

Introduction to Data Health Science Report

52533106

In Scotland, how have the percentages of the population prescribed drugs for anxiety, depression and psychosis changed between 2014 and 2023 at national and local authority level?

Summary

The target audience is government and public bodies, to help inform decision making at a local and national level. The key message to convey is how much the population prescribed drugs for anxiety, depression and psychosis has changed, and how that varies regionally.

The data used were from the Scottish Public Health Observatory (ScotPHO) profiles tool. The population percentage prescribed drugs for anxiety, depression and psychosis indicator were selected for the Inequalities 'Council Area' and 'Scotland' datasets, covering the years from 2014 to 2023, and the 'total' value was used.

One of the main limitations of the data is the granularity, it has the percentage of the population, but doesn't show the actual population, so it's difficult to gauge how many people this affects. For example, the Glasgow City is the most populous authority (at 650,300 according to the Mid-2024 population estimates - National Records of Scotland (2025) whereas Orkney has a population of 22,020).

It provides very broad data and doesn't stratify by any demographic status of the population aside from the Scottish Index of Multiple Deprivation (SIMD) quintiles.

From a policy perspective not knowing the absolute numbers or demographics makes it harder to know where potential interventions would be best placed, a small increase or decrease in the number of people prescribed drugs for anxiety, depression and psychosis would have a much bigger effect on the percentage for Orkney than for Glasgow City.

The analytical process of basic statistics and differences shows that there is a variation between the local authority areas, but that they are all trending upwards.

The line charts produced show the general trend for the national and local authority levels.

The national line graph shows that there is a consistent trend of an increasing percent of the population prescribed drugs for anxiety, depression and psychosis with a dip in 2020. I think this is clear and shows what it is intended to show.

The local authorities line graph shows that there is difference between the different authorities, but it is very busy and hard to differentiate the specific local authorities. This is mitigated slightly using `plotly` to allow for interaction in the live version of the graph, though this translates less well to the static nature of a PDF.

The final bar chart shows the difference between 2014 and 2023 in the percent population prescribed drugs for anxiety, depression and psychosis. This is effective in showing that the difference is not equal between each local authority. There is a lot of data shown, but in this instance, it is readable. The chart was ordered by the percentage (decreasing) to make it easier to see and compare the differences.

Libraries

```
library("tidyverse")
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.1      v stringr   1.5.2
## v ggplot2    4.0.0      v tibble    3.3.0
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.1.0
## -- 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("viridis")
```

```
## Loading required package: viridisLite
```

```
library('plotly')
```

```
##
## Attaching package: 'plotly'
##
## The following object is masked from 'package:ggplot2':
##
##   last_plot
##
## The following object is masked from 'package:stats':
##
##   filter
##
## The following object is masked from 'package:graphics':
##
##   layout
```

Data Acquisition - Read in data files

Local and national data from the The Scottish Public Health Observatory (ScotPHO) Online Profiles Tool (2025) (website)[https://scotland.shinyapps.io/ScotPHO_profiles_tool/].

```
df_all_data <- read_csv('./data/ScotPHO_pop_drugs_adp_2025-11-03.csv')
```

```
## Rows: 2906 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (9): area_code, area_type, area_name, period, indicator, quintile, measu...
## dbl (4): year, value, upper_confidence_interval, lower_confidence_interval
##
## 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.
```

Cleaning and transforming the data

We're only actually interested in the Total data, so we'll filter on that, and run a quick summary to see what is in the data.

```
df_all_data |>
  # Filtering by the 'total' quintile, as that's what we're interested in
  filter(quintile == 'Total') |>
  # Getting the count (`n`) and the min/max values for the `value` and `year`
  # columns to see if anything is out of the expected range
  # Also checking for missing (`na`) values in the columns
  summarise(
    n = n(),
    minValue = min(value, na.rm=TRUE), maxValue = max(value, na.rm=TRUE),
    missingValue = sum(is.na(value)),
    minYear = min(year), maxYear = max(year),
    missingYear = sum(is.na(year))
  )
```

```
## # A tibble: 1 x 7
##       n minValue maxValue missingValue minYear maxYear missingYear
##   <int>   <dbl>   <dbl>         <int>   <dbl>   <dbl>         <int>
## 1    330    13.9    25.4             0    2014    2023             0
```

- There are no missing values - so no need to filter where the value `!is.na(value)`
- Year is treated as a double, which is not correct

```
df_all_data <- df_all_data |>
  mutate(year = as.integer(year))
```

Now, we're creating two data frames, one for the national level data (where the `area_type` equals "Scotland") and one for the local authority level (where the `area_type` equals "Council area").

