Cohort define steps

1. From the entire database (**ED**), extract T2DM cohort with any codes in sheet “exposed”; then remove patients with any codes in sheet “exclude” before the end of the study (31/12/2018); use the first appearance diagnosis code of T2DM as the index date for exposed patient cohort (**T2C**).
2. From this cohort (**T2C**), remove patients with any codes in sheet “prevalent” or “cancer” appearing before the index date; this is the exposed cohort (**EC**).
3. From the entire database (**ED**), remove patients with any codes in sheet “non-exposed pool” before the end of the study (31/12/2018), to get the candidate non-exposed patients (**CNE**).
4. Exact matching the exposed cohort (**EC**) to the candidate non-exposed patients (**CNE**) with a ratio 1:10 (**EC:CNE**) by year of birth (+/-1year), sex, and practice, without replacement (each candidate non-exposed patient can be only matched once). This is the matched cohort (**MC**), and the index date is the same as the matched exposed patient.
5. From the matched cohort (**MC**), remove patients died before the index date; then remove patients with any codes in sheet “prevalent” or “cancer” appearing before the index date; then keep a randomly selected 5 matched non-exposed patients (good to set a seed to make the random selection replicable). This is the final non-exposed cohort (**FNE**).
6. The final study cohort is the combination of the exposed cohort (**EC**) and the final non-exposed cohort (**FNE**).

**La cohort defineix els passos**

1. De tota la base de dades (ED), extreu la cohort T2DM amb els codis del full "exposat"; després elimineu els pacients amb els codis del full "excloure" abans de finalitzar l'estudi (31/12/2018); utilitzeu el codi de diagnòstic de primera aparició de T2DM com a data índex de la cohort del pacient exposada (T2C).

2. D’aquesta cohort (T2C), traieu els pacients amb els codis del full “prevalent” o “càncer” que apareixin abans de la data de l’índex; es tracta de la cohort exposada (CE).

3. De tota la base de dades (ED), elimineu els pacients amb els codis del full "piscina no exposada" abans de finalitzar l'estudi (31/12/2018) per obtenir els pacients candidats no exposats (CNE).

4. Correspondre exactament a la cohort exposada (CE) als pacients candidats no exposats (CNE) amb una proporció 1:10 (EC: CNE) per any de naixement (+/- 1 any), sexe i pràctica, sense substitució ( cada pacient candidat no exposat només es pot combinar una vegada). Aquesta és la cohort coincident (MC) i la data de l’índex és la mateixa que el pacient exposat igualat.

5. De la cohort aparellada (MC), elimineu els pacients morts abans de la data de l’índex; a continuació, elimineu els pacients amb qualsevol codi en el full "prevalent" o "càncer" que aparegui abans de la data de l'índex; a continuació, mantingueu a 5 pacients no exposats seleccionats aleatòriament (és bo establir una llavor perquè la selecció aleatòria sigui replicable). Es tracta de la cohort final no exposada (FNE).

6. La cohort d'estudi final és la combinació de la cohorte exposada (EC) i la cohort final no exposada (FNE)

#--------------------------------#

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

#04.12.2019#

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

#--------------------------------#

citation("Epi")

citation("haven")

citation("dplyr")

#--------------------------------#

library(Epi)

#library(haven)

library(dplyr)

library(foreign)

#install.packages("broom")

library(broom)

library(popEpi)

# Llibreries necessaries

library("data.table")

library("SNPassoc")

library("htmlwidgets")

library("compareGroups")

library("foreign")

library("lattice")

library("Hmisc")

# library("ggplot2")

library("pander")

library("readxl")

library("knitr")

library("data.table")

library("MatchIt")

library("survival")

library("dplyr")

# library("survminer")

library("purrr")

library("stringr")

library("tidyr")

library("devtools")

library(lubridate)

#install.packages("here")

library("here")

#install.packages("data.table")

library("data.table")

#install.packages("remotes")

#remotes::install\_github("tagteam/heaven")

#install.packages("digest")

library("digest")

#install.packages("rlang")

library("rlang")

#--------------------------------#

#N O F U N C I O N A !!!

#--------------------------------#

#--------------------------------#

#install.packages("githubinstall")

#--------------------------------#

library("githubinstall")

#install.packages("remotes")

#remotes::install\_github("tagteam/heaven")

#--------------------------------#

#library(heaven)

#heaven::riskSetMatch

#--------------------------------#

#unlink("C:/Program Files/R/R-3.6.1/library/00LOCK-heaven, recursive = TRUE")

#--------------------------------##--------------------------------#

#Hola, et comento

#He estat investigant també, és el que dius tu,

#van fer dos llibreries més incidenceMatch() etc, i en les noves versions han eliminat la funció riskSetMatch()

#Ho he provat amb les meves dades i les coses que canvien són:

# 1. Els noms dels parámetres canvien, i han alguns arguments opocionals .....com la llavor ,etc...,

# 2. Va molt més lent ,( no sé perqué)

# 3. S'ha de vigilar amb els noms dels arguments dels paràmetres ja que si són els mateixos que el data.table que et genera, peta (p.e. el nom <event> no es pot posar com a paràmetre)

# 4. No et genera la data index pels grups a risc per tant després l'has d'assignar als grups a risc (casos i controls)

# 4. M'ha trobat uns quants més controls que l'anterior funció però ha tardat molt més. , nose si es per la llavor o que...

