

Assessment

Valentina Levi

2024-11-01

Investigation Question: Do regions in Northern Scotland, which experience shorter winter daylight hours, show higher seasonal increases in antidepressant prescriptions compared to Southern regions as a result of SAD?

Introduction Seasonal Affective Disorder (SAD) is a mood disorder characterised by depressive symptoms that occur at a specific time of year (usually autumn or winter) with full remission at other times of year (usually spring or summer) (Galima, Vogel and Kowalski, 2020). The key risk factors include family history, female sex and living at a more northern latitude. Although the etiology of SAD is still rather unclear, it is thought that the decreasing daylight/sunlight hours as a result of transitioning to winter triggers depressive episodes in individuals who are vulnerable to SAD (Praschak-Rieder and Willeit, 2003).

I want to assess whether the regions in Northern Scotland experience higher prescription rates of antidepressants than regions in Southern Scotland as a result of reduced daylight/sunlight causing SAD in individuals during the winter season.

The data that will be used is “Prescriptions in the Community” from NHS Public Health Scotland. I have chosen to use the year 2023 because it is the most recent complete year with all 4 seasons.

The Health Boards Area (2019) data from Scotland’s Census will be used and each healthboard will be defined as either part of the northern or southern regions in Scotland.

```
library(tidyverse)
library(janitor) # For cleaning the data
library(gt) # For tables
library(here) # For directory structure
library(readr)
library(dplyr)
library(sf)
library(viridis)
library(purrr)
library(tidyr)

#Reading in healthboards data
healthboards <- read_csv(here("data/healthboards.csv")) %>%
  clean_names()

# Define northern and southern regions based on Health Boards
northern_boards <- c("NHS Grampian", "NHS Highland", "NHS Orkney", "NHS Shetland", "NHS Western Isles")
southern_boards <- c("NHS Ayrshire and Arran", "NHS Borders", "NHS Dumfries and Galloway",
  "NHS Fife", "NHS Forth Valley", "NHS Greater Glasgow and Clyde",
```

```

      "NHS Lanarkshire", "NHS Lothian", "NHS Tayside")

# Add a Region column to the Health Board data
healthboards <- healthboards %>%
  mutate(Region = case_when(
    hb_name %in% northern_boards ~ "Northern",
    hb_name %in% southern_boards ~ "Southern",
    TRUE ~ "Other"
  ))

#Reading in prescriptions data from each month in the year 2023.
january <- read.csv(here("data/january.csv")) %>%
  clean_names()

february <- read.csv(here("data/february.csv")) %>%
  clean_names()

march <- read.csv(here("data/march.csv")) %>%
  clean_names()

april <- read.csv(here("data/april.csv")) %>%
  clean_names()

may <- read.csv(here("data/may.csv")) %>%
  clean_names()

june <- read.csv(here("data/june.csv")) %>%
  clean_names()

july <- read.csv(here("data/july.csv")) %>%
  clean_names()

august <- read.csv(here("data/august.csv")) %>%
  clean_names()

september <- read.csv(here("data/september.csv")) %>%
  clean_names()

october <- read.csv(here("data/october.csv")) %>%
  clean_names()

november <- read.csv(here("data/november.csv")) %>%
  clean_names()

december <- read.csv(here("data/december.csv")) %>%
  clean_names()

# List of monthly data frames
monthly_data <- list(
  january %>% mutate(Month = 1),
  february %>% mutate(Month = 2),
  march %>% mutate(Month = 3),
  april %>% mutate(Month = 4),