In both the `value` used is filtered by the `quintile` column, pulling out only the "Total" values.

The minimum columns needed are also selected, at the national level that is only the `year` and `value`, at the local authority level, the `area_name` is also selected.

In both cases, `value` is renamed to `percent_prescribed` to make it clearer when it is used later on what the value is.

```
df_national_data <- df_all_data |>
  # Filtering by 'Total' values and only for Scotland as a whole
  filter(quintile == "Total" & area_type=="Scotland") |>
  # Only selecting the 'year' and 'value' columns as this is all that is needed.
  select(c("year", "value")) |>
  # Rename value column to percent_prescribed for clarity
  rename(percent_prescribed = value)

df_local_authority_data <- df_all_data |>
  # Filtering by 'Total' values and where the 'area_type' is the
  # Council Area to get Local Authority data
  filter(quintile == "Total" & area_type=="Council area") |>
  # Selecting the 'year', 'value' and 'area_name' for the local authorities
```

```
select(c("year", "area_name", "value")) |>
# Rename value column to percent_prescribed for clarity
rename(percent_prescribed = value)
```

For one of my visualisations, I want to see the difference between 2014 and 2023 in the population percent prescribed drugs for anxiety, depression and psychosis, so I filter by those two years, `pivot_wider` the data frame to allow me to add a `diff_percent` column to be rendered.

```
df_local_authority_change <- df_local_authority_data |>
# First filter by the first and last years in the data set
filter(year %in% c(2014, 2023)) |>
# Group it by the area name
group_by(area_name) |>
# Pivot the data frame wider to allow the creation of a `diff_percent` column
pivot_wider(names_from = year, values_from = percent_prescribed) |>
# Create the `diff_percent` column subtracting the 2014 value from the 2023 one
mutate(diff_percent = `2023` - `2014`)
```

I want to show the local authorities that have increased *more* than the average from the National level, so I calculate that here finding the values for 2014 and 2023 of the `percent_prescribed` and subtracting the 2014 value from the 2023 one.

```
df_national_2014_2023 = df_national_data |>
# Filter years by 2014 and 2023
filter(year %in% c(2014, 2023) ) |>
# Pivot those into columns to allow for subtraction
pivot_wider(names_from = year, values_from = percent_prescribed)

# Set a variable to be the difference
national_diff = df_national_2014_2023$`2023` - df_national_2014_2023$`2014`
```

Analysing the data

Running some basic stats for the `national` and `local_authority` data sets.

```
df_national_data |>
  summarise(
    # Count
    n = n(),
    # Mean
    mean = mean(percent_prescribed),
    # Standard Deviation
    sd = sd(percent_prescribed),
    # Inter-quartile range
    iqr = IQR(percent_prescribed)
  )
```

```
## # A tibble: 1 x 4
##       n mean   sd   iqr
##   <int> <dbl> <dbl> <dbl>
## 1    10  19.3  1.15  1.43
```

```
df_local_authority_data |>
  summarise(
    # Count
    n = n(),
    # Mean
    mean = mean(percent_prescribed),
    # Standard Deviation
    sd = sd(percent_prescribed),
    # Inter-quartile range
    iqr = IQR(percent_prescribed)
  )
```

```
## # A tibble: 1 x 4
##       n mean   sd   iqr
##   <int> <dbl> <dbl> <dbl>
## 1   320  19.3  2.47   3.8
```