# 5. De totes maneres si vols utilitzar la versió antiga pots descarregar la llibreria amb el commit 964bbbd

# Installar versió de github 964bbbd (2018.08.09)

#githubinstall("heaven",ref="964bbbd",force=T)

#Qualsevol dubte em truques

#Fins aviat

#--------------------------------##--------------------------------#

#install.packages("heaven")

#--------------------------------------------------------------------------#

#library("heaven")

#--------------------------------------------------------------------------#

#githubinstall("heaven",ref="964bbbd",force=T)

#--------------------------------------------------------------------------#

# 1. Lectura de fitxers

#memory.size(max=160685)

memory.limit()

#

# Directori Font ==============================

#--------------------------------------------------------------------------#

link\_source<-paste0("https://github.com/jrealgatius/Stat\_codis/blob/master/funcions\_propies.R","?raw=T")

devtools::source\_url(link\_source)

#--------------------------------------------------------------------------#

#### DIRECTORI DE TREBALL

#### setwd en directori de treball

#--------------------------------#

#setwd("R:/LRWE\_Proj26/sl617/dataset/analysis")

#setwd("R:/LRWE\_Proj26/sl617/dataset/analysis")

#setwd("C:/Users/38122893W/Desktop/DataHarmonization")

#--------------------------------#

#testDAPCRMM\_entregable\_cmbdh\_diagnostics\_padris\_20190930\_093320

#--------------------------------#

# generar\_mostra\_fitxers() ----------------------

# Llegir tots els fitxers RDS dins d'un directori i generar una mostra aleatoria i salvar-lo en un directori "mostra"

# 1 Llegir fitxers sequencialment d'un directori

# 2 posarlos en una llista

# 3 Afafar la mostra i filtrar-los

# 4 Salvar-los en un directori

# mirara!!!

#generar\_mostra\_fitxers<-function(directori="dades/SIDIAP",

# fitxer\_poblacio="METPLUS\_entregable\_poblacio\_20181126\_190346.rds",

# mida\_mostra=10000,

# prefix="test",

# directori\_test="mostra\_test") {

#

# #directori="dades/SIDIAP"

# #fitxer\_poblacio="testDAPCRMM\_entregable\_cmbdh\_diagnostics\_padris\_20190930\_093320.rds"

# #mida\_mostra=10069

# #prefix="test"

# #directori\_test="mostra\_test"

#

#

# # Funció interna per llegir fitxer txt o rds

# LLEGIR.fitxer<-function(n,directori,fitxer) {

#

# if (stringr::str\_detect(fitxer,"\\.txt$")){

# dt<-data.table::fread(directori %>% here::here(fitxer)) %>% as\_tibble() %>% head(n)}

#

# if (stringr::str\_detect(fitxer,"\\.rds$")){

# dt<-readRDS(directori %>% here::here(fitxer)) %>% as\_tibble() %>% head(n)}

# dt}

#

#

# #k<-LLEGIR.fitxer(n=10069, directori="dades/SIDIAP",fitxer="testDAPCRMM\_entregable\_cmbdh\_diagnostics\_padris\_20190930\_093320.rds")

#

#

# # Llista de fitxers .rds | .txt

# llista\_de\_fitxers<-list.files(directori) [list.files(directori) %>% stringr::str\_detect("\\.rds$") |

# list.files(directori) %>% stringr::str\_detect("\\.txt$")]

#

# # Genero el directori mostra

# directori\_mostra<-paste0(directori,"/",directori\_test)

# if (!file.exists(directori\_mostra)) {

# # Crear directori si no existeix

# dir.create(file.path(directori,directori\_test), showWarnings = FALSE)

# }

#

#

#

#

# # Si NO existeix algun fitxer GENERAR LOS / Si EXISTEIX algun saltar

# if (!file.exists(paste0(directori\_mostra,"/",llista\_de\_fitxers)) %>% any()) {

#

# # Llegir ids mostra de fitxer poblacio

# dt\_ids<-LLEGIR.fitxer(mida\_mostra,directori,fitxer\_poblacio) %>% select(idp)

#

# # Posar noms per que els guardi

# llista\_de\_fitxers<-setNames(llista\_de\_fitxers,llista\_de\_fitxers)

# # Llegir fitxers complerts

# llista\_rds<-llista\_de\_fitxers %>% purrr::map(~LLEGIR.fitxer(n=Inf,directori = directori,fitxer=.x))

#

# # Filtrar via semijoint de tota la llista

# llista\_rds\_redux<-llista\_rds %>% purrr::map(~semi\_join(.x,dt\_ids))

#

# # Ara salvar-los en un surbdirectori amb el nom triat

#

# # Genero noms de fitxers dins directori test

# llista\_de\_fitxers<-str\_replace\_all(llista\_de\_fitxers, "\\.txt$", ".rds")

# llista\_de\_noms\_fitxers\_nous<-paste0(directori\_mostra,"/",prefix,llista\_de\_fitxers)

#

#

# # Salvo en format rds tots els fitxers en directori

# # saveRDS(llista\_rds\_redux[[1]],file=llista\_de\_fitxers\_nous[1])

# purrr::map2(llista\_rds\_redux,llista\_de\_noms\_fitxers\_nous,~saveRDS(.x,file=.y))

#

# }

#

# if (file.exists(paste0(directori\_mostra,"/",llista\_de\_fitxers)) %>% any()) {

# print ("Algun d'aquests fitxers ja existeix")

# }

#

#

# }

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

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

# generar\_mostra\_fitxers(

# directori="dades/SIDIAP",

# fitxer\_poblacio="testDAPCRMM\_entregable\_cmbdh\_diagnostics\_padris\_20190930\_093320.rds",

# mida\_mostra=1000,

# prefix="test",

# directori\_test="mostra\_test"

# )

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

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

#generar\_mostra:

