JACC study Milk intake and stroke mortality analysis

Chaochen Wang

2019-12-20 created, 2020-01-09 updated

Contents

1 Read in the data

```
library(readr)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.2.1     v dplyr   0.8.3
## v tibble  2.1.3     v stringr 1.4.0
## v tidyr   1.0.0     v forcats 0.4.0
## v purrr   0.3.3
```

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate) # for dealing with date time data
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
```

```
##
##     date
```

```
MILK <- read_csv("../data/StrokeMilk.csv",
                     progress = show_progress(),
                     col_types = cols(.default = "c"))

MILK %>%
  filter(tr_age > 39 & tr_age < 80) %>%
  group_by(tr_sex) %>%
  summarise(n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))
```

```
## # A tibble: 2 x 3
##   tr_sex     n rel.freq
##   <chr>  <int> <chr>
## 1 1      46395 41.95%
## 2 2      64190 58.05%
```

2 delete subjects outside of age range

```
MILK_0 <- MILK %>%
  filter(tr_age > 39 & tr_age < 80)
```

3 define total stroke mortality

```
MILK_0 <- MILK_0 %>%
  mutate(Tot_Stroke = if_else(grepl("I6[0-9][0-9]|I6[0-9]",
                                    ICD10), "I60_9",
                          if_else(!is.na(ICD10), "other_death",
                                    "Alive/Censor")))
MILK_0%>%
  group_by(tr_sex, Tot_Stroke) %>%
  summarise(n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex Tot_Stroke       n rel.freq
##   <chr>  <chr>        <int> <chr>
## 1 1      Alive/Censor 31110 67.05%
## 2 1      I60_9         1825 3.93%
## 3 1      other_death  13460 29.01%
## 4 2      Alive/Censor 52347 81.55%
## 5 2      I60_9         1777 2.77%
## 6 2      other_death  10066 15.68%
```

4 define different type of stroke mortality/CVD ?

I60 Nontraumatic subarachnoid hemorrhage

I61 Nontraumatic intracerebral hemorrhage

I62 Other and unspecified nontraumatic intracranial hemorrhage

I63 Cerebral infarction

I65 Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction

I66 Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction

I67 Other cerebrovascular diseases

I68 Cerebrovascular disorders in diseases classified elsewhere

I69 Sequelae of cerebrovascular disease

```
MILK_0 <- MILK_0 %>%
  mutate(HemoStroke = if_else(grepl("I6[0-2][0-9]|I6[0-2]",
                                    ICD10), "I60_2",
                              if_else(!is.na(ICD10), "other_death",
                                      "Alive/Censor"))) %>%
  mutate(IscheStroke = if_else(grepl("I63[0-9]|I63",
                                    ICD10), "I63",
                              if_else(!is.na(ICD10), "other_death",
                                      "Alive/Censor"))) %>%
  mutate(CHD = if_else(grepl("I2[0-5][0-9]|I2[0-5]",
                                    ICD10), "I20_5",
                               if_else(!is.na(ICD10), "other_death",
                                       "Alive/Censor"))) %>%
  mutate(HeartF = if_else(grepl("I50[0-9]|I50",
                                    ICD10), "I50",
                               if_else(!is.na(ICD10), "other_death",
                                       "Alive/Censor")))


MILK_0%>%
  group_by(tr_sex, HemoStroke) %>%
  summarise(n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex HemoStroke       n rel.freq
##   <chr>  <chr>        <int> <chr>
## 1 1      Alive/Censor 31110 67.05%
## 2 1      I60_2          556 1.2%
## 3 1      other_death  14729 31.75%
## 4 2      Alive/Censor 52347 81.55%
## 5 2      I60_2          666 1.04%
## 6 2      other_death  11177 17.41%
```

```
MILK_0%>%
  group_by(tr_sex, IscheStroke) %>%
  summarise(n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex IscheStroke      n rel.freq
```

```
##   <chr>  <chr>        <int> <chr>
## 1 1      Alive/Censor 31110 67.05%
## 2 1      I63            705 1.52%
## 3 1      other_death  14580 31.43%
## 4 2      Alive/Censor 52347 81.55%
## 5 2      I63            600 0.93%
## 6 2      other_death  11243 17.52%
```

```
MILK_0%>%
  group_by(tr_sex, CHD) %>%
  summarise(n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex CHD              n rel.freq
##   <chr>  <chr>        <int> <chr>
## 1 1      Alive/Censor 31110 67.05%
## 2 1      I20_5         1003 2.16%
## 3 1      other_death  14282 30.78%
## 4 2      Alive/Censor 52347 81.55%
## 5 2      I20_5          758 1.18%
## 6 2      other_death  11085 17.27%
```