```

```

    may %>% mutate(Month = 5),
    june %>% mutate(Month = 6),
    july %>% mutate(Month = 7),
    august %>% mutate(Month = 8),
    september %>% mutate(Month = 9),
    october %>% mutate(Month = 10),
    november %>% mutate(Month = 11),
    december %>% mutate(Month = 12)
  )

# Combine datasets for each season

# Winter: December, January, February
winter <- bind_rows(
  january %>% mutate(Month = 1),
  february %>% mutate(Month = 2),
  december %>% mutate(Month = 12)
) %>% mutate(Season = "Winter")

# Spring: March, April, May
spring <- bind_rows(
  march %>% mutate(Month = 3),
  april %>% mutate(Month = 4),
  may %>% mutate(Month = 5)
) %>% mutate(Season = "Spring")

#Summer: June, July, August
summer <- bind_rows(
  june %>% mutate(Month = 6, dmd_code = as.character(dmd_code)),
  july %>% mutate(Month = 7, dmd_code = as.character(dmd_code)),
  august %>% mutate(Month = 8, dmd_code = as.character(dmd_code))
) %>% mutate(Season = "Summer")

# Autumn: September, October, November
autumn <- bind_rows(
  september %>% mutate(Month = 9),
  october %>% mutate(Month = 10),
  november %>% mutate(Month = 11)
) %>% mutate(Season = "Autumn")

# Removing the `dmd_code` column from each seasonal dataset
winter <- winter %>% select(-dmd_code)
spring <- spring %>% select(-dmd_code)
summer <- summer %>% select(-dmd_code)
autumn <- autumn %>% select(-dmd_code)

# Combine all seasonal data sets into one
combined_2023 <- bind_rows(winter, spring, summer, autumn)

#Joining the combined seasonal data to the healthboard data
joined_2023 <- full_join(healthboards,combined_2023, by = join_by(hb == hbt))

```

Loading in the data sets, creating seasons and joining healthboard data to prescription data

Filtering the data to focus on antidepressants From research online I have discovered that the typical antidepressants prescribed for people diagnosed with SAD are selective serotonin reuptake inhibitors (SSRIs). In particular, sertraline and fluoxetine (Galima, Vogel and Kowalski, 2020).

```
# Filtering out only the data with fluoxetine and sertraline in the description of the drug
antidepressant_2023 <- joined_2023 %>%
  filter(
    str_detect(bnf_item_description, "FLUOXETINE") |
    str_detect(bnf_item_description, "SERTRALINE")
  )
```

Calculating the total antidepressant prescriptions for Fluoxetine and Sertraline by health-board, season and region It is important to calculate the prescription (rates) per person because some healthboards have more people living in them than others.

```
seasonal_totals <- antidepressant_2023 %>%
  group_by(hb_name, Region, Season) %>%
  summarise(Total_Prescriptions = sum(paid_quantity, na.rm = TRUE)) %>%
  ungroup()

#Reading in population data
population <- read_csv(here("data/UV103_age_health_board_census.csv"), skip = 10) %>%
# Rename the last column to avoid the messy name in column 6 and to match column names with the prescription data
rename(Spare = "...6",
  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) %>%
  # Change health board names so they match the prescription data
  mutate(hb_name = paste("NHS", hb_name))

#Joining population data and the total antidepressant prescriptions for Fluoxetine and Sertraline by healthboard
joined_data <- full_join(population,seasonal_totals, by = join_by(hb_name))

# Calculate the prescription rate per person by healthboard, season and region.
prescription_rates <- joined_data %>%
  mutate(prescriptions_per_person = Total_Prescriptions / hb_population)
```