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

#[[C:\Users\38122893W\Desktop\DataHarmonization\dades\SIDIAP]]#

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

#testtestDAPCRMM\_entregable\_cmbdh\_diagnostics\_padris\_20190930\_093320.rds

#testtestDAPCRMM\_entregable\_diagnostics\_20190926\_103409.rds

#testtestDAPCRMM\_entregable\_farmacs\_facturats\_20190926\_103409.rds

#testtestDAPCRMM\_entregable\_farmacs\_prescrits\_20190926\_103409.rds

#testtestDAPCRMM\_entregable\_poblacio\_20190926\_103409.rds

#testtestDAPCRMM\_entregable\_tabaquisme\_20190926\_103409.rds

#testtestDAPCRMM\_entregable\_variables\_analitiques\_20190926\_103409.rds

#testtestDAPCRMM\_entregable\_variables\_cliniques\_20190926\_103409.rds

#testtestDAPCRMM\_entregable\_variables\_geo\_sanitaries\_20190926\_103409.rds

#testtestDAPCRMM\_entregable\_variables\_socioeconomiques\_20190926\_103409.rds

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

#i [cmbdh\_diagnostics] mult

LLEGIR.cmbdh\_diagnostics\_padris<-readRDS("dades/SIDIAP/dades" %>% here::here("testDAPCRMM\_entregable\_cmbdh\_diagnostics\_padris\_20190930\_093320.rds")) %>% as\_tibble()

variable.names(LLEGIR.cmbdh\_diagnostics\_padris)

#[1] "idp" "cod" "dat" "dx\_pos" "dalta" "calta" "agr"

#----------------------------------------------#

#ii [dianostics] mult

LLEGIR.diagnostics<-readRDS("dades/SIDIAP/dades" %>% here::here("testDAPCRMM\_entregable\_diagnostics\_20190926\_103409.rds")) %>% as\_tibble()

variable.names(LLEGIR.diagnostics)

#[1] "idp" "cod" "dat" "dbaixa" "agr"

#----------------------------------------------#

#iii [farmacs\_facturats] mult

LLEGIR.farmacs\_facturat<-readRDS("dades/SIDIAP/dades" %>% here::here("testDAPCRMM\_entregable\_farmacs\_facturats\_20190926\_103409.rds")) %>% as\_tibble()

variable.names(LLEGIR.farmacs\_facturat)

#[1]"idp" "cod" "dat" "agr" "env"

#----------------------------------------------#

#iv [farmacs\_prescrits] mult

LLEGIR.farmacs\_prescrits<-readRDS("dades/SIDIAP/dades" %>% here::here("testDAPCRMM\_entregable\_farmacs\_prescrits\_20190926\_103409.rds")) %>% as\_tibble()

variable.names(LLEGIR.farmacs\_prescrits)

#[1] "idp" "cod" "dat" "dbaixa" "ics" "ap" "agr"

#----------------------------------------------#

#v [població] unic

LLEGIR.poblacio<-readRDS("dades/SIDIAP/dades" %>% here::here("testDAPCRMM\_entregable\_poblacio\_20190926\_103409.rds")) %>% as\_tibble()

variable.names(LLEGIR.poblacio)

#[1] "idp" "sexe" "dnaix" "entrada" "sortida" "situacio"

#----------------------------------------------#

#vi [tabaquisme] unic

LLEGIR.tabaquisme<-readRDS("dades/SIDIAP/dades" %>% here::here("testDAPCRMM\_entregable\_tabaquisme\_20190926\_103409.rds")) %>% as\_tibble()

variable.names(LLEGIR.tabaquisme)

#[1] "idp" "val" "dat" "dbaixa"

#----------------------------------------------#

#vii [analitiques] mult

LLEGIR.variables\_analitiques<-readRDS("dades/SIDIAP/dades" %>% here::here("testDAPCRMM\_entregable\_variables\_analitiques\_20190926\_103409.rds")) %>% as\_tibble()

variable.names(LLEGIR.variables\_analitiques)

#[1] "idp" "cod" "dat" "val" "agr"

#----------------------------------------------#

#viii [variables\_cliíniques] mult

LLEGIR.variables\_cliniques<-readRDS("dades/SIDIAP/dades" %>% here::here("testDAPCRMM\_entregable\_variables\_cliniques\_20190926\_103409.rds")) %>% as\_tibble()

variable.names(LLEGIR.variables\_cliniques)

#[1] "idp" "cod" "dat" "val" "agr"

#----------------------------------------------#

#ix [variables geo\_sanitàries] unic

LLEGIR.variables\_geo\_sanitaries<-readRDS("dades/SIDIAP/dades" %>% here::here("testDAPCRMM\_entregable\_variables\_geo\_sanitaries\_20190926\_103409.rds")) %>% as\_tibble()

variable.names(LLEGIR.variables\_geo\_sanitaries)

#[1] "idp" "idup" "idabs" "idrs" "iddap" "idambit"