```
MILK_0%>%
  group_by(tr_sex, HeartF) %>%
  summarise(n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex HeartF           n rel.freq
##   <chr>  <chr>        <int> <chr>
## 1 1      Alive/Censor 31110 67.05%
## 2 1      I50            711 1.53%
## 3 1      other_death  14574 31.41%
## 4 2      Alive/Censor 52347 81.55%
## 5 2      I50            799 1.24%
## 6 2      other_death  11044 17.21%
```

5 Define milk intake

```
MILK_0 <- MILK_0 %>%
  mutate(Milk_fre = as.numeric(MILK)) %>%
  mutate(Milk_fre = as.factor(Milk_fre)) %>%
  mutate(Mlkfre = fct_collapse(Milk_fre,
                               Never = "1",
                               Mon1_2 = "2",
                               Wek1_2 = "3",
                               Wek3_4 = "4",
                               Daily  = "5")) %>%
  mutate(MlkLogi = fct_collapse(Mlkfre,
                                Never = "Never",
                                Drinker = c("Mon1_2", "Wek1_2", "Wek3_4", "Daily")))
```

```
## Warning: NAs introduced by coercion
```

```
MILK_0 %>%
  group_by(tr_sex, Mlkfre) %>%
  summarise(n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))
```

```
## Warning: Factor `Mlkfre` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
## # A tibble: 12 x 4
## # Groups:   tr_sex [2]
##    tr_sex Mlkfre     n rel.freq
##    <chr>  <fct>  <int> <chr>
##  1 1      Never   8961 19.31%
##  2 1      Mon1_2  3691 7.96%
##  3 1      Wek1_2  6228 13.42%
##  4 1      Wek3_4  5862 12.63%
##  5 1      Daily  17110 36.88%
##  6 1      <NA>    4543 9.79%
##  7 2      Never  10960 17.07%
##  8 2      Mon1_2  3830 5.97%
##  9 2      Wek1_2  7975 12.42%
## 10 2      Wek3_4  8516 13.27%
## 11 2      Daily  26957 42%
## 12 2      <NA>    5952 9.27%
```

```
MILK_0 %>%
  group_by(tr_sex, MlkLogi) %>%
  summarise(n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))
```

```
## Warning: Factor `MlkLogi` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex MlkLogi     n rel.freq
##   <chr>  <fct>   <int> <chr>
## 1 1      Never    8961 19.31%
## 2 1      Drinker 32891 70.89%
## 3 1      <NA>     4543 9.79%
## 4 2      Never   10960 17.07%
## 5 2      Drinker 47278 73.65%
## 6 2      <NA>     5952 9.27%
```

6 Calculate person-years

```
MILK_0 <- MILK_0 %>%
  mutate(Age = as.numeric(tr_age)) %>%
  mutate(followpy = as.numeric(actual)/365.25)
```

7 Identify potential confounders: smoking, alcohol intake, BMI, DM/HYT/MI/APO/Cancer history, Exercise, Energy intake, Sleep duration, vegetable/fru/gretea/cofe intake, school education