Creating Mapped Graph to present the data This plot gives us a more general overview of the differences in antidepressant prescription rates across the northern and southern healthboards in Scotland.

```
# Load spatial data and standardise the column name
NHS_healthboards <- st_read("~/Desktop/data_science/B208015/data/NHS_healthboards_2019.shp") %>%
  rename(hb_name = HBName) %>%
  mutate(hb_name = paste("NHS", hb_name))
```

```
## Reading layer 'NHS_healthboards_2019' from data source
##   '/Users/valentialevi/Desktop/data_science/B208015/data/NHS_healthboards_2019.shp'
##   using driver 'ESRI Shapefile'
```

```

## Simple feature collection with 14 features and 4 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: 7564.996 ymin: 530635.8 xmax: 468754.8 ymax: 1218625
## Projected CRS:  OSGB36 / British National Grid

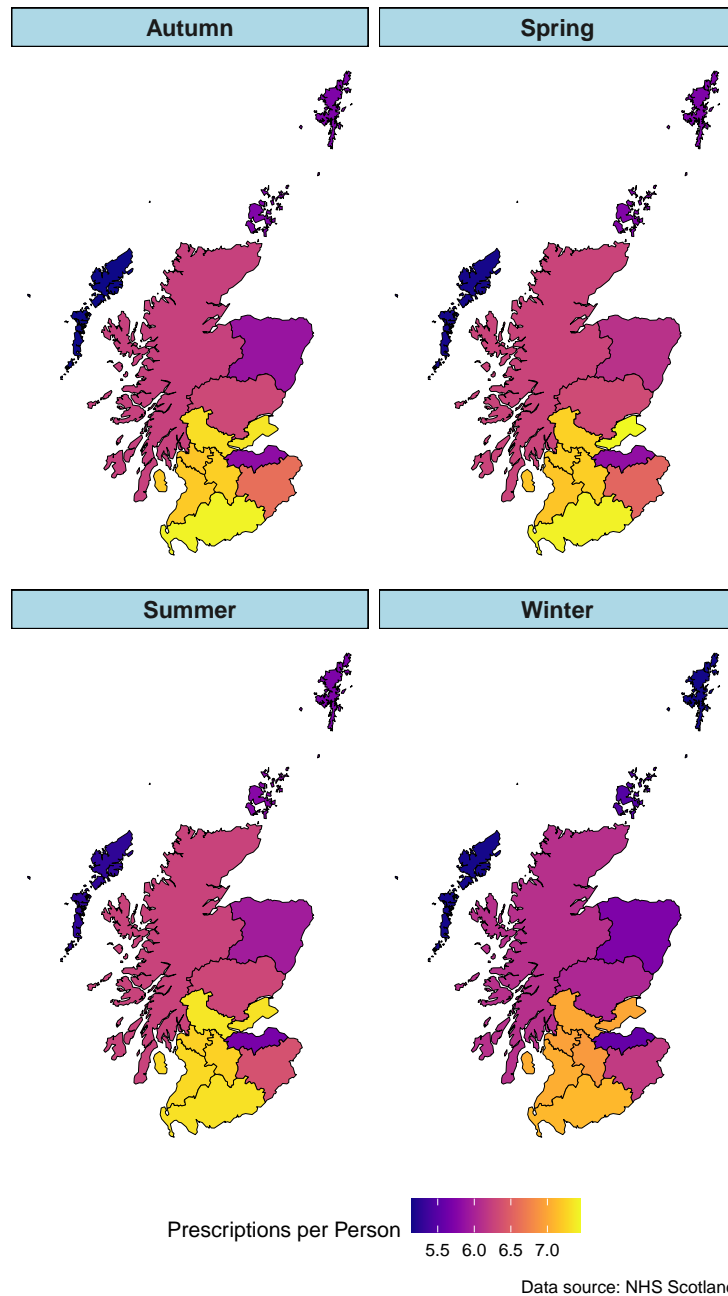
# Join data
prescription_rates_map <- NHS_healthboards %>%
  left_join(prescription_rates, by = "hb_name")

# Map Plot
map_plot <- ggplot(data = prescription_rates_map) +
  geom_sf(aes(fill = prescriptions_per_person), color = "black", linewidth = 0.2) +
  scale_fill_viridis(name = "Prescriptions per Person", option = "C", labels = scales::comma) +
  facet_wrap(~ Season, ncol = 2) +
  theme_minimal() +
  labs(
    title = "Seasonal Antidepressant Prescription Rates per\nPerson by Health Board",
    subtitle = "Comparison of Northern and Southern Scotland",
    caption = "Data source: NHS Scotland"
  ) +
  theme(
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank(),
    strip.background = element_rect(fill = "lightblue", color = "black"),
    strip.text = element_text(size = 12, face = "bold"),
    legend.position = "bottom",
  )

# Display the plot
print(map_plot)

