HMD Analysis To Calculate Premature Mortality Globally

Mathew Kiang

2/15/2018

Introduction

We want to be able to look at the county SMRs, convert them to age-standardized premature mortality rates, and compare the rates of counties with that of other countries. This file just takes the Human Mortality Database files and calculates the age-standardized premature mortality rate for a bunch of different counties over the same time period. Because not every county has 2010-2015 data (i.e., some countries end in 2013 or 2014), we will calculate whatever rate we have for that period and then take the "average" annual rate.

Set up

All countries have their own files. All files have the same structure.

Deaths

Note that for some years, the boundaries changed so there are duplicate counts of before (-) and after (+). I just remove the +/- and take the average. We also drop the female and male columns and just use total.

```
deaths_df <- bind_rows(
    map(death_files, ~read_hmd_file(.x))
) %>%
    rename(deaths = total) %>%
    mutate(year = as.integer(gsub("\\+|\\-", "", year))) %>%
    group_by(year, age, country) %>%
    mutate_all(mean) %>%
    ungroup() %>%
    select(-female, -male)

print(deaths_df, n = 10)
```

```
## # A tibble: 109,368 x 4
##     year age   deaths country
##    <int> <chr>  <dbl> <chr>
##  1  1921 0       8967 AUS
##  2  1921 1-4     3101 AUS
##  3  1921 5-9     1173 AUS
##  4  1921 10-14    799 AUS
##  5  1921 15-19   1000 AUS
##  6  1921 20-24   1375 AUS
##  7  1921 25-29   1735 AUS
##  8  1921 30-34   1948 AUS
##  9  1921 35-39   2188 AUS
## 10  1921 40-44   2161 AUS
## # ... with 1.094e+05 more rows
```

Population

Again, some years had boundary changes with - indicating before and + indicating after the boundary change for that year. We import year as a character, remove the symbol, and take the average for that year.

```
pops_df <- bind_rows(
    map(pop_files, ~read_hmd_file(.x))
) %>%
    rename(population = total) %>%
    mutate(year = as.integer(gsub("\\+|\\-", "", year))) %>%
    group_by(year, age, country) %>%
    mutate_all(mean) %>%
    ungroup() %>%
    select(-female, -male)

print(pops_df, n = 10)
```

```
## # A tibble: 111,456 x 4
##     year age   population country
##    <int> <chr>      <dbl> <chr>
##  1  1921 0          128699 AUS
##  2  1921 1-4        471964 AUS
##  3  1921 5-9        595034 AUS
##  4  1921 10-14      526404 AUS
##  5  1921 15-19      469291 AUS
##  6  1921 20-24      452418 AUS
##  7  1921 25-29      465016 AUS
##  8  1921 30-34      446717 AUS
##  9  1921 35-39      386709 AUS
## 10  1921 40-44      327825 AUS
## # ... with 1.114e+05 more rows
```

Combine them

```
hmd_data <- pops_df %>%
    left_join(deaths_df) %>%
    select(year, age, country, deaths, population)
```

```
## Joining, by = c("year", "age", "country")
```

```
print(hmd_data, n = 10)
```

```
## # A tibble: 111,456 x 5
##     year age   country deaths population
##    <int> <chr> <chr>    <dbl>      <dbl>
##  1  1921 0     AUS       8967     128699
##  2  1921 1-4   AUS       3101     471964
##  3  1921 5-9   AUS       1173     595034
##  4  1921 10-14 AUS        799     526404
##  5  1921 15-19 AUS       1000     469291
##  6  1921 20-24 AUS       1375     452418
##  7  1921 25-29 AUS       1735     465016
##  8  1921 30-34 AUS       1948     446717
##  9  1921 35-39 AUS       2188     386709
```

```
## 10  1921 40-44 AUS       2161     327825
## # ... with 1.114e+05 more rows
```