```
MILK_0 <-  MILK_0 %>%
  mutate(Smoking = replace_na(SM1, "unknown")) %>%
  mutate(Smoking = as_factor(Smoking)) %>%
  mutate(Smoking = fct_recode(Smoking, Never = "3", Past = "2", Current = "1")) %>%
  mutate(Smoking = factor(Smoking, levels = c("Never", "Past", "Current", "unknown"))) %>%  # Smoking
  mutate(Alc_Fre = if_else(as.numeric(DR1) >= 2, "Never or past",
                           if_else(as.numeric(DR1F) == 1, "Daily",
                                   if_else(as.numeric(DR1F) == 4, "< 1/week",
                                           if_else((as.numeric(DR1F) == 2) | (as.numeric(DR1F) == 3),
                                                   "1-4 /week", "Unknown"))))) %>%
  mutate(Alc_Fre = fct_explicit_na(Alc_Fre, na_level = "unknown")) %>%
  mutate(BMI = as.numeric(wt10)/(as.numeric(ht10)^2) * 100000) %>% # define BMI groups
  mutate(BMIgrp = cut(BMI, breaks = c(14, 18.5, 25, 30, 40), right = FALSE)) %>%
  mutate(BMIgrp = as.character(BMIgrp)) %>%
  replace_na(list(BMIgrp = "unknown")) %>%
  mutate(BMIgrp = factor(BMIgrp, levels = c("[18.5,25)",
                                            "[14,18.5)",
                                            "[25,30)",
                                            "[30,40)", "unknown"))) %>%
  mutate(DM_hist = if_else(as.numeric(p_DM) > 1, TRUE, FALSE)) %>%
  replace_na(list(DM_hist = "unknown")) %>% # recode DM history status
  mutate(HT_hist = if_else(as.numeric(p_HT) > 1, TRUE, FALSE)) %>%
  replace_na(list(HT_hist = "unknown")) %>% # recode hyt history status
  mutate(MI_hist = if_else(as.numeric(p_MI) > 1, TRUE, FALSE)) %>%
  replace_na(list(MI_hist = "unknown")) %>% # recode MI history status
  mutate(APO_hist = if_else(as.numeric(p_APO) > 1, TRUE, FALSE)) %>%
  replace_na(list(APO_hist = "unknown")) %>% # recode APO history status
  mutate(KID_hist = if_else(as.numeric(p_KID) > 1, TRUE, FALSE)) %>%
  replace_na(list(KID_hist = "unknown")) %>% # recode KID history status
  mutate(LIV_hist = if_else(as.numeric(p_APO) > 1, TRUE, FALSE)) %>%
  replace_na(list(LIV_hist = "unknown")) %>% # recode LIV history status
  mutate(Can_hist = if_else(as.numeric(p_can1) > 1 |
                            as.numeric(p_can2) > 1, TRUE, FALSE)) %>%
  replace_na(list(Can_hist = "unknown")) %>% # recode LIV history status
  mutate(Exercise = as.numeric(sport) != 4) %>% # define exercise habits
  mutate(Exercise = as.character(Exercise)) %>%
  replace_na(list(Exercise = "unknown")) %>%
  mutate(Exercise = factor(Exercise, levels = c("FALSE", "TRUE", "unknown"))) %>%
  mutate(Exercise = fct_recode(Exercise,
                               "> 1h/w" = "TRUE",
                               "Almost0" = "FALSE",
                               unknown   = "unknown")) %>%
  mutate(Engy = log(as.numeric(ENERGY))) %>%
  mutate(Sleep = as.numeric(SLEEP)/10) %>%
  mutate(Slepgrp = cut(Sleep, breaks = c(0, 6.9, 7.9, 8.9, 23), right = FALSE)) %>%
  mutate(Slepgrp = as.character(Slepgrp)) %>%
  replace_na(list(Slepgrp = "unknown")) %>%
```

```
  mutate(Slepgrp = factor(Slepgrp, levels = c("[0,6.9)",
                                              "[6.9,7.9)",
                                              "[7.9,8.9)",
                                              "[8.9,23)", "unknown"))) %>%
  mutate(Spi = as.factor(SPI)) %>% # define vegetable intake
  mutate(Spi = fct_collapse(Spi,
                            unknown = "X",
                            daily = "5",
                            Thre4tw = "4",
                            One2tw = "3",
                            Less1tm = c("1", "2"))) %>%
  mutate(Spi = fct_explicit_na(Spi, na_level = "unknown")) %>%
  mutate(Fru = as.factor(FRU)) %>% # define fruit intake
  mutate(Fru = fct_collapse(Fru,
                            unknown = "X",
                            daily = "5",
                            Thre4tw = "4",
                            One2tw = "3",
                            Less1tm = c("1", "2"))) %>%
  mutate(Fru = fct_explicit_na(Fru, na_level = "unknown")) %>%
  mutate(Gretea = as.factor(GreTEA1)) %>% # define greentea intake
  mutate(Gretea = fct_collapse(Gretea,
                               unknown = "X",
                               Thre3tw = "2",
                               Thre3tw = "3",
                               Thre3tw = "4",
                               Never = "5",
                               daily = "1")) %>%
  mutate(Gretea = fct_explicit_na(Gretea, na_level = "unknown"))  %>%
  mutate(Cofe = as.factor(COFE)) %>% # define greentea intake
  mutate(Cofe = fct_collapse(Cofe,
                               unknown = "X",
                               Thre3tw = "2",
                               Thre3tw = "3",
                               Thre3tw = "4",
                               Never = "5",
                               daily = "1")) %>%
  mutate(Cofe = fct_explicit_na(Cofe, na_level = "unknown"))  %>%
  mutate(Educ = as.numeric(MILK_0$SCHOOL)) %>%
  mutate(Educgrp = cut(Educ, breaks = c(0, 18, 70), right = FALSE)) %>%
  mutate(Educgrp = as.character(Educgrp)) %>%
  replace_na(list(Educgrp = "unknown")) %>%
  mutate(Educgrp = factor(Educgrp, levels = c("[0,18)",
                                              "[18,70)",
                                              "unknown")))
```

```
## Warning in if_else(as.numeric(DR1F) == 1, "Daily", if_else(as.numeric(DR1F) == :
## NAs introduced by coercion
```