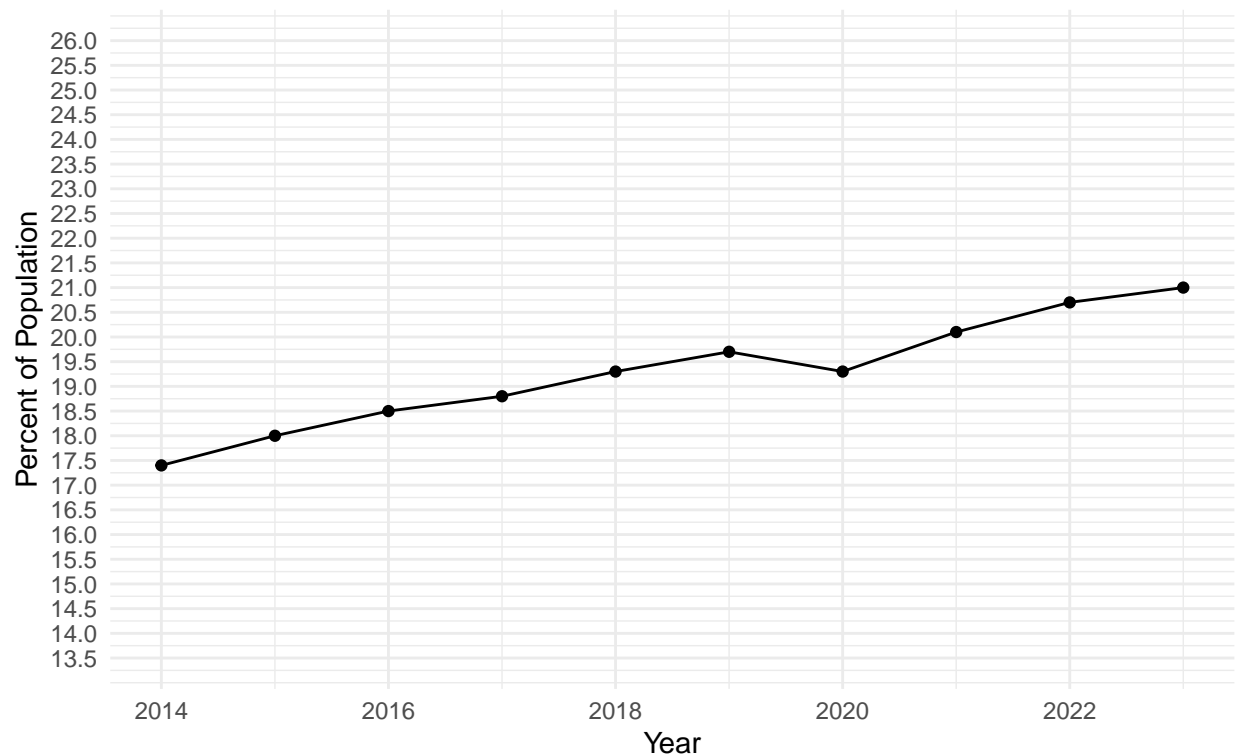
The biggest differences are the standard deviation (`sd`) and inter-quartile range (`iqr`) which shows there is a much wider range of values for the population prescribed drugs for anxiety, depression and psychosis on a local authority basis compared to national. The `mean` is the same which is what would be expected.

Visualisation

First we render a line graph of the national level data, giving a single line.

```
df_national_data |>
  # Plotting the `percent_prescribed` by `year`
  ggplot(aes(x=year, y=percent_prescribed)) +
  # Adding the points to the graph
  geom_point() +
  # Plot the line
  geom_path() +
  # Setting the limits to the same as will be used for the 'local authority' data so
  # the scales will match Also adding a tick at every 0.5 percent to make it a bit
  # clearer to read
  scale_y_continuous(limits=c(13.5,26), breaks = seq(13.5, 26, 0.5)) +
  # Setting the labels for the graph
  labs(
    title = paste(
      "Population prescribed drugs for anxiety,",
      "\ndepression and psychosis in Scotland"
    ),
    x = "Year",
    y = "Percent of Population"
  ) +
  # Using a minimal theme to make the graph less cluttered
  theme_minimal()
```

Population prescribed drugs for anxiety, depression and psychosis in Scotland



We then do the same but for the local authorities. The scale is kept the same to allow for comparison.