Recode age to match the standard pops

```
std_pop <- narcan::std_pops %>%
    filter(standard == "s204") %>%
    select(age, age_cat, pop_std)

print(std_pop)
```

```
## # A tibble: 18 x 3
##      age age_cat  pop_std
##    <dbl> <ord>      <int>
##  1  0    0-4     18986520
##  2  5.00 5-9     19919840
##  3 10.0  10-14   20056779
##  4 15.0  15-19   19819518
##  5 20.0  20-24   18257225
##  6 25.0  25-29   17722067
##  7 30.0  30-34   19511370
##  8 35.0  35-39   22179956
##  9 40.0  40-44   22479229
## 10 45.0  45-49   19805793
## 11 50.0  50-54   17224359
## 12 55.0  55-59   13307234
## 13 60.0  60-64   10654272
## 14 65.0  65-69    9409940
## 15 70.0  70-74    8725574
## 16 75.0  75-79    7414559
## 17 80.0  80-84    4900234
## 18 85.0  85+      4259173
```

So we need to collapse the `0` and `1-4` groups of the HMD data into one group and all the `85+` groups into one group.

```
## First, make a new column with the first number of the age category
hmd_data <- hmd_data %>%
    mutate(age_orig = age,
           age = as.integer(
               unlist(
                   lapply(
                       str_split(age_orig,
                                 pattern = "\\-|\\+"),
                       function(x) x[[1]])
                   )
               )
           )

head(hmd_data)
```

```
## # A tibble: 6 x 6
##    year   age country deaths population age_orig
##   <int> <int> <chr>    <dbl>      <dbl> <chr>
```

```
## 1  1921     0 AUS       8967     128699 0
## 2  1921     1 AUS       3101     471964 1-4
## 3  1921     5 AUS       1173     595034 5-9
## 4  1921    10 AUS        799     526404 10-14
## 5  1921    15 AUS       1000     469291 15-19
## 6  1921    20 AUS       1375     452418 20-24
```

```
## Now create a new collapsed age column
hmd_data <- hmd_data %>%
    mutate(new_age = case_when(
        age <= 1 ~ 0L,
        age >= 85 ~ 85L,
        TRUE ~ age
    ))

print(hmd_data, n = 25)
```

```
## # A tibble: 111,456 x 7
##     year   age country deaths population age_orig new_age
##    <int> <int> <chr>    <dbl>      <dbl> <chr>      <int>
##  1  1921     0 AUS     8967    128699    0              0
##  2  1921     1 AUS     3101    471964    1-4            0
##  3  1921     5 AUS     1173    595034    5-9            5
##  4  1921    10 AUS      799    526404    10-14         10
##  5  1921    15 AUS     1000    469291    15-19         15
##  6  1921    20 AUS     1375    452418    20-24         20
##  7  1921    25 AUS     1735    465016    25-29         25
##  8  1921    30 AUS     1948    446717    30-34         30
##  9  1921    35 AUS     2188    386709    35-39         35
## 10  1921    40 AUS     2161    327825    40-44         40
## 11  1921    45 AUS     2405    283018    45-49         45
## 12  1921    50 AUS     2892    252872    50-54         50
## 13  1921    55 AUS     3526    217011    55-59         55
## 14  1921    60 AUS     4135    166940    60-64         60
## 15  1921    65 AUS     3969    103756    65-69         65
## 16  1921    70 AUS     3702     64799    70-74         70
## 17  1921    75 AUS     3718     40360    75-79         75
## 18  1921    80 AUS     2817     22034    80-84         80
## 19  1921    85 AUS     1749      8902    85-89         85
## 20  1921    90 AUS      570      2371    90-94         85
## 21  1921    95 AUS      123       398    95-99         85
## 22  1921   100 AUS       23.0      63.8  100-104       85
## 23  1921   105 AUS        0         4.38 105-109       85
## 24  1921   110 AUS        0         1.00 110+          85
## 25  1922     0 AUS     7259    131945    0              0
## # ... with 1.114e+05 more rows
```

```
## Now collapse according to new age column
hmd_data <- hmd_data %>%
    group_by(country, year, new_age) %>%
    summarize(population = sum(population),
              deaths = sum(deaths)) %>%
    rename(age = new_age) %>%
    left_join(std_pop, by = "age") %>%
    ungroup()
```

```
head(hmd_data)
```

```
## # A tibble: 6 x 7
##   country  year   age population deaths age_cat  pop_std
##   <chr>   <int> <dbl>      <dbl>  <dbl> <ord>      <int>
## 1 AUS      1921  0        600663  12068 0-4     18986520
## 2 AUS      1921  5.00     595034   1173 5-9     19919840
## 3 AUS      1921 10.0      526404    799 10-14   20056779
## 4 AUS      1921 15.0      469291   1000 15-19   19819518
## 5 AUS      1921 20.0      452418   1375 20-24   18257225
## 6 AUS      1921 25.0      465016   1735 25-29   17722067
```