```
## Warning in if_else(as.numeric(DR1F) == 4, "< 1/week", if_else((as.numeric(DR1F)
## == : NAs introduced by coercion
```

```
## Warning in if_else((as.numeric(DR1F) == 2) | (as.numeric(DR1F) == 3), "1-4 /
## week", : NAs introduced by coercion
```

```
## Warning in if_else((as.numeric(DR1F) == 2) | (as.numeric(DR1F) == 3), "1-4 /
## week", : NAs introduced by coercion
```

```
## Warning in if_else(as.numeric(p_KID) > 1, TRUE, FALSE): NAs introduced by
## coercion
```

```
## Warning in if_else(as.numeric(p_can1) > 1 | as.numeric(p_can2) > 1, TRUE, : NAs
## introduced by coercion
```

```
## Warning in if_else(as.numeric(p_can1) > 1 | as.numeric(p_can2) > 1, TRUE, : NAs
## introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
MILK_0 %>%
  group_by(tr_sex, Smoking) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 8 x 4
## # Groups:   tr_sex [2]
##   tr_sex Smoking     n rel.freq
##   <chr>  <fct>   <int> <chr>
## 1 1      Never    9027 19.46%
## 2 1      Past    11668 25.15%
## 3 1      Current 23444 50.53%
## 4 1      unknown  2256 4.86%
## 5 2      Never   51457 80.16%
## 6 2      Past      963 1.5%
## 7 2      Current  3066 4.78%
## 8 2      unknown  8704 13.56%
```

```
MILK_0 %>%
  group_by(tr_sex, Alc_Fre) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 10 x 4
## # Groups:   tr_sex [2]
##    tr_sex Alc_Fre           n rel.freq
##    <chr>  <fct>         <int> <chr>
##  1 1      < 1/week       2027 4.37%
##  2 1      1-4 /week      7251 15.63%
##  3 1      Daily         22178 47.8%
##  4 1      Never or past 11118 23.96%
##  5 1      unknown        3821 8.24%
##  6 2      < 1/week       4106 6.4%
##  7 2      1-4 /week      6142 9.57%
##  8 2      Daily          2901 4.52%
##  9 2      Never or past 43908 68.4%
## 10 2      unknown        7133 11.11%
```

```r
MILK_0 %>%
  group_by(tr_sex, BMIgrp) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 10 x 4
## # Groups:   tr_sex [2]
##    tr_sex BMIgrp        n rel.freq
##    <chr>  <fct>     <int> <chr>
##  1 1      [18.5,25) 33340 71.86%
##  2 1      [14,18.5)  2443 5.27%
##  3 1      [25,30)    7670 16.53%
##  4 1      [30,40)     451 0.97%
##  5 1      unknown    2491 5.37%
##  6 2      [18.5,25) 42523 66.25%
##  7 2      [14,18.5)  3774 5.88%
##  8 2      [25,30)   12391 19.3%
##  9 2      [30,40)    1271 1.98%
## 10 2      unknown    4231 6.59%
```

```r
MILK_0 %>%
  group_by(tr_sex, DM_hist) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex DM_hist     n rel.freq
##   <chr>  <chr>   <int> <chr>
## 1 1      FALSE   37631 81.11%
## 2 1      TRUE     2879 6.21%
## 3 1      unknown  5885 12.68%
## 4 2      FALSE   53167 82.83%
## 5 2      TRUE     2404 3.75%
## 6 2      unknown  8619 13.43%
```

```r
MILK_0 %>%
  group_by(tr_sex, HT_hist) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex HT_hist     n rel.freq
##   <chr>  <chr>   <int> <chr>
## 1 1      FALSE   32476 70%
## 2 1      TRUE     8990 19.38%
## 3 1      unknown  4929 10.62%
## 4 2      FALSE   43772 68.19%
## 5 2      TRUE    13541 21.1%
## 6 2      unknown  6877 10.71%
```

```r
MILK_0 %>%
  group_by(tr_sex, MI_hist) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex MI_hist     n rel.freq
##   <chr>  <chr>   <int> <chr>
## 1 1      FALSE   39063 84.2%
## 2 1      TRUE     1310 2.82%
## 3 1      unknown  6022 12.98%
## 4 2      FALSE   53826 83.85%
## 5 2      TRUE     1684 2.62%
## 6 2      unknown  8680 13.52%
```

```r
MILK_0 %>%
  group_by(tr_sex, APO_hist) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex APO_hist     n rel.freq
##   <chr>  <chr>    <int> <chr>
## 1 1      FALSE    39336 84.78%
## 2 1      TRUE       915 1.97%
## 3 1      unknown   6144 13.24%
## 4 2      FALSE    54642 85.13%
## 5 2      TRUE       581 0.91%
## 6 2      unknown   8967 13.97%
```