```

Seasonal Antidepressant Prescription Rates per
Person by Health Board
Comparison of Northern and Southern Scotland



Creating two tables (northern region and southern region) to show the average prescription per person for each healthboard across the 4 seasons. Here I have included the average sunlight hours from the MET office data. I have used the North Scotland data for the northern region and combined the East and West of Southern Scotland for the southern region.

```
# Function to process Met Office sunshine data
process_sunshine <- function(url, year_filter = 2023) {
  read_table(url, skip = 5) %>%
```

```

    filter(year == year_filter) %>%
    pivot_longer(
      cols = jan:dec,
      names_to = "month",
      values_to = "sunshine_hours"
    ) %>%
    mutate(
      month = factor(month, levels = c("jan", "feb", "mar", "apr", "may", "jun", "jul", "aug", "sep", "oct", "nov", "dec"),
      season = case_when(
        month %in% c("dec", "jan", "feb") ~ "Winter",
        month %in% c("mar", "apr", "may") ~ "Spring",
        month %in% c("jun", "jul", "aug") ~ "Summer",
        month %in% c("sep", "oct", "nov") ~ "Autumn",
      )
    )
  }

# Function to generate a GT table
gt_table <- function(prescription_data, title) {
  prescription_data %>%
    select(hb_name, Season, prescriptions_per_person, avg_sunlight_hours) %>%
    pivot_wider(names_from = hb_name, values_from = prescriptions_per_person) %>%
    arrange(Season) %>%
    gt() %>%
    tab_header(
      title = title,
      subtitle = "Including Average Sunlight Hours per Season"
    ) %>%
    cols_label(
      Season = "Season",
      avg_sunlight_hours = "Avg Sunlight (Hours)"
    ) %>%
    fmt_number(columns = everything(), decimals = 2)
}

# Northern Scotland Sunshine Data URL
northern_sunshine_url <- "https://www.metoffice.gov.uk/pub/data/weather/uk/climate/datasets/Sunshine/data"

# Process Northern sunshine data
northern_sunshine <- process_sunshine(northern_sunshine_url)

# Calculate average sunshine for each season in Northern Scotland
northern_seasonal_sunshine <- northern_sunshine %>%
  group_by(season) %>%
  summarise(avg_sunlight_hours = mean(sunshine_hours, na.rm = TRUE))

# Filter Northern prescription rates
northern_prescription_rates <- prescription_rates %>%
  filter(Region == "Northern") %>%
  left_join(northern_seasonal_sunshine, by = c("Season" = "season"))

# Generate GT table for Northern Scotland
northern_gt_table <- gt_table(

```

Northern Scotland NHS Healthboards Antidepressant Prescription Data per Person Including Average Sunlight Hours per Season

Season	Avg Sunlight (Hours)	NHS Grampian	NHS Highland	NHS Orkney	NHS Shetland	NHS
Autumn	76.73	5.89	6.25	5.74	5.72	
Spring	140.67	6.13	6.28	5.74	5.73	
Summer	155.43	5.95	6.26	5.78	5.72	
Winter	32.50	5.73	6.11	5.48	5.17	

```

northern_prescription_rates,
  "Northern Scotland NHS Healthboards Antidepressant Prescription Data per Person"
)

# Southern Scotland Sunshine Data URLs
east_sunshine_url <- "https://www.metoffice.gov.uk/pub/data/weather/uk/climate/datasets/Sunshine/date/S"
west_sunshine_url <- "https://www.metoffice.gov.uk/pub/data/weather/uk/climate/datasets/Sunshine/date/S"

# Process East and West sunshine data
east_sunshine <- process_sunshine(east_sunshine_url)
west_sunshine <- process_sunshine(west_sunshine_url)

# Combine East and West sunshine data and calculate seasonal averages
southern_sunshine <- east_sunshine %>%
  inner_join(west_sunshine, by = c("year", "month", "season")) %>%
  mutate(
    avg_sunshine_hours = (sunshine_hours.x + sunshine_hours.y) / 2
  ) %>%
  group_by(season) %>%
  summarise(avg_sunlight_hours = mean(avg_sunshine_hours, na.rm = TRUE))

# Filter Southern prescription rates
southern_prescription_rates <- prescription_rates %>%
  filter(Region == "Southern") %>%
  left_join(southern_sunshine, by = c("Season" = "season"))

# Generate GT table for Southern Scotland
southern_gt_table <- gt_table(
  southern_prescription_rates,
  "Southern Scotland NHS Healthboards Antidepressant Prescription Data per Person"
)

# Display GT tables
northern_gt_table

