

# dataset36\_\_withmed

Fay Yan

2024-07-22

load data dataset: newDatabase\_36\_ANON

```
library(readxl)
newDatabase_36_ANON <- read_excel("~/Desktop/newDatabase_36_ANON.xlsx",
  sheet = "Asymptomatic_CAC", skip = 2)
```

```
## New names:
## * 'Date' -> 'Date...135'
## * 'Date' -> 'Date...137'
## * 'Date' -> 'Date...139'
## * 'Date' -> 'Date...141'
## * 'Excision' -> 'Excision...149'
## * 'Excision Date' -> 'Excision Date...150'
## * 'Radiation' -> 'Radiation...151'
## * 'Radiation Date' -> 'Radiation Date...152'
## * 'Chemo' -> 'Chemo...153'
## * 'Chemo Date' -> 'Chemo Date...154'
## * 'Excision' -> 'Excision...158'
## * 'Excision Date' -> 'Excision Date...159'
## * 'Radiation' -> 'Radiation...160'
## * 'Radiation Date' -> 'Radiation Date...161'
## * 'Chemo' -> 'Chemo...162'
## * 'Chemo Date' -> 'Chemo Date...163'
## * 'Excision' -> 'Excision...166'
## * 'Excision Date' -> 'Excision Date...167'
## * 'Radiation' -> 'Radiation...168'
## * 'Radiation Date' -> 'Radiation Date...169'
## * 'Chemo' -> 'Chemo...170'
## * 'Chemo Date' -> 'Chemo Date...171'
## * 'Chemo Date' -> 'Chemo Date...179'
```

select column: (including variables for medication)

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

```
ANON36_selected <- newDatabase_36_ANON[, c(1:6,11,40,41,43,44,45,47,49,
                                           50:57,61:64,65:66,96,107:115)]
```

For risk factors, encode to binary (change NA to 0)

```
ANON36_selected <- ANON36_selected %>%
  mutate(across(15:22, ~ ifelse(is.na(.) | . == "", 0, .)))
```

add column Male:

```
ANON36_selected <- ANON36_selected %>%
  mutate(Male = ifelse(Gender == "Male", 1, 0))
```

handle Race column: "American Indian or Alaskan Native" = 1 "Asian" = 2 "Black" = 3 "Hispanic" = 4 "Other" = 5 "Unknown" = 6 "White" = 7 NA = missing

```
# Step 1: Convert "NA" strings to actual NA values
ANON36_selected$Race[ANON36_selected$Race == "NA"] <- NA
# Step 2: Convert Race to a factor
ANON36_selected$Race <- as.factor(ANON36_selected$Race)
# Step 3: Manually encode Race as numeric, keeping NA as NA
ANON36_selected$Race_encoded <- ifelse(is.na(ANON36_selected$Race), NA, as.numeric(ANON36_selected$Race))
```

For medication columns, encode to binary (change NA to 0)

```
ANON36_selected_withmed <- ANON36_selected %>%
  mutate(across(c(`Beta Blockers`, `Antianginal Agents`, `Antihyperlipidemic`, `Calcium Blockers`, `Antidiabetic`, `Antiarrhythmic`),
    ~ ifelse(grepl("prior|on date", ., ignore.case = TRUE), 1, 0)))
```

MACE dataset: when either MACE or Death = 1, result = 1; otherwise 0

```
ANON36_MACE <- ANON36_selected_withmed[, c(1,4,7:26,29:40)]
ANON36_MACE <- ANON36_MACE %>%
  mutate(result = if_else(!is.na(`First MACE or Death (1=MACE, 2=Death)`), 1, 0))
ANON36_MACE_output <- ANON36_MACE[, -c(23)]
```

```
write.csv(ANON36_MACE_output, "/Users/Student/Desktop/ANON36_MACE_output.csv", row.names = FALSE)
```

HF dataset: when CHF date after exam date, result = 1; otherwise 0

```
ANON36_HF <- ANON36_selected_withmed[, c(1:2,4,7:28,30:40)]
ANON36_HF$`Exam Completion Timestamp` <- as.Date(ANON36_HF$`Exam Completion Timestamp`, format = "%Y-%m-%d")
ANON36_HF$`CHF Dx Date` <- as.Date(ANON36_HF$`CHF Dx Date`, format = "%Y-%m-%d")
ANON36_HF <- ANON36_HF %>%
  mutate(is_later = case_when(
    !is.na(`Exam Completion Timestamp`) & !is.na(`CHF Dx Date`) ~ `CHF Dx Date` > `Exam Completion Timestamp`
```

```

      TRUE ~ NA
    ))
  ANON36_HF <- ANON36_HF %>%
    mutate(result = if_else(is.na(is_later) | is_later == FALSE, 0, 1))
  ANON36_HF_output <- ANON36_HF[, -c(2,24,25,37)]

write.csv(ANON36_HF_output, "/Users/Student/Desktop/ANON36_HF_output.csv", row.names = FALSE)

```

death dataset: when Death = 1, result = 1; otherwise 0

```

ANON36_death <- ANON36_selected_withmed[, -c(2,3,5,6,27:28)]
ANON36_death <- ANON36_death %>%
  mutate(`First MACE or Death (1=MACE, 2=Death)` = ifelse(is.na(`First MACE or Death (1=MACE, 2=Death)`),
    mutate(result = ifelse(`First MACE or Death (1=MACE, 2=Death)` == 2, 1, 0))
  ANON36_death_output <- ANON36_death[, -c(23)]

write.csv(ANON36_death_output, "/Users/Student/Desktop/ANON36_death_output.csv", row.names = FALSE)

```