```r
MILK_0 %>%
  group_by(tr_sex, KID_hist) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex KID_hist     n rel.freq
##   <chr>  <chr>    <int> <chr>
## 1 1      FALSE    34759 74.92%
## 2 1      TRUE      1603 3.46%
## 3 1      unknown  10033 21.63%
## 4 2      FALSE    47752 74.39%
## 5 2      TRUE      2668 4.16%
## 6 2      unknown  13770 21.45%
```

```r
MILK_0 %>%
  group_by(tr_sex, LIV_hist) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex LIV_hist     n rel.freq
##   <chr>  <chr>    <int> <chr>
## 1 1      FALSE    39336 84.78%
## 2 1      TRUE       915 1.97%
## 3 1      unknown   6144 13.24%
## 4 2      FALSE    54642 85.13%
## 5 2      TRUE       581 0.91%
## 6 2      unknown   8967 13.97%
```

```
MILK_0 %>%
  group_by(tr_sex, Can_hist) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex Can_hist     n rel.freq
##   <chr>  <chr>    <int> <chr>
## 1 1      FALSE     5899 12.71%
## 2 1      TRUE       411 0.89%
## 3 1      unknown  40085 86.4%
## 4 2      FALSE     8453 13.17%
## 5 2      TRUE      1050 1.64%
## 6 2      unknown  54687 85.2%
```

```
MILK_0 %>%
  group_by(tr_sex, Exercise) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex Exercise     n rel.freq
##   <chr>  <fct>    <int> <chr>
## 1 1      Almost0  25559 55.09%
## 2 1      > 1h/w   11697 25.21%
## 3 1      unknown   9139 19.7%
## 4 2      Almost0  38842 60.51%
## 5 2      > 1h/w   12172 18.96%
## 6 2      unknown  13176 20.53%
```

```
MILK_0 %>%
  group_by(tr_sex, Slepgrp) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 10 x 4
## # Groups:   tr_sex [2]
##    tr_sex Slepgrp       n rel.freq
##    <chr>  <fct>     <int> <chr>
```

```
##  1 1      [0,6.9)    7804 16.82%
##  2 1      [6.9,7.9) 14248 30.71%
##  3 1      [7.9,8.9) 16512 35.59%
##  4 1      [8.9,23)   5384 11.6%
##  5 1      unknown    2447 5.27%
##  6 2      [0,6.9)   17064 26.58%
##  7 2      [6.9,7.9) 22008 34.29%
##  8 2      [7.9,8.9) 16749 26.09%
##  9 2      [8.9,23)   4307 6.71%
## 10 2      unknown    4062 6.33%
```

```
MILK_0 %>%
  group_by(tr_sex, Spi) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 10 x 4
## # Groups:   tr_sex [2]
##    tr_sex Spi         n rel.freq
##    <chr>  <fct>   <int> <chr>
##  1 1      Less1tm  3977 8.57%
##  2 1      One2tw  11352 24.47%
##  3 1      Thre4tw 10688 23.04%
##  4 1      daily   11008 23.73%
##  5 1      unknown  9370 20.2%
##  6 2      Less1tm  3670 5.72%
##  7 2      One2tw  14111 21.98%
##  8 2      Thre4tw 15711 24.48%
##  9 2      daily   18067 28.15%
## 10 2      unknown 12631 19.68%
```

```
MILK_0 %>%
  group_by(tr_sex, Fru) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 10 x 4
## # Groups:   tr_sex [2]
##    tr_sex Fru         n rel.freq
##    <chr>  <fct>   <int> <chr>
##  1 1      Less1tm  6511 14.03%
##  2 1      One2tw   9449 20.37%
##  3 1      Thre4tw  8221 17.72%
##  4 1      daily    9099 19.61%
##  5 1      unknown 13115 28.27%
##  6 2      Less1tm  5168 8.05%
##  7 2      One2tw   9534 14.85%
##  8 2      Thre4tw 11900 18.54%
##  9 2      daily   20390 31.77%
## 10 2      unknown 17198 26.79%
```

```
MILK_0 %>%
  group_by(tr_sex, Gretea) %>%
  summarise (n= n()) %>%
```

```
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 8 x 4
## # Groups:   tr_sex [2]
##   tr_sex Gretea      n rel.freq
##   <chr>  <fct>   <int> <chr>
## 1 1      daily   35374 76.25%
## 2 1      Thre3tw  4112 8.86%
## 3 1      Never    2765 5.96%
## 4 1      unknown  4144 8.93%
## 5 2      daily   47366 73.79%
## 6 2      Thre3tw  6185 9.64%
## 7 2      Never    4505 7.02%
## 8 2      unknown  6134 9.56%
```

