

Relationship between antidepressant prescriptions and education across Scotland

2025-11-04

Background

The relationship between mental health and education level has been studied extensively, with a number of theories emerging on how the two interact. Correlative trends have been observed in PhD students, possibly owing to increased stress (1.). While other studies have shown a significantly reduced incidence of mental illness in educated individuals, hypothesising a protective effect (2.).

This report aims to explore the relationship between antidepressant prescriptions and education level across Scotland, using Public Health Scotland prescribing data, and data from Scotland's most recent census in 2022. The prescribing data will be taken from January to August 2025 and a monthly average calculated, minimising any recording of repeat prescriptions following the NHS' prescribing recommendations indicating an optimal 28-day repeat prescription duration (3, 4.).

Part 1: Exploring the relationship

Prescribing data can be found here: <https://www.opendata.nhs.scot/dataset/prescriptions-in-the-community>

```
library(tidyverse)
library(stringr)
library(janitor)
library(here)
library(ggrepel)
library(readr)
library(dplyr)
library(gt)

urls <- c(
  "https://www.opendata.nhs.scot/dataset/84393984-14e9-4b0d-a797-b288db64d088/resource/cdd756fa-566e-4c",
  "https://www.opendata.nhs.scot/dataset/84393984-14e9-4b0d-a797-b288db64d088/resource/54ec3c6d-f4c5-46",
  "https://www.opendata.nhs.scot/dataset/84393984-14e9-4b0d-a797-b288db64d088/resource/9849b4dd-b130-4a",
  "https://www.opendata.nhs.scot/dataset/84393984-14e9-4b0d-a797-b288db64d088/resource/a36a1902-da40-40",
  "https://www.opendata.nhs.scot/dataset/84393984-14e9-4b0d-a797-b288db64d088/resource/a4410a3c-0e75-4b",
  "https://www.opendata.nhs.scot/dataset/84393984-14e9-4b0d-a797-b288db64d088/resource/0f2931af-3d08-4a",
  "https://www.opendata.nhs.scot/dataset/84393984-14e9-4b0d-a797-b288db64d088/resource/12448904-f11c-4a",
  "https://www.opendata.nhs.scot/dataset/84393984-14e9-4b0d-a797-b288db64d088/resource/381166dd-3a07-4c"

months <- seq(as.Date("2025-01-01"), as.Date("2025-08-01"), by = "month")

all_prescribing <- map2_df(urls, months, ~ read_csv(.x, col_types = cols(DMDCode = col_character(), .de
  mutate(month = .y))
```

First, we need to find all the antidepressants being prescribed in Scotland. This is made possible by the inclusion of BNF item codes in the Public Health Scotland prescribing data set. By identifying the relevant codes for antidepressants we can filter the data set down to just the drugs we want to look at.

In order to identify the code for antidepressants, we can look at the BNF code information provided by the NHS Business Services Authority.

BNF information available at: <https://opendata.nhsbsa.net/dataset/bnf-code-information-current-year>

```
BNF_info <- read_csv(here("data/bnf_code_current_202506_version_88.csv"))

# We filter the data to show just the antidepressants
BNF_info_antidepressants <- BNF_info %>%
  filter(BNF_SECTION == "Antidepressant drugs")

unique(BNF_info_antidepressants$BNF_SECTION_CODE)
```

```
## [1] "0403"
```

Now we can see the section code, 0403, and we could also get the codes for specific antidepressant drugs and even formulations or dosages if needed.

Let's use this information to filter the original data set, and we'll calculate the average monthly prescribing numbers to set us up for exploring the research question.

```
antidepressants <- all_prescribing %>%
  filter(str_starts(BNFItemCode, "0403"), HBT != "SB0806") # remove special health board as we filter f

hb_month <- antidepressants %>%
  group_by(HBT, month) %>%
  summarise(total_items = sum(NumberOfPaidItems, na.rm = TRUE), .groups = "drop") # summing the number

hb_avg <- hb_month %>%
  group_by(HBT) %>%
  summarise(mean_items = mean(total_items, na.rm = TRUE), .groups = "drop") # prescription averages for
```

Next, let's add some names for the healthboard to our table, to allow for clearer visualisation down the line, and we can get an idea of the range of total prescriptions across Scotland.