#----------------------------------------------#

#x [variables socioeconòmiques] unic

LLEGIR.variables\_socioeconomiques<-readRDS("dades/SIDIAP/dades" %>% here::here("testDAPCRMM\_entregable\_variables\_socioeconomiques\_20190926\_103409.rds")) %>% as\_tibble()

variable.names(LLEGIR.variables\_socioeconomiques)

#[1] "idp" "qmedea" "ruralitat"

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

#Cohort define steps#

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

#1. From the entire database (ED), extract T2DM cohort with any codes in sheet “exposed”;

# then remove patients with any codes in sheet “exclude” before the end of the study (31/12/2018);

# use the first appearance diagnosis code of T2DM as the index date for exposed patient cohort (T2C).

# --------------------------------------------------------------------------------------------------------

# De tota la base de dades (ED), extreu la cohort T2DM amb els codis del full "exposat";

# després elimineu els pacients amb els codis del full "excloure" abans de finalitzar l'estudi (31/12/2018);

# utilitzeu el codi de diagnòstic de primera aparició de T2DM com a data índex de la cohort del pacient exposada (T2C).

# --------------------------------------------------------------------------------------------------------

#2. From this cohort (T2C),

# remove patients with any codes in sheet “prevalent” or “cancer” appearing before the index date;

# this is the exposed cohort (EC).

# --------------------------------------------------------------------------------------------------------

# D’aquesta cohort (T2C), traieu els pacients amb els codis del full “prevalent” o “càncer”

# que apareixin abans de la data de l’índex; es tracta de la cohort exposada (CE).

# --------------------------------------------------------------------------------------------------------

#3. From the entire database (ED),

# remove patients with any codes in sheet “non-exposed pool” before the end of the study (31/12/2018),

# to get the candidate non-exposed patients (CNE).

# --------------------------------------------------------------------------------------------------------

# De tota la base de dades (ED), elimineu els pacients amb els codis del full "piscina no exposada"

# abans de finalitzar l'estudi (31/12/2018) per obtenir els pacients candidats no exposats (CNE).

# --------------------------------------------------------------------------------------------------------

#4. Exact matching the exposed cohort (EC) to the candidate non-exposed patients (CNE) with a ratio 1:10 (EC:CNE)

# by year of birth (+/-1year), sex, and practice,

# without replacement (each candidate non-exposed patient can be only matched once).

# This is the matched cohort (MC), and the index date is the same as the matched exposed patient.

# --------------------------------------------------------------------------------------------------------

# Correspondre exactament a la cohort exposada (CE)

# als pacients candidats no exposats (CNE) amb una proporció 1:10 (EC: CNE)

# per any de naixement (+/- 1 any), sexe i pràctica, sense substitució

# ( cada pacient candidat no exposat només es pot combinar una vegada). Aquesta és la cohort coincident (MC)

# i la data de l’índex és la mateixa que el pacient exposat igualat.

# --------------------------------------------------------------------------------------------------------

#5. From the matched cohort (MC), remove patients died before the index date;

# then remove patients with any codes in sheet “prevalent” or “cancer” appearing before the index date;

# then keep a randomly selected 5 matched non-exposed patients

# (good to set a seed to make the random selection replicable). This is the final non-exposed cohort (FNE).

# --------------------------------------------------------------------------------------------------------

# De la cohort aparellada (MC), elimineu els pacients morts abans de la data de l’índex;

# a continuació, elimineu els pacients amb qualsevol codi en el full "prevalent" o "càncer"

# que aparegui abans de la data de l'índex; a continuació,

# mantingueu a 5 pacients no exposats seleccionats aleatòriament

# (és bo establir una llavor perquè la selecció aleatòria sigui replicable).

# Es tracta de la cohort final no exposada (FNE).

# --------------------------------------------------------------------------------------------------------

#6. The final study cohort is the combination of the exposed cohort (EC) and the final non-exposed cohort (FNE).

# --------------------------------------------------------------------------------------------------------

# La cohort d'estudi final és la combinació de la cohorte exposada (EC) i la cohort final no exposada (FNE)

# --------------------------------------------------------------------------------------------------------

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

# Generar data index -----------

dt\_diagnostics<-LLEGIR.cmbdh\_diagnostics\_padris %>%

transmute(idp,cod=as.character(cod),dat,agr) %>%

bind\_rows(select(LLEGIR.diagnostics,idp,cod,dat,agr))

#data minima[]#

dt\_cataleg<-read\_excel("Spain\_codes.xls") %>% select(cod,agr,exposed)

dt\_diagnostics<-dt\_diagnostics %>% left\_join(dt\_cataleg,by="cod") %>% filter(exposed=="exposed")

#-----#

DINDEX<-dt\_diagnostics%>% group\_by(idp)%>%summarise(data\_index=min(dat,na.rm = TRUE))%>%ungroup()

DINDEX

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

#-----------------------------------------------------------------------------------------------------------------------#

#[ Crearé una base de dades dels exposat( TOTS ELS DELS DINDEX!!), amb ANY DE NAIXAMENT+SEXE]

#[[població]]

C\_EXPOSATS<-DINDEX%>%left\_join(LLEGIR.poblacio,by="idp")%>%mutate(Edat=as.numeric(ymd(data\_index)-ymd(dnaix))/365.25 )

C\_EXPOSATS<-C\_EXPOSATS%>%filter(sortida<=20181231) #Excluits finalitazar després de (31/12/2018)

variable.names(C\_EXPOSATS)

#[1] "idp" "data\_index" "sexe" "dnaix" "entrada" "sortida" "situacio" "Edat"

#-----------------------------------------------------------------------------------------------------------------------#

#[ Crearé una altre base de dades , que seran els No exposats,tots menys els exposats!!]