```
MILK_0 %>%
  group_by(tr_sex, Cofe) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 8 x 4
## # Groups:   tr_sex [2]
##   tr_sex Cofe        n rel.freq
##   <chr>  <fct>   <int> <chr>
## 1 1      daily   21804 47%
## 2 1      Thre3tw 12264 26.43%
## 3 1      Never    9642 20.78%
## 4 1      unknown  2685 5.79%
## 5 2      daily   28693 44.7%
## 6 2      Thre3tw 16977 26.45%
## 7 2      Never   15026 23.41%
## 8 2      unknown  3494 5.44%
```

```
MILK_0 %>%
  group_by(tr_sex, Educgrp) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex Educgrp     n rel.freq
##   <chr>  <fct>   <int> <chr>
## 1 1      [0,18)  19209 41.4%
## 2 1      [18,70) 14470 31.19%
## 3 1      unknown 12716 27.41%
## 4 2      [0,18)  29683 46.24%
## 5 2      [18,70) 17917 27.91%
## 6 2      unknown 16590 25.85%
```

```
# 02-04 AREA 地区 (施設番号 + 地区番号)
# - touhoku: (1, 2, 3, 4, 17, 29)
# - kanto: (5, 6, 8, 9, 11, 13, 31)
```

```
# - chubu: (15, 18)
# - kinki: (10, 20, 21, 22, 24)
# - chugoku: (25, 26)
# - kyushiu: (27, 30)

MILK_0 <- MILK_0 %>%
  mutate(areano = as.numeric(areano)) %>%
  mutate(Area = if_else(areano %in% c(11, 22, 23, 24, 41, 30,
                                      170, 178, 179, 298, 299), "Touhoku",
                  if_else(areano %in% c(51, 61, 81, 91, 92, 93,
                                        110, 130, 311), "Kanto",
                    if_else(areano %in% c(151, 181), "Chubu",
                      if_else(areano %in% c(100, 108, 109, 201, 211, 212, 213,
                                214, 221, 241, 242, 243), "Kinki",
                        if_else(areano %in% c(250, 261), "Chugoku",
                          if_else(areano %in% c(271, 272, 273, 274, 300, 301, 302, 303, 304,
                            305, 306, 307, 308, 309), "Kyushiu", "else")))))) %>%
  mutate(Area = factor(Area))
```

8 Exclusion: history of stroke, cancer, MI, angina pectoris, other ischemic heart disease (ICD9)

410-414 Ischemic Heart Disease

415-417 Diseases Of Pulmonary Circulation

420-429 Other Forms Of Heart Disease

```
MILK_0 %>%
  group_by(tr_sex, APO_hist) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex APO_hist     n rel.freq
##   <chr>  <chr>    <int> <chr>
## 1 1      FALSE    39336 84.78%
## 2 1      TRUE       915 1.97%
## 3 1      unknown   6144 13.24%
## 4 2      FALSE    54642 85.13%
## 5 2      TRUE       581 0.91%
## 6 2      unknown   8967 13.97%
```

```
MILK_0 %>%
  group_by(tr_sex, Can_hist) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex Can_hist     n rel.freq
##   <chr>  <chr>    <int> <chr>
```

```
## 1 1      FALSE     5899 12.71%
## 2 1      TRUE       411 0.89%
## 3 1      unknown  40085 86.4%
## 4 2      FALSE     8453 13.17%
## 5 2      TRUE      1050 1.64%
## 6 2      unknown  54687 85.2%
```

```
MILK_0 %>%
  group_by(tr_sex, MI_hist) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex MI_hist     n rel.freq
##   <chr>  <chr>   <int> <chr>
## 1 1      FALSE   39063 84.2%
## 2 1      TRUE     1310 2.82%
## 3 1      unknown  6022 12.98%
## 4 2      FALSE   53826 83.85%
## 5 2      TRUE     1684 2.62%
## 6 2      unknown  8680 13.52%
```

```
MILK_0 <- MILK_0 %>%
  mutate(p_Oth1 = as.numeric(p_oth1c)) %>%
  mutate(p_Oth2 = as.numeric(p_oth2c)) %>%
  mutate(IscheHeart = if_else((p_Oth1 >=410 & p_Oth1 <=414) |
                                (p_Oth2 >=410 & p_Oth2 <=414), TRUE, FALSE)) %>%
  replace_na(list(IscheHeart = "unknown")) %>% # recode IscheHeart history status
  mutate(OtheHeart = if_else((p_Oth1 >=420 & p_Oth1 <=429) |
                                (p_Oth2 >=420 & p_Oth2 <=429), TRUE, FALSE)) %>%
  replace_na(list(OtheHeart = "unknown")) #%>% # recode Otherheart history status
```