Health board names available at: <https://www.opendata.nhs.scot/dataset/geography-codes-and-labels/resource/652ff726-e676-4a20-abda-435b98dd7bdc>

```
hb_names <- read_csv("https://www.opendata.nhs.scot/dataset/9f942fdb-e59e-44f5-b534-d6e17229cc7b/resourc

data_antidepressants <- hb_names %>%
  inner_join(hb_avg, by = c("HB" = "HBT")) %>%
  clean_names() %>%
  select(!c(hb_date_enacted, hb_date_archived, country)) %>%
  mutate(hb_name = gsub("^NHS ", "", hb_name)) # removes NHS prefix for each health board to match the
```

Using the flexible table builder on the Scottish census website, I have created a table with the numbers for the highest level of education (columns) achieved by healthboard (rows). We can join this with the data for antidepressants, using the healthboard names we just added.

Resource is available at: <https://www.scotlandscensus.gov.uk/search-the-census#/>

```

# we have to load the census data as raw text to extract just the information we need
raw_lines <- read_lines(here("data/education_level_healthboard.csv"))
info <- raw_lines[10:25] # keep only the headers and the data (lines 10-25), also excluding the extra t
info <- info[-2] # remove the extra header line

census_data22 <- read_csv(
  I(info), # feed character vector as a connection
  col_names = TRUE, show_col_types = FALSE) %>%
  clean_names() %>%
  rename(hb = highest_level_of_qualification) %>%
  select(hb, degree_level_qualifications_or_above_education_qualifications_not_already_mentioned_includ

education_antidepressants <- data_antidepressants %>%
  full_join(census_data22, by = c("hb_name" = "hb"))

```

Because all we have right now are flat numbers, we need to show these as proportions to properly gauge the relative prescribing rate of antidepressants. Luckily, we have a “total” column from the census data that shows us the population of each healthboard, so we can create a new column with the antidepressant prescriptions per capita from it.

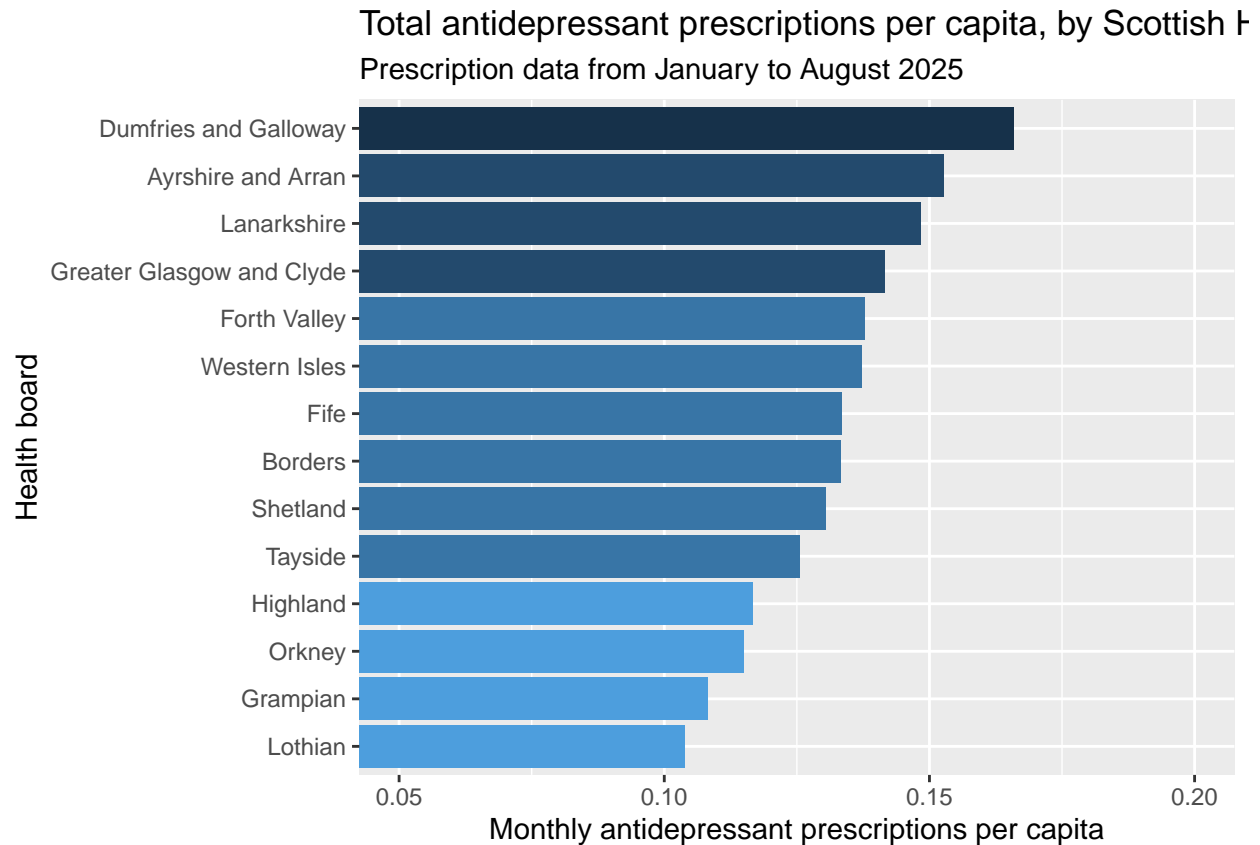
Once we’ve done that, we can investigate the prescribing rates of the health boards.

```

education_antidepressants <- education_antidepressants %>%
  mutate(ratio = mean_items / total, degree_pct = degree_level_qualifications_or_above_education_qualif

education_antidepressants %>%
  ggplot(aes(ratio, reorder(hb_name, ratio), fill = ratio)) +
  geom_col(stat="identity") +
  coord_cartesian(xlim=c(0.05, 0.2)) + # zoom in on the disparities between boards
  labs(x = "Monthly antidepressant prescriptions per capita", y = "Health board", title = "Total antidep
  scale_fill_steps(low = "#56B1F7", high = "#132B43") +
  guides(fill = "none")