C\_NO\_EXPOSATS<-LLEGIR.poblacio %>% anti\_join(C\_EXPOSATS,by="idp")

C\_NO\_EXPOSATS<-C\_NO\_EXPOSATS%>%filter(sortida<=20181231) #Excluits finalitazar després de (31/12/2018)

variable.names(C\_NO\_EXPOSATS)

#[1] "idp" "sexe" "dnaix" "entrada" "sortida" "situacio"

#-----------------------------------------------------------------------------------------------------------------------#

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

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

# (demà fer-ho!)

#[ DESPRÉS FARÉ UN MATCHED COHORT [ 1 Exposició: 10 No Exposició], ANY DE NAIXAMENT (+/- 1 ANY), SEXE]

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

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

#[ De la coohrt aparellada, eliminarem els pacients Morts abans de la data Index,eliminem els pacients amb qualsevol

# codi en el full "prevalent" o "Cancer"]

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

#-----#

#mirar si a la data del dia Index tenen 35 anys o mes!, tenir un minim un any d'història clínica prèvia

#abans de de dia index ?

# MATCH amb els controls!

#Introducció

#Al present document tractarem de traduir el protocol original del projecte a l’estructura actual del SIDIAP,

#convertint-se aquest en el manual fet servir pels Data Managers del SIDIAP durant l’extracció de les dades.

#És un document INTERN del SIDIAP.

#Objectius del projecte

# Objectius principals #

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

#-----------------------------------------------------------------------------------------------------------------------------#

#● Avaluar les tendències al llarg del temps en mortalitat de totes les causes entre el 20060101 i

# la captura més recent del SIDIAP (20181231) en persones amb recent diabetis tipus 2

# ien comparació amb una població control sense diabetis.

#-----------------------------------------------------------------------------------------------------------------------------#

#● Avaluar les diferències i les proporcions de les taxes de mortalitat de totes les causes en persones amb diabetis tipus 2

# i sense diabetis entre el 20060101 i la captura més recent del SIDIAP (20181231).

#-----------------------------------------------------------------------------------------------------------------------------#

# Objectius secundaris #

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

#-----------------------------------------------------------------------------------------------------------------------------#

#● Avaluar les tendències al llarg del temps en la mortalitat cardiovascula r entre 20060101

# i la captura més recent del SIDIAP en persones amb diabetis tipus 2 i sense diabetis.

#-----------------------------------------------------------------------------------------------------------------------------#

#● Avaluar les diferències i les proporcions

# de les taxes de mortalitat per causes determinades en persones amb diabetis tipus 2 i sense diabetis entre el 20060101

# i la captura més recent del SIDIAP.

#-----------------------------------------------------------------------------------------------------------------------------#

#● Comparar les tendències de les taxes de mortalitat i de ràtio entre els diferents països.

#-----------------------------------------------------------------------------------------------------------------------------#

#● Avaluar les tendències del temps en condicions renals cardiometabòliques.

#-----------------------------------------------------------------------------------------------------------------------------#

# 20060101 la captura més recent del SIDIAP (20181231)

#[1.1.2006---31.12.2018]

# data.index []

#[La data del primer diagnòstic de DM2 passarà a ser la data d’índex]: [[DINDEX]]

# 0. Inicialització de parametres -----------------------------

# N test mostra a seleccionar (Nmostra=Inf)

# Nmostra=Inf # Seria tota la mostra

Nmostra=Inf

# fitxer conductor cataleg

fitxer\_conductor\_cataleg<-"Spain\_codes.xls"

# fitxer conductor variables

fitxer\_conductor\_variables<-"variables\_precav.xls"

# write.xlsx(CATALEG,file="cataleg.xlsx")

CATALEG<-readxl::read\_excel(fitxer\_conductor\_cataleg,col\_types = "text")

# Funcions lectura de fitxers

LLEGIR.PACIENTS<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_pacients\_20190517\_101801.rds")) %>% as\_tibble() %>% head(n)}

LLEGIR.PROBLEMES<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_problemes\_20181123\_172533.rds"))%>% as\_tibble() %>% head(n)}

LLEGIR.CMBDH<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_cmbd\_dx\_20181123\_172533.rds"))%>% as\_tibble() %>% head(n)}

LLEGIR.padris<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_cmbd\_dx\_padris\_20181123\_172533.rds"))%>% as\_tibble() %>% head(n)}

LLEGIR.PROC<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_cmbd\_px\_padris\_20181123\_172533.rds"))%>% as\_tibble() %>% head(n)}

LLEGIR.TABAC<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_tabaquisme\_20181123\_172533.rds"))%>% as\_tibble() %>% head(n) }

# Llegit pacients i problemes de salut

library(dplyr)

# Llegir PACIENTS, I PROBLEMES DE SALUT

PACIENTS<-Inf %>% LLEGIR.PACIENTS()

PROBLEMES<-Nmostra %>% LLEGIR.PROBLEMES()