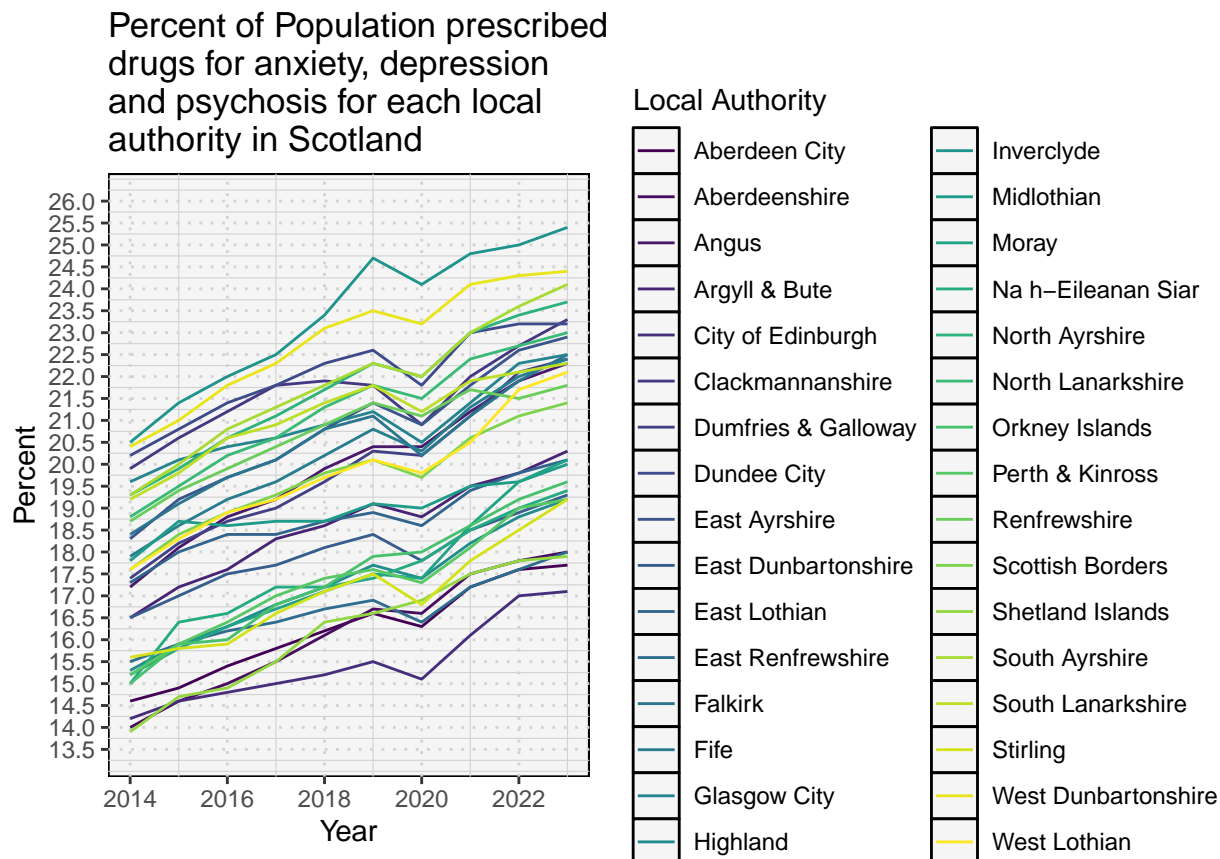
```
local_authority_plot <- df_local_authority_data |>
  # Plotting the `percent_prescribed` by the `year` using the `area_name` for the colour
  ggplot(aes(x=year, y=percent_prescribed, colour=area_name)) +
  # Plot the lines
  geom_path() +
  # From the min/max analysis above - I know
  # these ranges are outside the min/max of the data
  scale_y_continuous(limits = c(13.5,26), breaks = seq(13.5, 26, 0.5)) +
  # Setting the labels
  labs(
    title = paste(
      "Percent of Population prescribed",
      "\ndrugs for anxiety, depression",
      "\nand psychosis for each local",
      "\nauthority in Scotland"
    ),
    x = "Year",
    y = "Percent",
    colour = "Local Authority"
  ) +
  # Using the Viridis colour scale for colour blindness safety
  scale_color_viridis(discrete = TRUE) +
  # I'm setting the panel background to be a different colour to make the
  # lighter Viridis colours show up more clearly
```

```

theme(
  panel.background = element_rect(fill = '#f5f5f5ff', color = 'black'),
  panel.grid.major = element_line(color = 'lightgrey', linetype = 'dotted'),
  panel.grid.minor = element_line(color = 'lightgrey', linewidth = .2)
)

# Rendering using Plotly to allow for an easier way to see which local authority is which
#ggplotly(local_authority_plot)
local_authority_plot