```

southern_gt_table

Line graph showing how the average prescription per person changes across the 4 seasons for each healthboard Here we can more specifically see which healthboards have the highest and lowest rates of antidepressant prescriptions.

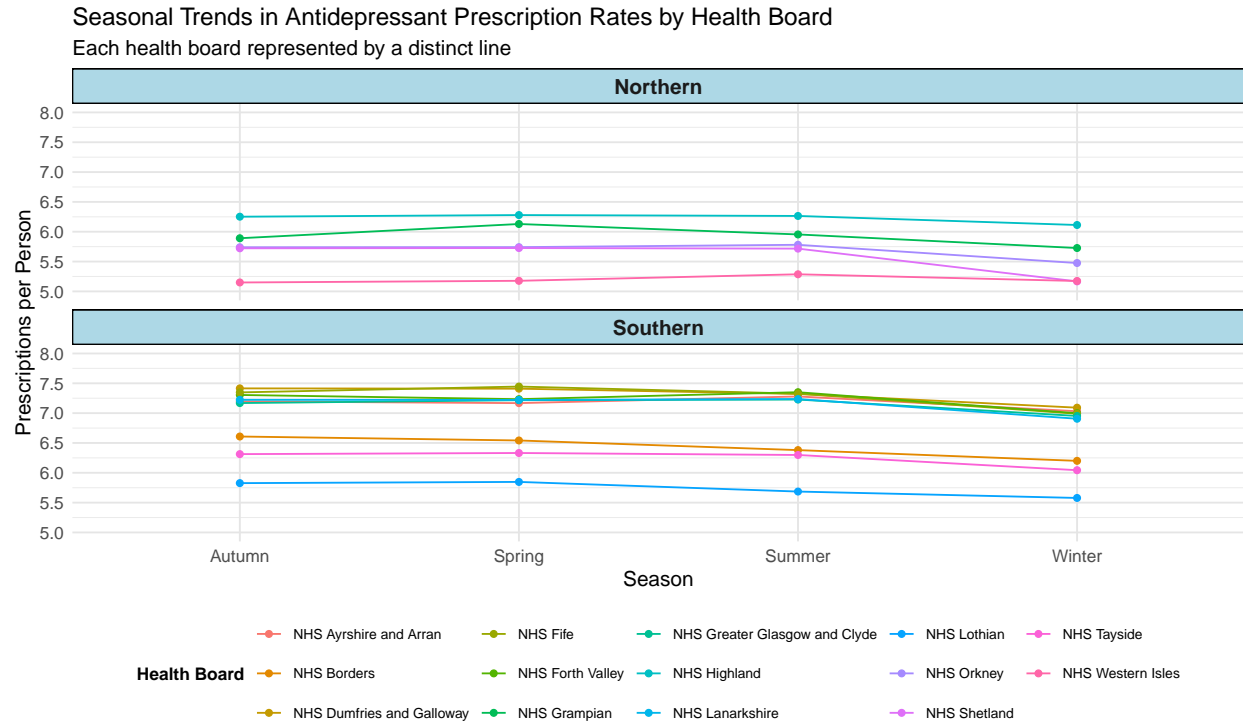
Southern Scotland NHS Healthboards Antidepressant Prescription Data per Person
Including Average Sunlight Hours per Season