# Criteris d'Inclusió 1 (Any naixament)

PACIENTS<-PACIENTS %>%

filter (dnaix<19720101) # Excluye Nacidos a partir del 1972

# Seleccionar tots els pacients que han estat DM durant periode

# 1. Extreure T2DM (exposed) Agrego problemes amb límit data máxima (31/12/2016)

#### Fusiono agregador (+agr)

dadesDM\_exposats<-PROBLEMES %>%

agregar\_problemes\_agr(bd.dindex="20161231",agregador ="exposed",dt.agregadors=CATALEG,finestra.dies=c(-Inf,0),prefix="",camp\_agregador="exposed") %>% # agrego problemes de salut

select(idp,dtevent="exposed",codindex=cod)

dades\_exclos<-PROBLEMES %>%

agregar\_problemes\_agr(bd.dindex="20161231",agregador ="exclude",dt.agregadors=CATALEG,finestra.dies=c(-Inf,0),prefix="",camp\_agregador="exclude") %>% # agrego problemes de salut

select(idp,dtevent="exclude",codindex=cod)

# filtro dadesDM\_exposats (Elimino dades excloses)

dadesDM\_exposats<-dadesDM\_exposats %>% anti\_join(dades\_exclos,by="idp")

# 2. From this cohort (T2C), remove patients with any codes in sheet “prevalent” or “cancer” appearing before the index date; this is the exposed cohort (EC).

bd\_dtindex<-dadesDM\_exposats %>% select(idp,dataindex=dtevent) %>% mutate(dataindex=as.Date(dataindex,origin="1970-01-01"))

historic\_problemes<-PROBLEMES %>% mutate(dat=as.Date(as.character(dat),format="%Y%m%d")) %>% select(idp,cod,dat)

dades\_exclos2<-historic\_problemes %>%

agregar\_problemes\_agr(bd.dindex=bd\_dtindex,agregador ="prevalent",dt.agregadors=CATALEG,finestra.dies=c(-Inf,0),prefix="",camp\_agregador="prevalent")

#[]

dadesDM\_exposats<-dadesDM\_exposats %>% anti\_join(dades\_exclos2,by="idp")

# 3. From the entire database (ED), remove patients with any codes in sheet “non-exposed pool” before the end of the study (31/12/2018), to get the candidate non-exposed patients (CNE).

## Eliminar de població (PACIENTS) dades\_exclos

PACIENTS<-PACIENTS %>% anti\_join(dades\_exclos,by="idp")

# 4. Exact matching the exposed cohort (EC) to the candidate non-exposed patients (CNE) with a ratio 1:10 (EC:CNE)

# by year of birth (+/-1year), sex, and practice, without replacement (each candidate non-exposed patient can be only matched once).

# This is the matched cohort (MC), and the index date is the same as the matched exposed patient.

# Juntar casos a PACIENTS

PACIENTS<-PACIENTS %>% left\_join(dadesDM\_exposats,by="idp") %>% mutate(grup=ifelse(is.na(dtevent),0,1))

# Preparar matching i setriskmatching #

dt\_matching<-PACIENTS %>% select(idp,idup,dnaix,sexe,grup,dtevent,sortida)

# 5.2.1 Generar data de sortida (Data event / Data de censura) -----------------

## dtindex\_case

dt\_matching<-dt\_matching %>% mutate(dtindex\_case=ifelse(grup==1, as.Date(as.character(dtevent),format="%Y%m%d"),NA))

## dtindex\_control

dt\_matching<-dt\_matching %>% mutate (dtindex\_control=as.Date(as.character(sortida),format="%Y%m%d")%>% as.numeric())

## Generar any de naixament i grups cada 10

dt\_matching<-dt\_matching %>% mutate (

any\_naix=lubridate::year(lubridate::ymd(dnaix)))

#### Parametres d'aparellament

llistaPS=c("sexe","any\_naix","idup")

num\_controls<-5

llavor<-125

set.seed(llavor)

# 5.4.1. Aplicar algoritme -----------

dades\_match<-heaven::riskSetMatch(ptid="idp" # Unique patient identifier

,event="grup" # 0=Control, 1=case

,terms=llistaPS # terms c("n1","n2",...) - list of vairables to match by

,dat=dt\_matching # dataset with all variables

,Ncontrols=num\_controls # number of controls to provide

,oldevent="oldevent" # To distinguish cases used as controls

,caseid="caseid" # variable to group cases and controls (case-ptid)

,reuseCases=F # T og F or NULL - can a case be a control prior to being a case?

,reuseControls=F # T or F or NULL - can controls be reused?

,caseIndex="dtindex\_case" # Integer or date, date where controls must be prior

,controlIndex="dtindex\_control" # controlIndex - Index date for controls

,NoIndex=FALSE # If T ignore index

,cores=1 # Number of cores to use, default 1

,dateterms=NULL # character list of date variables

)

# Report matchreport -------------

heaven::matchReport(dades\_match, id="idp",case="grup",caseid="caseid")

# Número de controls per conjunt a risk ------------

dades\_match[,numControls:=.N,by=caseid]

dades\_match<- dades\_match %>% mutate(numControls=numControls-1)

library("tagteam")

#--------------------------------#

#db <-read\_dta("db.dta")

#nrow(db)

