```
---
title: "Assessment"
author: "Qimiao Li"
date: "2024-11-01"
output: html_document
---
```

````
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
````

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

````
```{r}
# Install and load necessary packages
library(tidyverse)
library(dplyr)
library(ggplot2)
library(janitor)
library(gt)
library(here)
library(data.table)
library(purrr)
library(sf)
```
````

Import the data

Prescription data
Data manipulation and cleaning
````
```{r, echo=FALSE}
monthly_prescription <- list.files(path = here("data", "prescription"),
               pattern = "csv",
               full.names = TRUE)
# Read and combine all CSV files, while changing DMDCode to character
prescription_summary <- map_df(monthly_prescription, ~fread(.x) %>%
  mutate(DMDCode = as.character(DMDCode))) %>%
  clean_names()
# Check the combined data
summary(prescription_summary)
str(prescription_summary)
```
````

Health board and population data
````
```{r include=TRUE}
heathboard_list <- read_csv("https://www.opendata.nhs.scot/dataset/9f942fdb-e59e-44f5-b534-d6e17229cc7b/resource/652ff726-e676-4a20-abda-435b98dd7bdc/download/hb14_hb19.csv") %>%
  clean_names()

census_data <- read_csv(here("data", "UV103_age_health_board_census.csv"), skip = 10) %>%
  select(-...6) %>%
  rename(
        hb_name = "Health Board Area 2019",
        hb_population = Count)%>%
  # filter the data so that we get the population of the entire health board
  filter(Age == "All people" & Sex == "All people") %>%
  # select only the relevant columns
  select(hb_name, hb_population) %>%
  # match the prescription data
  mutate(hb_name = paste("NHS", hb_name))
```
````

```
```

Focus on the Glucagon-like peptide-1 (GLP-1) analogues medication
```{r}
joined_hb_data <- prescription_summary%>%
  full_join(heathboard_list, join_by(hbt == hb))

prescribed_GLP1RA <- joined_hb_data %>%
  filter(str_detect(bnf_item_description,
"OZEMPIC|WEGOVY|SEMAGLUTIDE|TRULICITY|SAXENDA|RYBELSUS|VICTOZA|BAYETTA|BYDUREON|LYXUMIA"))
prescribed_GLP1RA_summary <- prescribed_GLP1RA %>%
  group_by(bnf_item_description) %>%
  summarise(total_prescription = sum(paid_quantity,na.rm = TRUE),
            total_paid_items = sum(number_of_paid_items, na.rm = TRUE))

```

```{r}
most_prescribed <- prescribed_GLP1RA_summary %>%
  slice_max(total_prescription , n = 10) %>%
  gt() %>%
# Set table title and subtitle
  tab_header(
    title = "Top 10 Most Prescribed GLP-1 Agonist Medications",
    subtitle = "Data from NHS Scotland"
  ) %>%
  cols_label(bnf_item_description = "Medication Brand Name",
             total_prescription = "Total Prescription",
             total_paid_items = "Number of Paid Items") %>%
  # Format the 'total_prescription' column with commas
  fmt_number(
    columns = c(total_prescription),
    decimals = 0,
    use_seps = TRUE
  ) %>%

  # Add styling to the header for emphasis
  tab_style(
    style = list(
      cell_text(weight = "bold", color = "white"),
      cell_fill(color = "#2C3E50") # Dark background for the header
    ),
    locations = cells_title(groups = "title")
  ) %>%

  # Add background color to rows for visual clarity
  data_color(
    columns = c(total_prescription),
    colors = scales::col_numeric(
      palette = c("#f7fbff", "#2171b5"),
      domain = NULL
    )
  ) %>%

  # Adjust column alignment
  cols_align(
    align = "center",
    columns = everything()
  ) %>%
  opt_row_striping()
most_prescribed