```



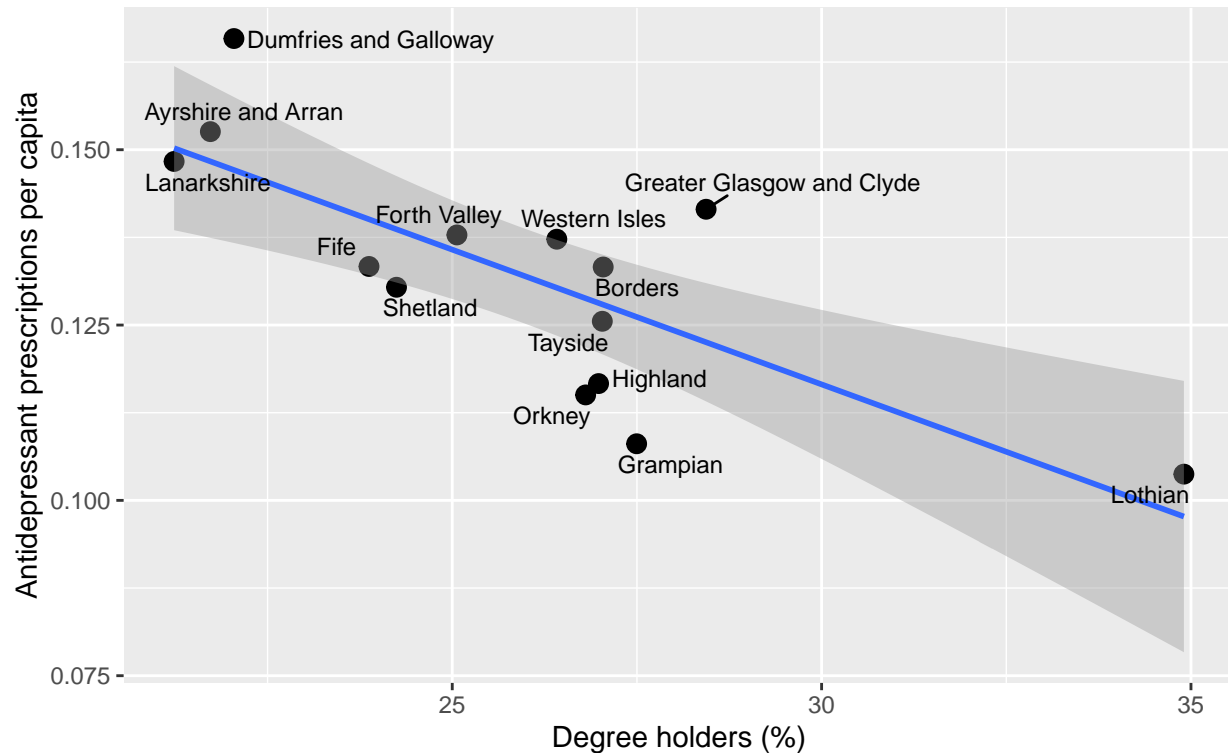
This graph shows the average monthly antidepressant prescriptions per capita from January to August 2025 in all 14 Scottish health boards. Dumfries and Galloway is shown to have the highest prescribing rate, with a relatively noticable gap between it and Ayrshire and Arran following in second, while Lothian has the lowest rate of the lot. The total range of prescriptions per capita is noticeable, with the lowest at 0.104 and the highest at 0.17, a near 70% increase. This indicates meaningful geographical variation in antidepressant prescribing across Scotland, with many factors possibly contributing to the inequalities, such as: local policy, socioeconomic disparities, and baseline characteristic differences.