#--------------------------------#

######all-cause mortality in DM####

db1 <-Lexis(entry = list(period = yearin,

age = agein),

exit = list(period = outm),

exit.status = acm,

id = patid,

data = subset(db, DM == 1))

dbs1 <- splitMulti(db1, age = seq(35,100,1), period= seq(1998,2018,1))

a.kn <- with(subset(dbs1, lex.Xst==1), quantile(age+lex.dur,(1:5-0.5)/5))

p.kn <- with(subset(dbs1, lex.Xst==1), quantile(period+lex.dur,(1:5-0.5)/5))

r1 <- glm((lex.Xst==1)~Ns(age, knots = a.kn)\*Ns(period, knots = p.kn)\*gender,

family = poisson,

offset = log(lex.dur),

data = dbs1)

age <- c(35:100)

period <- seq(1998,2018,0.1)

gender <- c(1:2)

nd <- expand.grid(age, period, gender)

colnames(nd) <- c("age","period","gender")

nd <- cbind(nd, lex.dur=1000)

p1 <- ci.pred(r1, newdata = nd, Exp = FALSE)

colnames(p1) <- c("es\_d", "lb\_d", "ub\_d")

acm\_DM <- cbind(nd,p1, out="acm")

#####cardio-renal-mortality in DM#####

db1 <-Lexis(entry = list(period = yearin,

age = agein),

exit = list(period = outm),

exit.status = crm,

id = patid,

data = subset(db, DM == 1))

dbs1 <-splitMulti(db1, age = seq(35,100,1), period= seq(1998,2018,1))

a.kn <- with(subset(dbs1, lex.Xst==1), quantile(age+lex.dur,(1:5-0.5)/5))

p.kn <- with(subset(dbs1, lex.Xst==1), quantile(period+lex.dur,(1:5-0.5)/5))

r1 <- glm((lex.Xst==1)~Ns(age, knots = a.kn)\*Ns(period, knots = p.kn)\*gender,

family = poisson,

offset = log(lex.dur),

data = dbs1)

age <- c(35:100)

period <- seq(1998,2018,0.1)

gender <- c(1:2)

nd <- expand.grid(age, period, gender)

colnames(nd) <- c("age","period","gender")

nd <- cbind(nd, lex.dur=1000)

p1 <- ci.pred(r1, newdata = nd, Exp = FALSE)

colnames(p1) <- c("es\_d", "lb\_d", "ub\_d")

crm\_DM <- cbind(nd,p1, out="crm")

res\_DM <-rbind(acm\_DM,crm\_DM)

res\_DM <-cbind(res\_DM, rateD=exp(res\_DM$es\_d), rateD\_lb=exp(res\_DM$lb\_d), rateD\_ub=exp(res\_DM$ub\_d))

write.dta(res\_DM, file="res\_DM.dta")

# 0. Inicialització de parametres -----------------------------

# N test mostra a seleccionar (Nmostra=Inf)

# Nmostra=Inf # Seria tota la mostra

Nmostra=Inf

# fitxer conductor cataleg

fitxer\_conductor\_cataleg<-"Spain\_codes.xls"

# fitxer conductor variables

fitxer\_conductor\_variables<-"variables\_precav.xls"

# write.xlsx(CATALEG,file="cataleg.xlsx")

CATALEG<-readxl::read\_excel(fitxer\_conductor\_cataleg,col\_types = "text")

# Funcions lectura de fitxers

LLEGIR.PACIENTS<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_pacients\_20190517\_101801.rds")) %>% as\_tibble() %>% head(n)}

LLEGIR.PROBLEMES<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_problemes\_20181123\_172533.rds"))%>% as\_tibble() %>% head(n)}

LLEGIR.CMBDH<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_cmbd\_dx\_20181123\_172533.rds"))%>% as\_tibble() %>% head(n)}

LLEGIR.padris<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_cmbd\_dx\_padris\_20181123\_172533.rds"))%>% as\_tibble() %>% head(n)}

LLEGIR.PROC<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_cmbd\_px\_padris\_20181123\_172533.rds"))%>% as\_tibble() %>% head(n)}

LLEGIR.TABAC<<-function(n=Nmostra) {

readRDS("dades/sidiap\_precav" %>% here::here("ECV\_CAT\_entregable\_tabaquisme\_20181123\_172533.rds"))%>% as\_tibble() %>% head(n) }

# Llegit pacients i problemes de salut

library(dplyr)

# Llegir PACIENTS, I PROBLEMES DE SALUT

PACIENTS<-Inf %>% LLEGIR.PACIENTS()

PROBLEMES<-Nmostra %>% LLEGIR.PROBLEMES()

# Criteris d'Inclusió 1 (Any naixament)

PACIENTS<-PACIENTS %>%

filter (dnaix<19720101) # Excluye Nacidos a partir del 1972

# Seleccionar tots els pacients que han estat DM durant periode

# 1. Extreure T2DM (exposed) Agrego problemes amb límit data máxima (31/12/2016)

#### Fusiono agregador (+agr)

dadesDM\_exposats<-PROBLEMES %>%

agregar\_problemes\_agr(bd.dindex="20161231",agregador ="exposed",dt.agregadors=CATALEG,finestra.dies=c(-Inf,0),prefix="",camp\_agregador="exposed") %>% # agrego problemes de salut

select(idp,dtevent="exposed",codindex=cod)

dades\_exclos<-PROBLEMES %>%

agregar\_problemes\_agr(bd.dindex="20161231",agregador ="exclude",dt.agregadors=CATALEG,finestra.dies=c(-Inf,0),prefix="",camp\_agregador="exclude") %>% # agrego problemes de salut

select(idp,dtevent="exclude",codindex=cod)

# filtro dadesDM\_exposats (Elimino dades excloses)

dadesDM\_exposats<-dadesDM\_exposats %>% anti\_join(dades\_exclos,by="idp")

# 2. From this cohort (T2C), remove patients with any codes in sheet “prevalent” or “cancer” appearing before the index date; this is the exposed cohort (EC).

