

RMarkdownAssessment

Abhinaya

2025-12-01

Aim

To examine the trend in age–sex standardised **Coronary Heart Disease hospitalisation rates** in Scotland between financial years **2016/17–2018/19** and **2021/22–2023/24** according to the 3-year rolling averages.

The target audience will be primarily the public health boards, epidemiologists, hospital administration and research scientists in the fields of cardiac medicine. The insights obtained from this study will enable them to identify patterns, increase resource allocation and create awareness regarding this disease among the public. It will also contribute to research as trends form an important analysis in any study.

The data obtained has mainly the numbers of hospitalisation and rates for the given years. This data does not allow any causal inference, as no other factors are mentioned. Also, since it is secondary data, there is no control over the quality and the methodology of obtaining it.

The strengths of the summary statistics are that it is easy to interpret, and gives a clear snapshot of the mean hospitalisation cases in the period. The limitation is that there is no data on the variation during individual years. The strength of the bar chart is that it clearly shows the difference in the rates over the years, considering there are few groups to track. The limitation is that since we are considering 3-year rolling averages, it is difficult to track individual year rates and plot it on a trend line.

Research Question

“How have age–sex standardised CHD hospitalisation rates in Scotland changed between 2016/17–2018/19 and 2021/22–2023/24 according to the 3-year rolling averages”

Load Packages

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.5.2
```

```
## Warning: package 'ggplot2' was built under R version 4.5.2
```

```
## Warning: package 'tibble' was built under R version 4.5.2
```

```
## Warning: package 'tidyr' was built under R version 4.5.2
```

```
## Warning: package 'readr' was built under R version 4.5.2
```

```
## Warning: package 'purrr' was built under R version 4.5.2

## Warning: package 'dplyr' was built under R version 4.5.2

## Warning: package 'forcats' was built under R version 4.5.2

## Warning: package 'lubridate' was built under R version 4.5.2

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.6
## v forcats    1.0.1      v stringr    1.5.2
## v ggplot2    4.0.1      v tibble     3.3.0
## v lubridate  1.9.4      v tidyr      1.3.1
## v purrr      1.2.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readr)
library(tidyr)
library(dplyr)
library(here)
```

```
## Warning: package 'here' was built under R version 4.5.2

## here() starts at C:/Users/abhih/OneDrive/Documents/GitHub/RMarkdownAssessment
```

Read in data

The dataset used for the study is the **coronary heart disease (CHD) patient hospitalisations**, in *Scotland*, for the years 2006-2022, from the Scottish Public Health Observatory Online profiles Tool.

-**Numerator**= Number of patients admitted due to coronary heart disease each year.

-**Denominator** = Total population each year.

-**Measure** = Rate of admissions per 100,000 persons in the population.

```
CHD_data <- read_csv("ScotPHO_datatab_extract_2025-11-29.csv")
```

```
## Rows: 20 Columns: 12
## -- Column specification -----
## Delimiter: ","
## chr (7): area_code, area_type, area_name, period, type_definition, indicator...
## dbl (5): year, numerator, measure, upper_confidence_interval, lower_confiden...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
glimpse (CHD_data)
```

```
## Rows: 20
## Columns: 12
## $ area_code      <chr> "S000000001", "S000000001", "S000000001", "S000~
## $ area_type      <chr> "Scotland", "Scotland", "Scotland", "Scotlan~
## $ area_name      <chr> "Scotland", "Scotland", "Scotland", "Scotlan~
## $ year           <dbl> 2003, 2004, 2005, 2006, 2007, 2008, 2009, 20~
## $ period         <chr> "2002/03 to 2004/05 financial years; 3-year ~
## $ type_definition <chr> "Age-sex standardised rate per 100,000", "Ag~
## $ indicator       <chr> "Coronary heart disease (CHD) patient hospit~
## $ numerator       <dbl> 27182.0, 26465.0, 25603.3, 24778.3, 23717.0,~
## $ measure         <dbl> 637.6, 614.5, 587.3, 561.0, 529.3, 500.4, 47~
## $ upper_confidence_interval <dbl> 645.6, 622.2, 594.8, 568.3, 536.4, 507.2, 48~
## $ lower_confidence_interval <dbl> 629.8, 606.8, 579.8, 553.8, 522.4, 493.7, 47~
## $ data_source     <chr> "Public Health Scotland (SMR01)", "Public He~
```

Tidy data

Select main columns (year, period, numerator, measure)

```
CHD_main <- CHD_data %>% select(year, period, numerator, measure) %>% filter(year>= 2017) %>% rename(Hosp_per_year = numerator, HospRate_100k = measure)
```

Renaming the variables to shorten the names in “Period” column

```
CHD_main <- CHD_main %>% mutate(period = gsub(" financial years; 3-year aggregates","", period )) %>% mutate(period = View(CHD_main))
```

Data Analysis

Summary statistics of the data

```
# To check the mean of the hospitalisation numbers and rates for the period
summary(CHD_main)
```

##	year	period	Hosp_per_year	HospRate_100k
##	Min. :2017	Length:6	Min. :17952	Min. :326.7
##	1st Qu.:2018	Class :character	1st Qu.:18253	1st Qu.:330.0
##	Median :2020	Mode :character	Median :18676	Median :346.1
##	Mean :2020		Mean :18801	Mean :349.0
##	3rd Qu.:2021		3rd Qu.:19463	3rd Qu.:366.8
##	Max. :2022		Max. :19667	Max. :376.8

Data Visualisation

Plot a bar chart

Plotting a bar chart to show the trend line of the change in hospitalisation rates over the years.

```
ggplot(CHD_main, aes(x= str_wrap(period, 10), y= HospRate_100k)) + geom_col(fill = "blue4") + labs(title =
  x = "Year (3 year rolling average)",
  y = "Rate per 100,000") + theme_classic() +
  coord_cartesian(ylim = c(310, 380))
```