Now, we can explore the relationship between these antidepressant prescription rates and the education data. We'll do this by using the percentage of each health board's population that are degree holders as calculated earlier, and running a linear regression to determine the correlation between the two variables.

```
ggplot(education_antidepressants, aes(x = degree_pct, y = ratio)) +
  geom_point(size = 3) +
  geom_smooth(method = "lm", se = TRUE) +
  geom_text_repel(aes(label = hb_name), size = 3) +
  labs(
    x = "Degree holders (%)",
    y = "Antidepressant prescriptions per capita",
    title = "Association between education level and antidepressant prescribing by Scottish Health board",
    subtitle = "Prescribing data from January to August 2025"
  )
```

Association between education level and antidepressant prescribing by Scottish Health Board

Prescribing data from January to August 2025



A surface level interpretation of this graph says that there is a strong negative correlation between the variables: health boards with a higher percentage of degree holders have a lower antidepressant prescribing rate. To take just the extremes, Dumfries and Galloway with the highest prescribing rate sits at the far left of the graph, among the bottom health boards in degree attainment, while Lothian has the lowest prescribing rate and is directly juxtaposed, with the highest degree attainment by a large margin. In isolation, this is a strong point backing the negative correlation.

However, the standard error is of moderate size, and the majority of the health boards are distributed quite close to each other in education level, while still displaying a range of prescription rates, detracting from the credibility of the suggested relationship.

It is clear there is variation in the variables, and possibly a relationship between them, but the geographical stratification is too large at the level of health boards. To try and get a clearer image of the relationship, let's do the same analysis, but with Scottish data zones, focusing in on much smaller localities.

Part 2: Focusing in

The original prescribing data already had GP practice codes, so we can map the prescriptions to data zones using a practice lookup data set from Public Health Scotland. After that we'll do the same as we did to the prescribing data before, finding the total prescriptions per month and then averaging, just this time by data zone instead of health board.

GP practices and list sizes October 2025 available from: https://www.opendata.nhs.scot/dataset/gp-practice-contact-details-and-list-sizes/resource/47557411-7eda-4278-9d6d-d26ed2ceab5a?inner_span=True

```
## load in practice details
practice_info <- read_csv("https://www.opendata.nhs.scot/dataset/f23655c3-6e23-4103-a511-a80d998adb90/r")
```

```
## Rows: 884 Columns: 15
## -- Column specification -----
## Delimiter: ","
## chr (12): GPPracticeName, AddressLine1, AddressLine2, AddressLine3, AddressL...
## dbl (2): PracticeCode, PracticeListSize
## lgl (1): Dispensing
##
## 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.
```

```
antidepressants_practice <- antidepressants %>%
  inner_join(practice_info, by = c("GPPractice" = "practice_code")) %>%
  select(c(GPPractice, NumberOfPaidItems, month, data_zone))

dz_month <- antidepressants_practice %>%
  group_by(data_zone, month) %>%
  summarise(total_items = sum(NumberOfPaidItems, na.rm = TRUE), .groups = "drop") # total prescriptions

dz_avg <- dz_month %>%
  group_by(data_zone) %>%
  summarise(mean_items = mean(total_items, na.rm = TRUE), .groups = "drop") # average monthly prescript
```

Next, we'll load in the census data set with the education levels as columns and data zones as rows. Again, we're using the flexible table builder so we'll do some pre-processing to get the data ready.

```
## load in census data by data zone instead of health board
read_lines(here("data/census_education_dz.csv"), n_max = 20)
```

```
## [1] "SuperWEB2(tm)"
## [2] ""
## [3] "\"Person\""
## [4] "\"Data Zone (2011) by Highest level of qualification\""
## [5] "\"Counting: Individuals\""
## [6] ""
## [7] "Filters:"
## [8] "\"Default Summation\", \"Individuals\""
## [9] ""
## [10] "\"Highest level of qualification\", \"Not applicable (aged less than 16) \", \"No qualifications\""
## [11] "\"Data Zone (2011)\", "
## [12] "\"S01006506\", 147,84,136,80,65,102,355,969,"
## [13] "\"S01006507\", 125,118,124,54,45,77,215,758,"
## [14] "\"S01006508\", 92,37,88,42,28,62,192,542,"
## [15] "\"S01006509\", 108,67,80,40,55,56,154,560,"
## [16] "\"S01006510\", 95,74,110,55,56,54,170,621,"
## [17] "\"S01006511\", 111,65,87,59,44,77,265,703,"
## [18] "\"S01006512\", 91,54,72,56,37,55,232,598,"
## [19] "\"S01006513\", 266,126,78,47,37,83,492,1129,"
## [20] "\"S01006514\", 226,49,90,82,34,80,479,1040,"
```

```
census_dz <- read_csv(here("data/census_education_dz.csv"), skip = 9, col_names = TRUE) %>%
  slice(-1) %>% # removing misformatted row
  remove_empty("cols") %>%
  rename(data_zone = "Highest level of qualification") %>%
```

```
filter(str_detect(data_zone, "^S\\d{8}$")) %>% # remove irrelevant rows/data zones not complicit with
clean_names()
```

```
## New names:
## * ' ' -> '...10'
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 6980 Columns: 10
## -- Column specification -----
## Delimiter: ","
## chr (1): Highest level of qualification
## dbl (8): Not applicable (aged less than 16) , No qualifications, Lower schoo...
## lgl (1): ...10
##
## 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.
```

```
antidepressants_education_dz <- inner_join(dz_avg, census_dz, by = c("data_zone" = "data_zone"))
```

```
raw_lines_dz <- read_lines(here("data/census_education_dz.csv"))
dz_rows <- grep('^"S\\d{8}"', raw_lines_dz) # use regex to find just the rows with the data we need
core_text_dz <- c(raw_lines_dz[10], raw_lines_dz[dz_rows]) # combine header row with the data
```

```
census_dz <- read_csv(
  I(core_text_dz), # character vector fed in
  show_col_types = FALSE) %>%
  clean_names() %>%
  rename(data_zone = highest_level_of_qualification) %>%
  select(data_zone, degree_level_qualifications_or_above_education_qualifications_not_already_mentioned)
```

```
## New names:
## * ' ' -> '...10'
```

```
antidepressants_education_dz <- inner_join(dz_avg, census_dz)
```

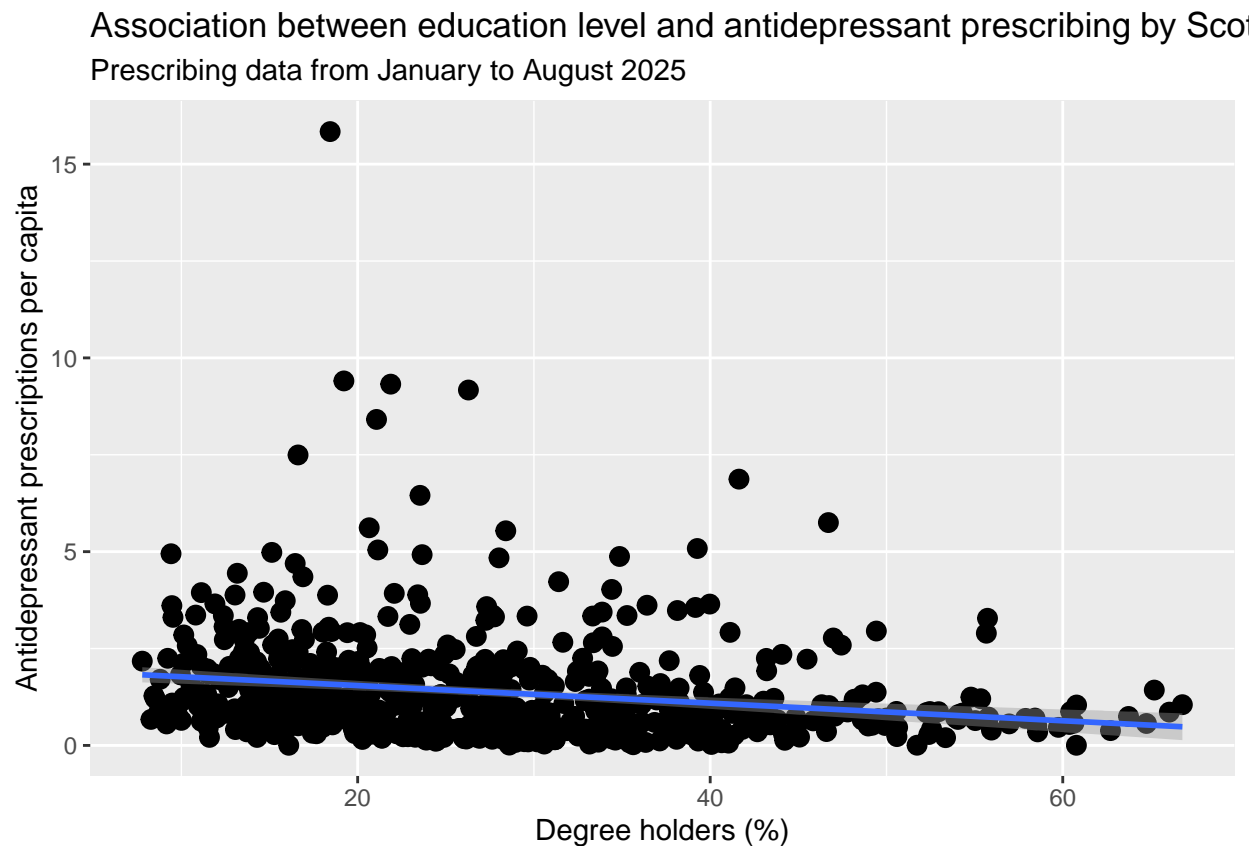
```
## Joining with 'by = join_by(data_zone)'
```

Now, we can just add the antidepressants per capita and degree holder % to the data set and we can get a similar linear regression to what we had above, with data zones this time.

```
antidepressants_education_dz <- antidepressants_education_dz %>%
  mutate(apc = mean_items/total, # antidepressants per capita
         degree_pct = (degree_level_qualifications_or_above_education_qualifications_not_already_mentioned/total))
ggplot(antidepressants_education_dz, aes(x = degree_pct, y = apc)) +
```

```
geom_point(size = 3) +
geom_smooth(method = "lm", se = TRUE) +
labs(
  x = "Degree holders (%)",
  y = "Antidepressant prescriptions per capita",
  title = "Association between education level and antidepressant prescribing by Scottish Data Zone",
  subtitle = "Prescribing data from January to August 2025"
)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



Next Steps

- table of contents with headings for each figure/visualisation
- explore dosages?
- could find geometric dataset for healthboards in scotland, then colour code the prescriptions to population ratio
- warning in census data read (line 100)
- chunk titles
- table of contents with navigation, headings to split the report up a bit(may need to revise flow somewhat)
- gt table somewhere maybe
- look over joins
- do i have to remove outlier in final graph?

- do i have to set a seed for linear regression?
- gen ai acknowledgement

References

1. Bergvall, S., Fernström, C., Ranehill, E., & Sandberg, A. (2025). The impact of PhD studies on mental health-a longitudinal population study. *Journal of health economics*, 104, 103070. Advance online publication. <https://doi.org/10.1016/j.jhealeco.2025.103070>
2. Maguire, A., Moriarty, J., O'Reilly, D. et al. (2017). Education as a predictor of antidepressant and anxiolytic medication use after bereavement: a population-based record linkage study. *Qual Life Res* 26, 1251–1262. <https://doi.org/10.1007/s11136-016-1440-1>
3. NHS Business Services Authority. Electronic Repeat Dispensing Handbook. 2020. Available at: https://www.nhsbsa.nhs.uk/sites/default/files/2020-07/Electronic%20Dispensing%20Handbook_Digital_WEB_S-1589995676.pdf [Accessed 21 Nov 2025].
4. Hertfordshire & West Essex Integrated Care Board. (n.d.) Prescription duration guidance. Available at: <https://www.hweclinicalguidance.nhs.uk/prescribing-guidance/prescription-duration-guidance-hwe-icb/> [Accessed 21 Nov 2025].