```

Classfication and Adjustment for Defined Daily Dose Index
```{r}
prescription_classification <- prescribed_GLP1RA %>%
  mutate(medication_type = case_when(
      grepl("Ozempic|Wegovy|Semaglutide", bnf_item_description, ignore.case = TRUE) ~
"Semaglutide",
```

```r
    grepl("Rybelsus", bnf_item_description, ignore.case = TRUE) ~ "Oral semaglutide",
    grepl("Saxenda|Victoza", bnf_item_description, ignore.case = TRUE) ~
"Liraglutide",
    grepl("Trulicity", bnf_item_description, ignore.case = TRUE) ~ "Dulaglutide",
    grepl("Bydureon|Bayetta", bnf_item_description, ignore.case = TRUE) ~ "Exenatide",
    grepl("Lyxumia", bnf_item_description, ignore.case = TRUE) ~ "Lixisenatide",
    TRUE ~ "Other"
  ))

prescription_dosage <- prescription_classification %>%
  mutate(
    dosage_per_unit_mg = case_when(
      str_detect(bnf_item_description, "MG") ~
as.numeric(str_extract(bnf_item_description, "\\d+(\\.\\d+)?(?=MG)")),
      str_detect(bnf_item_description, "MICROGRAMS") ~
as.numeric(str_extract(bnf_item_description, "\\d+(\\.\\d+)?(?=MICROGRAMS)")) / 1000,
      TRUE ~ NA
    )) %>%
  mutate(total_quantity_dispensed_mg = number_of_paid_items * dosage_per_unit_mg
  )

# Join with the DDD reference to normalize by DDD
  # Read the DDD index CSV file
DDD_index <- read_csv(here("data", "DDD_GLP-1_analogues.csv"))
prescriptions_with_ddd <- prescription_dosage %>%
  full_join(DDD_index, join_by (medication_type == Name)) %>%
  mutate(number_of_DDDs_dispensed = total_quantity_dispensed_mg / DDD) %>%
  group_by(medication_type) %>%
  summarise (adjusted_prescription_by_DDDs = sum(number_of_DDDs_dispensed)) %>%
  arrange(desc(adjusted_prescription_by_DDDs))

```


```{r}
#Calculate the number of paid items per 10,000 patients in each health board.
adjusted_hb_prescription <- prescription_classification %>%
  group_by(hb_name, medication_type, paid_date_month) %>%
  summarise(total_paid_items = sum(number_of_paid_items, na.rm = TRUE)) %>%
  left_join(census_data, by = "hb_name") %>%
  mutate(prescriptions_per_10000 = (total_paid_items / hb_population) * 10000) %>%
  arrange(-prescriptions_per_10000)

adjusted_hb_prescription_Oral <- adjusted_hb_prescription %>%
  filter(str_detect(medication_type, "Oral"))

```


```{r}
# load the Healthboard Shapefile
NHS_healthboards <- st_read(here("data", "NHS_healthboards_2019.shp")) %>%
  mutate(HBName = paste("NHS", HBName))
prescription_hb<- NHS_healthboards %>%
  full_join(adjusted_hb_prescription_Oral, join_by(HBName == hb_name))%>%
   mutate(month = as.factor(paid_date_month))

map_Oral_per_10k <- prescription_hb %>%
  ggplot(aes(fill = prescriptions_per_10000)) +
  geom_sf(size = 0.1, colour = "black") +
   scale_fill_distiller(
     palette = "Greens",
     direction = 1,
     name = "Oralsemaglutide\nPrescriptions\nper 10,000"
   ) +
   labs(
     title = "Oral Semaglutide Prescriptions per 10,000 People by Health Board",
     subtitle = "Scotland",
     caption = "Data Source: NHS Scotland"
   ) +
   theme_minimal() +
   theme(
```

```
    legend.position = "right",
    plot.title = element_text(hjust = 0.5, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5)
  )+
  facet_wrap(~ month, ncol = 3)
map_Oral_per_10k
```