1: Create a database utilizing survey weights and design variables **************
nhis_svyWeights_all <- svydesign(id = ~new_psu,
                                 strata  = ~new_stratum,
                                 weights = ~new_weight,
                                 nest    = TRUE,
                                 data    = nhis_all)

  # Create subset with no missing data, correcting for survey weights
  nhis_svyWeights <- subset(nhis_svyWeights_all,
                            !is.na(yrs_followup) & !is.na(mortstat) &
                            !is.na(alcohol5v2) & !is.na(bmi_cat) & !is.na(smoking4) &
                            !is.na(phy_act3) & !is.na(edu) & !is.na(age) &
                            !is.na(female) & !is.na(married) & !is.na(ethnicity) &
                            (age>=25 & age <85))

  # Create subset with males or females only
  nhis_svyWeights_female <- subset(nhis_svyWeights,  female==1)
  nhis_svyWeights_male <- subset(nhis_svyWeights,  female==0)


# Step 2: Create function to extract and format results
cox_HR <- function(model, x) {
  mu <- model$coefficients[x]
  var <-  model$var[x,x]
  confint.lower <- round((exp(mu - (1.96 * sqrt(var)))),2)
  confint.upper <- round((exp(mu + (1.96 * sqrt(var)))),2)
  mu <- round(exp(mu),2)
  output<-paste0(mu, " (",confint.lower,", ", confint.upper, ")")
  return(cat(output, "\n"))}


# Step 3: Run Multiplicative (Cox) Hazard model WITH survey weights

# Interaction model
cox_wt <- svycoxph(Surv(bl_age, end_age, allcause_death) ~ edu.factor*alcohol5v2.factor +
                   married.factor + ethnicity.factor + factor(srvy_yr),
                   design=nhis_svyWeights_female)

# Joint effect model
cox_wt2 <- svycoxph(Surv(bl_age, end_age, allcause_death) ~ edu.alc +
                    married.factor + ethnicity.factor + factor(srvy_yr),
                    design=nhis_svyWeights_female)

# Get combined and formatted results using the function created earlier
cox_HR(cox_wt, 1) ; cox_HR(cox_wt, 6) ; cox_HR(cox_wt2, 13); cox_HR(cox_wt, 34)


    # Assessing proportional hazards
    cox_wt_check <- cox.zph(cox_wt)
    print(cox_wt_check)
    plot(cox_wt_check, col = "red")
```

```r
# Step 4: Run Multiplicative (Cox) Hazard model WITHOUT survey weights

    # Interaction model
    cox <- coxph(Surv(bl_age, end_age, allcause_death) ~ edu.factor*alcohol5v2.factor +
                     married.factor + ethnicity.factor + factor(srvy_yr), data=nhis_female)

    # Joint effect model
    cox2 <- coxph(Surv(bl_age, end_age, allcause_death) ~ edu.alc + married.factor +
                     ethnicity.factor + factor(srvy_yr), data=nhis_female)

    # Get combined and formatted results using the function created earlier
    cox_HR(cox, 1) ; cox_HR(cox, 6) ; cox_HR(cox2, 13); cox_HR(cox, 34)

    # Assessing proportional hazards
    cox_check <- cox.zph(cox)
    print(cox_check)
    plot(cox_check, col = "red")
```