```



We now render the change from the first year (2014) to the last (2023) for each of the local authorities and show the which of those authorities have changed the population prescribed drugs for anxiety, depression and psychosis more than the national average.

```

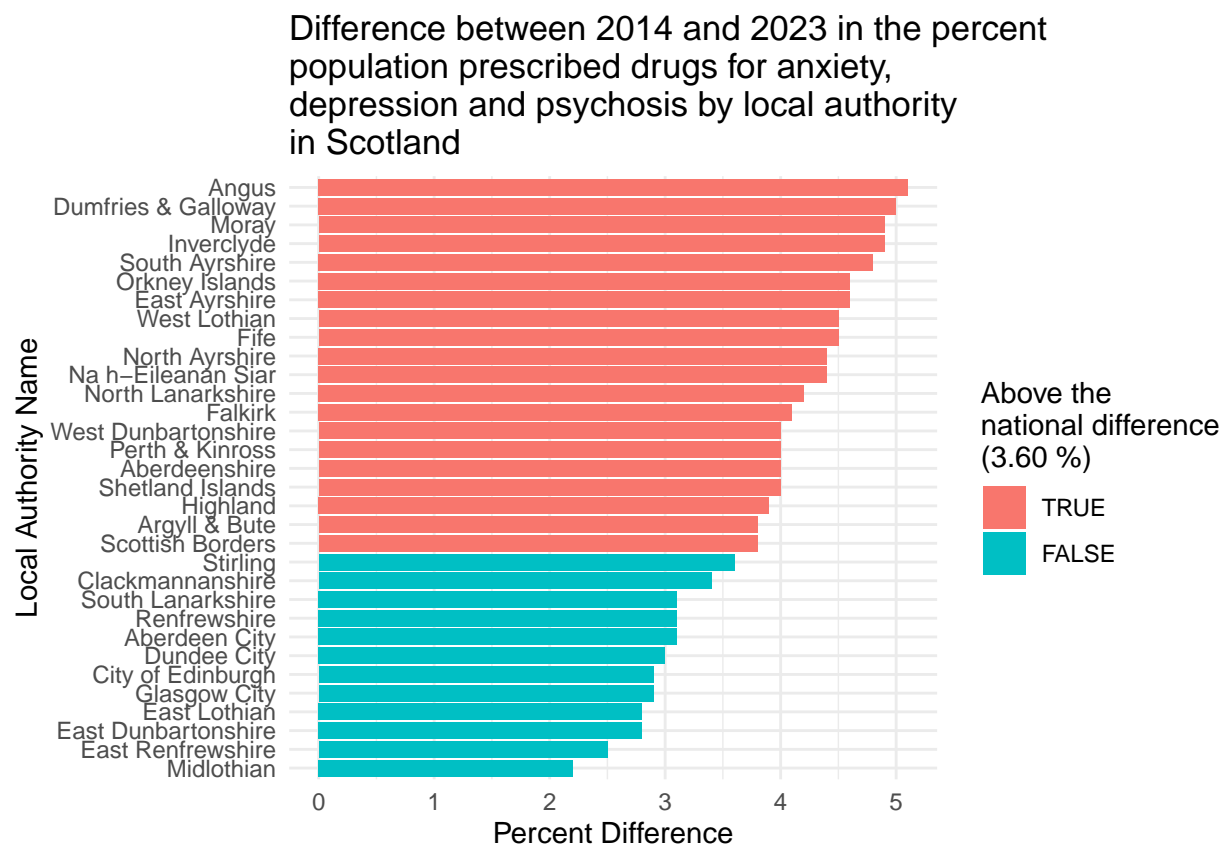
df_local_authority_change |>
  # Adding a field `isAbove` to be TRUE or FALSE if the change in percent
  # is above (bigger) than the national average.
  mutate(isAbove = diff_percent > national_diff) |>
  # Plot the percent difference by the area name. The fill is set to be
  # by the `isAbove` field calculated above
  ggplot(aes(
    x=reorder(area_name, diff_percent),
    y=diff_percent,
    fill=factor(isAbove, c(TRUE, FALSE)))
  ) +

```

```

# Rendering a bar chart to show the data
geom_bar( stat="identity" ) +
# Rendering it in a vertical form as there are
# too many areas to show in a horizontal way
coord_flip() +
# Adding the labels
labs(
  title = paste(
    "Difference between 2014 and 2023 in the percent",
    "\npopulation prescribed drugs for anxiety,",
    "\ndepression and psychosis by local authority",
    "\nin Scotland"
  ),
  x = "Local Authority Name",
  y = "Percent Difference",
  # Showing the legend, and what the national difference is (to 2 decimal places)
  fill = sprintf("Above the\nnational difference\n(%.2f %%)", national_diff)
) +
theme_minimal()

```



References

- The Scottish Public Health Observatory. Available at: https://scotland.shinyapps.io/ScotPHO_profiles_tool/ (Accessed: 3rd November 2025)

- Mid-2024 population estimates - National Records of Scotland (NRS). Available at: <https://www.nrscotland.gov.uk/publications/mid-2024-population-estimates/> (Accessed: 27th November 2025)