```
## Warning: NAs introduced by coercion
```

```
## Warning: NAs introduced by coercion
```

```
MILK_0 %>%
  group_by(tr_sex, IscheHeart) %>%
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex IscheHeart     n rel.freq
##   <chr>  <chr>      <int> <chr>
## 1 1      FALSE       1774 3.82%
## 2 1      TRUE          91 0.2%
## 3 1      unknown    44530 95.98%
## 4 2      FALSE       2614 4.07%
## 5 2      TRUE          95 0.15%
## 6 2      unknown    61481 95.78%
```

```
MILK_0 %>%
  group_by(tr_sex, OtheHeart) %>%
```

```
  summarise (n= n()) %>%
  mutate(rel.freq = paste0(round(100 * n/sum(n), 2), "%"))  %>%
  print(n=Inf)
```

```
## # A tibble: 6 x 4
## # Groups:   tr_sex [2]
##   tr_sex OtheHeart     n rel.freq
##   <chr>  <chr>     <int> <chr>
## 1 1      FALSE      1743 3.76%
## 2 1      TRUE        204 0.44%
## 3 1      unknown   44448 95.8%
## 4 2      FALSE      2566 4%
## 5 2      TRUE        314 0.49%
## 6 2      unknown   61310 95.51%
```

```
MData <- MILK_0 %>%
  filter(APO_hist != "TRUE" & IscheHeart != "TRUE" &
           OtheHeart != "TRUE" & Can_hist != "TRUE" & MI_hist != "TRUE" & !is.na(Mlkfre)) %>%
  select(Area, Age, tr_sex, Tot_Stroke, HemoStroke, IscheStroke, CHD, HeartF, MlkLogi,
         Mlkfre, followpy, Smoking, Alc_Fre, BMI, BMIgrp, DM_hist, HT_hist, KID_hist,
         LIV_hist, Exercise, Engy, ENERGY, Sleep, Slepgrp, Spi, Fru, Gretea, Cofe, Educ,
         Educgrp)

# data preparation done
```

8.1 before entering the analyses ordered, we need to explore by preliminary analyses

```
# Number of subjects, number of cases, person years
# by frequency

MData %>%
  group_by(Mlkfre) %>%
  summarise(pyear = sum(followpy), n = n()) %>%
  mutate_if(is.numeric, format, 2)
```

```
## # A tibble: 5 x 3
##   Mlkfre pyear    n
##   <fct>  <chr>    <chr>
## 1 Never  308925.7 18915
## 2 Mon1_2 116454.8 7162
## 3 Wek1_2 226331.5 13518
## 4 Wek3_4 232071.9 13671
## 5 Daily  671288.9 41119
```

```
epiDisplay::tabpct(MData$Mlkfre, MData$Tot_Stroke,
                   percent = "row", graph = FALSE)
```

```
##
## Row percent
##              MData$Tot_Stroke
## MData$Mlkfre  Alive/Censor  I60_9  other_death  Total
##       Never          14064    626         4225  18915
##                     (74.4)  (3.3)       (22.3)  (100)
```

```
##       Mon1_2          5647    206         1309   7162
##                      (78.8)  (2.9)       (18.3)  (100)
##       Wek1_2         10695    363         2460  13518
##                      (79.1)  (2.7)       (18.2)  (100)
##       Wek3_4         10975    349         2347  13671
##                      (80.3)  (2.6)       (17.2)  (100)
##       Daily          31692   1131         8296  41119
##                      (77.1)  (2.8)       (20.2)  (100)
```

```
#############################################################################################################
## survival object
#############################################################################################################
library(survival)
library(ggplot2)
library(survminer)
```

```
## Loading required package: ggpubr
```

```
## Loading required package: magrittr
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##     set_names
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(cowplot)
```

```
##
## ************************************************************
```

```
## Note: As of version 1.0.0, cowplot does not change the
```

```
##   default ggplot2 theme anymore. To recover the previous
```

```
##   behavior, execute:
##   theme_set(theme_cowplot())
```

```
## ************************************************************
```

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:ggpubr':
##
##     get_legend
```

```
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```
library(ggsci)

su_obj <- Surv(MData$followpy, MData$Tot_Stroke == "I60_9")
```

```
#########################################################################################################
## Kaplan-Meier plots and log rank tests for TotStroke and Milk intake  (frequency)
#########################################################################################################
```