bd\_dtindex<-dadesDM\_exposats %>% select(idp,dataindex=dtevent) %>% mutate(dataindex=as.Date(dataindex,origin="1970-01-01"))

historic\_problemes<-PROBLEMES %>% mutate(dat=as.Date(as.character(dat),format="%Y%m%d")) %>% select(idp,cod,dat)

dades\_exclos2<-historic\_problemes %>%

agregar\_problemes\_agr(bd.dindex=bd\_dtindex,agregador ="prevalent",dt.agregadors=CATALEG,finestra.dies=c(-Inf,0),prefix="",camp\_agregador="prevalent")

dadesDM\_exposats<-dadesDM\_exposats %>% anti\_join(dades\_exclos2,by="idp")

# 3. From the entire database (ED), remove patients with any codes in sheet “non-exposed pool” before the end of the study (31/12/2018), to get the candidate non-exposed patients (CNE).

## Eliminar de població (PACIENTS) dades\_exclos

PACIENTS<-PACIENTS %>% anti\_join(dades\_exclos,by="idp")

# 4. Exact matching the exposed cohort (EC) to the candidate non-exposed patients (CNE) with a ratio 1:10 (EC:CNE)

# by year of birth (+/-1year), sex, and practice, without replacement (each candidate non-exposed patient can be only matched once).

# This is the matched cohort (MC), and the index date is the same as the matched exposed patient.

# Juntar casos a PACIENTS

PACIENTS<-PACIENTS %>% left\_join(dadesDM\_exposats,by="idp") %>% mutate(grup=ifelse(is.na(dtevent),0,1))

# Preparar matching i setriskmatching #

dt\_matching<-PACIENTS %>% select(idp,idup,dnaix,sexe,grup,dtevent,sortida)

# 5.2.1 Generar data de sortida (Data event / Data de censura) -----------------

## dtindex\_case

dt\_matching<-dt\_matching %>% mutate(dtindex\_case=ifelse(grup==1, as.Date(as.character(dtevent),format="%Y%m%d"),NA))

## dtindex\_control

dt\_matching<-dt\_matching %>% mutate (dtindex\_control=as.Date(as.character(sortida),format="%Y%m%d")%>% as.numeric())

## Generar any de naixament i grups cada 10

dt\_matching<-dt\_matching %>% mutate (

any\_naix=lubridate::year(lubridate::ymd(dnaix)))

#### Parametres d'aparellament

llistaPS=c("sexe","any\_naix","idup")

num\_controls<-5

llavor<-125

set.seed(llavor)

# 5.4.1. Aplicar algoritme -----------

dades\_match<-heaven::riskSetMatch(ptid="idp" # Unique patient identifier

,event="grup" # 0=Control, 1=case

,terms=llistaPS # terms c("n1","n2",...) - list of vairables to match by

,dat=dt\_matching # dataset with all variables

,Ncontrols=num\_controls # number of controls to provide

,oldevent="oldevent" # To distinguish cases used as controls

,caseid="caseid" # variable to group cases and controls (case-ptid)

,reuseCases=F # T og F or NULL - can a case be a control prior to being a case?

,reuseControls=F # T or F or NULL - can controls be reused?

,caseIndex="dtindex\_case" # Integer or date, date where controls must be prior

,controlIndex="dtindex\_control" # controlIndex - Index date for controls

,NoIndex=FALSE # If T ignore index

,cores=1 # Number of cores to use, default 1

,dateterms=NULL # character list of date variables

)

# Report matchreport -------------

heaven::matchReport(dades\_match, id="idp",case="grup",caseid="caseid")

# Número de controls per conjunt a risk ------------

dades\_match[,numControls:=.N,by=caseid]

dades\_match<- dades\_match %>% mutate(numControls=numControls-1)

#

bdades\_index<-dades\_match %>%

select(idp,dataindex=dtindex\_case)%>%

mutate (dataindex=as.Date(dataindex,origin = "1970-01-01")) %>%

as\_tibble()

dades\_match<-dades\_match %>% rename(dtindex=dtindex\_case) %>% mutate (dtindex=as.Date(dtindex,origin = "1970-01-01"))

dades\_match<-dades\_match %>% as\_tibble()

gc()

# Verificació de Matching aprox -----------------------

table(dades\_match$numControls,dades\_match$grup)

descrTable(formula\_vector(llistaPS,y="grup"),data=dades\_match)

# extreure\_OR("event~sexe+any\_naix",dades=dades\_match,conditional = T,str

# any\_dg i filtrar per any

dades\_match<-dades\_match %>% mutate(yearindex=lubridate::year(dtindex)) %>% select(-c(idup,dnaix,sexe,sortida)) %>%

filter(yearindex>=2006)

# filtrar per Pacients

PACIENTS\_MATCH<-dades\_match %>% left\_join(select(PACIENTS,-c(grup,dtevent)), by="idp")

table(PACIENTS\_MATCH$yearindex,PACIENTS\_MATCH$grup)

descrTable(formula\_vector(llistaPS,y="grup"),data=dades\_match)

# Filtrar per edat=35-100 anys en data index

library(lubridate)

PACIENTS\_MATCH<-PACIENTS\_MATCH %>%

mutate(edat = year(as.period(interval(start = ymd(dnaix), end = dtindex))))

PACIENTS\_MATCH<-PACIENTS\_MATCH %>% filter(edat>=35 & edat<=100)