Calculate average premature mortality between 2010-2015

Because not every country has all observations between 2010 and 2015, we'll just calculate premature mortality for this period for all years that are observed and then take the average.

```
sub_df <- hmd_data %>%
    filter(year >= 2010,
           year <= 2015,
           age < 65)

age_specific_rates <- sub_df %>%
    narcan::calc_asrate_var(prem_death,
                            death_col = deaths,
                            pop_col = population)

head(age_specific_rates)
```

```
## # A tibble: 6 x 9
##   country  year   age population deaths age_cat  pop_std prem_death_rate
##   <chr>   <int> <dbl>      <dbl>  <dbl> <ord>      <int>           <dbl>
## 1 AUS      2010  0       1439964   1439 0-4     18986520            99.9
## 2 AUS      2010  5.00    1353199    143 5-9     19919840            10.6
## 3 AUS      2010 10.0     1385373    141 10-14   20056779            10.2
## 4 AUS      2010 15.0     1461233    509 15-19   19819518            34.8
## 5 AUS      2010 20.0     1593312    712 20-24   18257225            44.7
## 6 AUS      2010 25.0     1602488    823 25-29   17722067            51.4
## # ... with 1 more variable: prem_death_var <dbl>
```

```
age_std_rates <- age_specific_rates %>%
    group_by(country, year) %>%
    summarize(prem_std_rate = weighted.mean(prem_death_rate, pop_std)) %>%
    ungroup()
```

```
average_premature_mortality <- age_std_rates %>%
    group_by(country) %>%
    summarize(avg_prem = mean(prem_std_rate, na.rm = TRUE))
```

Add useful names

To get more useful names, we will cycle through the death files and extract the country name which should be on the first line.

```
country_name <- NULL
country_abb <- NULL

for (f in death_files) {
    cname <- strsplit(
        read_lines(f, n_max = 1),
        split = ",")[[1]][[1]]

    cabb <-  str_split(f, "\\.|//")[[1]][[3]]

    country_name <- c(country_name, cname)
    country_abb  <- c(country_abb, cabb)
}

c_map <- tibble(country_name, country_abb)
```

```
average_premature_mortality <- average_premature_mortality %>%
    left_join(c_map, by = c("country" = "country_abb")) %>%
    arrange(avg_prem)

knitr::kable(average_premature_mortality)
```

country	avg_prem	country_name
ISL	111.6077	Iceland
ITA	115.4745	Italy
CHE	115.8801	Switzerland
SWE	118.3970	Sweden
JPN	123.7484	Japan
ISR	125.6659	Israel
ESP	128.2573	Spain
AUS	128.5769	Australia
NOR	129.4965	Norway
NLD	133.3110	Netherlands
IRL	142.3838	Ireland
LUX	144.2029	Luxembourg
NZL_NP	144.4984	New Zealand
GBRCENW	145.9554	England and Wales
GBRTENW	145.9554	England and Wales
CAN	146.0141	Canada
GBR_NP	150.3870	United Kingdom
AUT	150.6851	Austria
DEUTW	152.9155	West Germany
GRC	154.3994	
DEUTNP	157.1675	Germany
DNK	159.9691	Denmark
FIN	160.8036	Finland
PRT	161.1180	Portugal
BEL	161.9363	Belgium
FRACNP	161.9381	France
FRATNP	161.9381	France (Total Population)
GBR_NIR	166.4287	Northern Ireland
SVN	167.7820	Slovenia
DEUTE	176.6837	East Germany

country	avg_prem	country_name
GBR_SCO	190.2100	Scotland
CZE	197.0062	Czech Republic
TWN	201.5856	Taiwan
HRV	211.1057	Croatia
USA	220.0430	The United States of America
SVK	252.7573	Slovakia
POL	261.4465	Poland
EST	271.0093	Estonia
HUN	302.7214	Hungary
BGR	315.8587	Bulgaria
LVA	365.8070	Latvia
LTU	378.8249	Lithuania
BLR	434.9740	Belarus
UKR	450.1296	Ukraine
RUS	510.0960	Russia

Save

```
write.csv(average_premature_mortality,
          "./data_working/hmd_countries_average_prem.csv",
          row.names = FALSE)
```