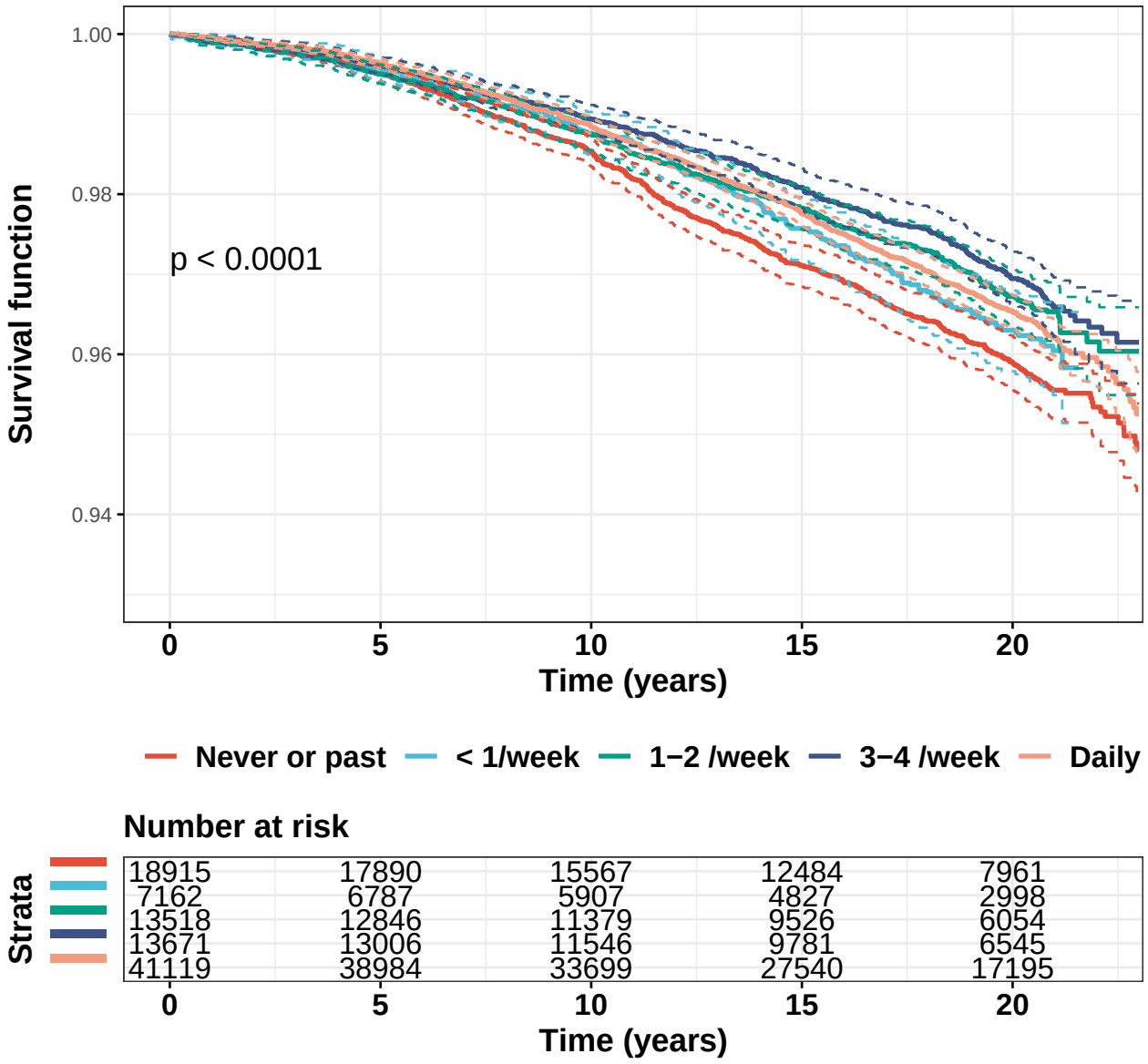

Figure 1: Kaplan-Meier survival curves for total stroke mortality by drinking frequency (P value was obtained from log-rank tests)

```
# empty.cox<-coxph(su_obj~1,data=MData)
# mgale_res<-resid(empty.cox,type="martingale")
# plot(MData$Age,mgale_res, ylim = c(-0.06, 0.01))
# lines(lowess(MData$Age,mgale_res)) # not bad
# cox1<-coxph(su_obj~ Age,data=MData)
# mgale_res<-resid(cox1,type="martingale")
# plot(MData$Age,mgale_res, ylim = c(-0.1, 0.01))
#
# # the relationship seems not linear age should be changed
```

```
#
#
# # check the proportional hazard assumption with time
# # test for interactions between the explanatory variable and time as below
#
# age.cox.tt<-coxph(su_obj~Age + tt(Age),data=MData, tt=function(x,t,...) {x*t})
# summary(age.cox.tt)
```