Season	Avg Sunlight (Hours)	NHS Ayrshire and Arran	NHS Borders	NHS Dumfries and Galloway
Autumn	84.15	7.20	6.61	7.42
Spring	144.23	7.17	6.54	7.41
Summer	167.88	7.28	6.38	7.32
Winter	43.75	7.03	6.20	7.09

```
# Generating bright colour for each healthboard
bright_colors <- scales::hue_pal()(length(unique(prescription_rates$hb_name)))

# Line plot
line_plot <- prescription_rates %>%
  ggplot(aes(x = Season, y = prescriptions_per_person, group = hb_name, color = hb_name)) +
  geom_line(linewidth = 0.5) +
  geom_point(size = 1.5) +
  scale_color_manual(values = bright_colors, name = "Health Board") +
  scale_y_continuous(
    name = "Prescriptions per Person",
    breaks = seq(floor(min(prescription_rates$prescriptions_per_person)),
                  ceiling(max(prescription_rates$prescriptions_per_person)),
                  by = 0.5),
    limits = c(floor(min(prescription_rates$prescriptions_per_person)),
               ceiling(max(prescription_rates$prescriptions_per_person)))
  ) +
  labs(
    title = "Seasonal Trends in Antidepressant Prescription Rates by Health Board",
    subtitle = "Each health board represented by a distinct line",
    x = "Season",
    y = "Prescriptions per Person",
    color = "Health Board"
  ) +
  facet_wrap(~ Region, ncol = 1, scales = "fixed") +
  theme_minimal() +
  theme(
    legend.position = "bottom",
    legend.text = element_text(size = 8),
    legend.title = element_text(size = 10, face = "bold"),
    axis.text.x = element_text(size = 10),
    axis.text.y = element_text(size = 10),
    strip.background = element_rect(fill = "lightblue", color = "black"),
    strip.text = element_text(size = 12, face = "bold"),
    panel.grid.major = element_line(color = "gray90"),
    text = element_text(size = 12)
  )

# Display the plot
print(line_plot)
```



Analysis of Results: Interestingly the results demonstrates that antidepressant prescription rates actually tend to be higher in Southern Scotland and are actually lowest in the winter season. This reveals that although the occurrence of Seasonal Affective Disorder (SAD) is typically associated with reduced daylight during winter months, its prevalence and the prescription of antidepressants do not consistently increase with latitude. This highlights that factors beyond daylight duration (as a result of seasons), such as genetic predispositions, environmental conditions, and individual sensitivity to light changes, play significant roles in the occurrence of SAD.

Further research suggests that some people actually experience “reverse SAD,” where they feel low during the summer months instead of winter. This could be linked to disruptions in their circadian rhythm due to extended daylight hours, higher temperatures, or changes in routine. Therefore, its a complex combination of biological, psychological, and environmental factors that cause SAD to affect people differently.

Limitations: A key limitation of this research question is that it assumes that SSRI antidepressant prescriptions are all a result of SAD when in fact it is likely to be only a small proportion. It is difficult to therefore establish causation rather than just a correlation.

The reasons for higher antidepressant prescription rates could be linked to several factors rather than SAD. For example, Southern Scotland has more urban areas, particularly around cities like Glasgow and Edinburgh, where people might have greater access to mental health services and are therefore more likely to receive antidepressant prescriptions. Rural regions often have less access to mental health services, potentially lowering prescription rates.

Additionally, although northern Scotland experiences longer winter nights, factors like lifestyle, and occupational stress in the southern areas might contribute to a different type of depression, resulting in a need for antidepressants.

Another key limitation is that there are alternative therapies used to treat SAD other than just antidepressant prescriptions such as light therapy and cognitive behavior therapy (Galima, Vogel and Kowalski, 2020).

Next Steps: Analysing the other potential factors that influence antidepressant prescriptions. Exploring datasets on sex as SAD is more prevalent in females, weather patterns (e.g. rainfall), population genetics, levels of stress, access to mental health services and therefore receiving diagnosis / prescription.

References: