Contraceptives across Space, Age Groups and it's correlation with Neonate Mortality in Scotland

Maria Amoah

2024-11-01

This report is investigating the correlation between prescription of contraception across Scotland in 2022 and various social factors such as geographical location, age and neonate mortality. I am focusing on this issue because globally contraceptive prevalence rate is an indicator of health, women's empowerment and can be used as a proxy measure of reproductive service accessibility.(CITE). Although Scotland is devloped in blah blah blah increasing contraception has reduced the number of unintended pregnancies and infant deaths(CITE) and so this report is investigating if there is a correlation seen on a micor-level in Scotland.

the ability to decide to have fewer children . Contraceptive prevalence rate is an indicator of health, population, development and women's empowerment.Allowing individuals to be able to decide to have fewer children also has the potential to facilitate better investment in the overall health and well-being of families and communities.

```r
#Loading libraries
library(tidyverse)
library(here)
library(janitor)
library(gt)
library(sf)
library(knitr)
library(plotly)
```

Setting up the data for analysis

The Combined 2022 prescribing data comes from every month in 2022 from the Public Health Scotalnd Open Prescribing Dataset. Data on Healthboards and population data on sex and age has been loaded in here.

```r
# Loading in all the data:

# Creating a file path reduces the amount of code as there are 12 files(one per month) that need to be
prescribing_2022_files <-list.files(path = "data", pattern = "*.csv")

combined_2022_prescribing_data <- prescribing_2022_files %>%
  map_dfr(~read_csv(here("data", . ))) %>%
  clean_names()

health_boards <- read.csv("https://www.opendata.nhs.scot/dataset/9f942fdb-e59e-44f5-b534-d6e17229cc7b/r
  clean_names()

# Eliminating the archived Healthboard repeats in the dataset
```

```r
health_boards <- health_boards[-c (4,7,9,13),]

sex_data <- read.csv("https://www.opendata.nhs.scot/dataset/7f010430-6ce1-4813-b25c-f7f335bdc4dc/resour
  clean_names() %>%
  filter(year == "2022")

combined_2022_prescribing_data <- combined_2022_prescribing_data %>%
  filter(!is.na(bnf_item_code)) %>%
  filter(!is.na(bnf_item_description))

# This data was collected by Financial Year. I chose 22/23 rather than 21/22 as it contains more months
births_by_hospital <-  read.csv("https://www.opendata.nhs.scot/dataset/df10dbd4-81b3-4bfa-83ac-b14a5ec6
  filter(FinancialYear == "2022/23" )%>%
  clean_names()

hospital_and_hb <- read.csv("https://www.opendata.nhs.scot/dataset/cbd1802e-0e04-4282-88eb-d7bdcfb120f0/
  select(HospitalCode, HealthBoard) %>%
  clean_names()
```

Grouping commonly prescribed contraceptive drugs into five categories:

- Combined oral contraceptive
- Progesterone only pill
- Intrauterine contraceptives
- Contraceptive injection
- Hormone patch

The prescribed drug brand names come from NICE who list the UK brand names of drugs.(CITE)

```r
birth_control_coc <- c("MICROGYON","RIGEVIDON","OVRANETTE", "CILEST", "CILIQUE", "YASMIN", "MARVELON",

birth_control_prog <- c("NORGESTON","NORIDAY","CERAZETTE", "ZELLETA", "CERELLE", "DESOGESTROL" )

iuc <- c("MIRENA", "KYLEENA", "JAYDESS", "LEVOSERT", "BENILEXA", "COPPER T380 A", "NOVAPLUS")

# Evra is the hormone patch and Depo-provera is the contraceptive injection prescribed in the UK

all_contraceptives <- c(birth_control_coc, birth_control_prog, iuc, "EVRA", "DEPO-PROVERA")

collapsed_contraceptives <- paste(all_contraceptives, collapse = "|")

# Filtering the data set to only include contraceptive prescriptions named above

all_contraceptive_data <- combined_2022_prescribing_data %>%
  filter(str_detect(bnf_item_description,collapsed_contraceptives))
```

Figure 1 :

Investigating the difference in contraception prescription rate across the Scottish NHS Healthboards.

The use of 2019 Spatial Data structure files is appropriate as the borders used in 2022 were determined in 2013(CITE)and have not changed(CITE where i got the map from)

```r
#Removing aggregated and Male data
joined_by_hb_name <-full_join(all_contraceptive_data, health_boards, by = join_by(hbt == hb))
sex_data <- sex_data %>%
  filter(sex != "All")

joined_with_sex <- full_join(joined_by_hb_name, sex_data, by = join_by("hbt"=="hb"))%>%
  filter(sex != "Male")

contraception_proportion <- joined_with_sex %>%
  group_by(hbt) %>%
  filter(!is.na(number_of_paid_items)) %>%
  summarise(per_hb =sum(number_of_paid_items)/mean(all_ages))

#Limitation: As I have divided the number of paid items by mean of all ages the results may be skewed b

nhs_healthboard <- st_read(here("data/NHS_healthboards_2019.shp"))
```

```
## Reading layer `NHS_healthboards_2019' from data source
##   `C:\Users\maria\OneDrive\Documents\data science\week_7\B230189\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
```

```r
joined_with_polygon <-full_join(nhs_healthboard, contraception_proportion, by = join_by("HBCode"=="hbt")

contraception_distribution_hb <- joined_with_polygon %>%
  ggplot(aes(fill = per_hb))+
   scale_fill_viridis_c( name = "Contraception prescription proportion", option = "D") +
  geom_sf()+
  labs(title = "Contraception Prescribed to Women Across Scotland's Healthboards",
       subtitle = "Measured Proportionally Against Each Healthboard's Female Population") +
  theme_minimal()

contraception_distribution_hb
```

Figure 2 :

Observing whether or not there is a trend between age group and choice of contraceptive method

```r
collapsed_coc <- paste(birth_control_coc, collapse = "|")
collapsed_p <- paste(birth_control_prog, collapse = "|")
collapsed_iuc <-  paste(iuc, collapse = "|")

# Renaming bnf_item_description column so I can group by the contraceptive categories

renaming_column_function <- function(dataset, column_name, replacement_word, word_to_be_replaced) {
  dataset %>%
    mutate({{ column_name }} := if_else(str_starts({{ column_name }}, word_to_be_replaced), replacement_
```

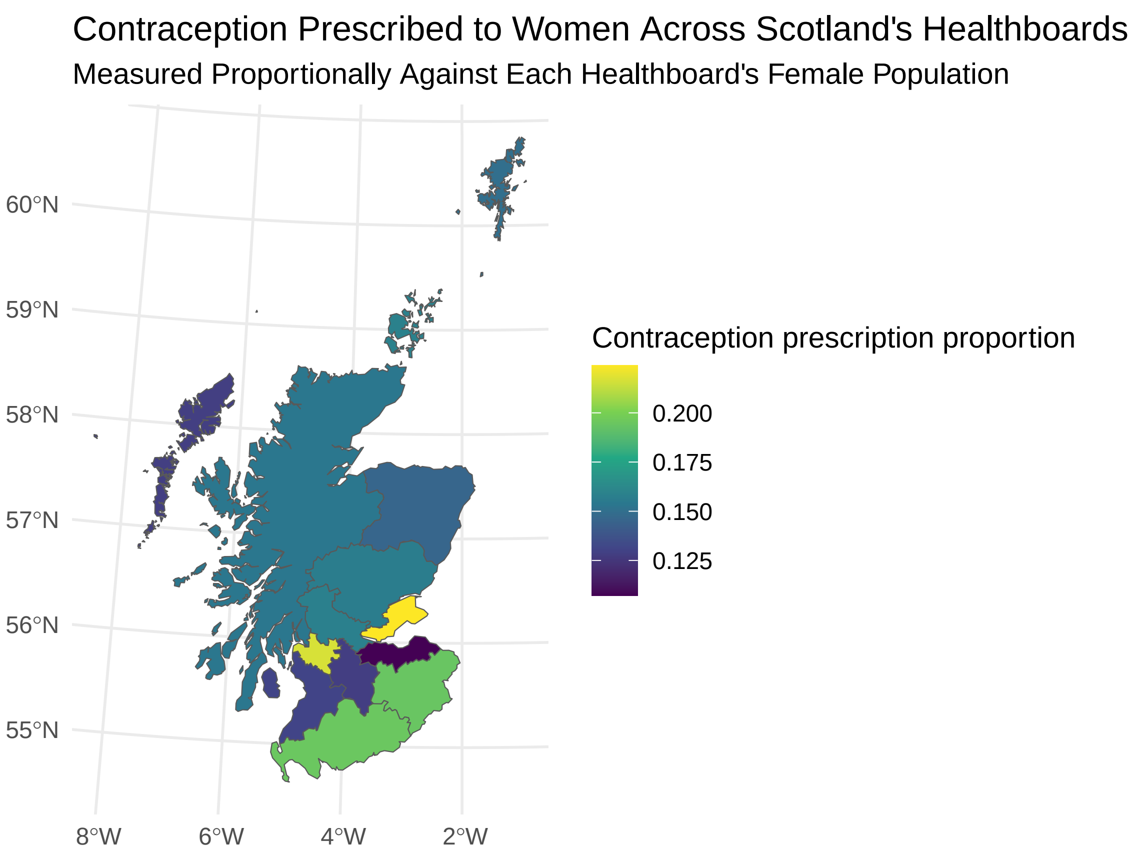


Figure 1: The distribution shows the highest rates of contraception prescription in Lothian and Glasgow & Clyde. This may be because these healthboards contain Scotland's two most densely popualted, big cities and people may be more sexually active here. The lowest rates are observed in Tayside. There is a general trend of higher levels in the South compared to the North.

```
}

grouped_contraceptive_data <- joined_with_sex %>%  renaming_column_function( bnf_item_description, "com
  renaming_column_function(bnf_item_description, "progesterone", collapsed_p) %>%
  renaming_column_function(bnf_item_description, "iuc", collapsed_iuc) %>%
  renaming_column_function(bnf_item_description, "hormone_patch", "EVRA") %>%
  renaming_column_function(bnf_item_description, "injection", "DEPO-PROVERA")

#Categorising each one by age-group categories, although people take contraceptives under the age of 16

grouped_contraceptive_data_1 <- grouped_contraceptive_data %>%
  group_by(bnf_item_description) %>%
  summarize(
    "16-25" = sum(rowSums(across(age16:age25) * number_of_paid_items), na.rm = TRUE),
    "26-35" = sum(rowSums(across(age26:age35) * number_of_paid_items), na.rm = TRUE),
    "36-45" = sum(rowSums(across(age36:age45) * number_of_paid_items), na.rm = TRUE),
    "46-55" = sum(rowSums(across(age46:age55) * number_of_paid_items), na.rm = TRUE)
  )

grouped_contraceptive_data_1 <- grouped_contraceptive_data_1 %>%
  pivot_longer(cols = `16-25`:`46-55`, values_to = "weighted_pop", names_to = "age")

grouped_contraceptive_data_1 %>%
  filter(!is.na(bnf_item_description)) %>%
  ggplot(aes(x = age, y = weighted_pop, fill = bnf_item_description)) +
  geom_col() +
  facet_wrap(~ bnf_item_description, labeller = labeller(bnf_item_description = c(
    "combined_pill" = "Combined Pill",
    "progesterone" = "Progesterone Only",
    "injection" = "Contraceptive Injection",
    "hormone_patch" = "Hormone Patch",
    "iuc" = "Intrauterine Contraceptive"
  ))) +
  labs(title = "Population Distribution by Age Group and Contraceptive Type",
       x = "Age Group",
       y = "Weighted Population",
       fill = "Contraceptive Type") +
  theme_minimal(base_size = 8)
```

```
grouped_contraceptive_data_1
```

```
## # A tibble: 24 x 3
##    bnf_item_description age   weighted_pop
##    <chr>                <chr>        <dbl>
##  1 combined_pill        16-25   9898562186
##  2 combined_pill        26-35  10925718666
##  3 combined_pill        36-45  10191473061
##  4 combined_pill        46-55  10678330135
##  5 hormone_patch        16-25    654246216
##  6 hormone_patch        26-35    718919443
##  7 hormone_patch        36-45    678140723
##  8 hormone_patch        46-55    710611449
```

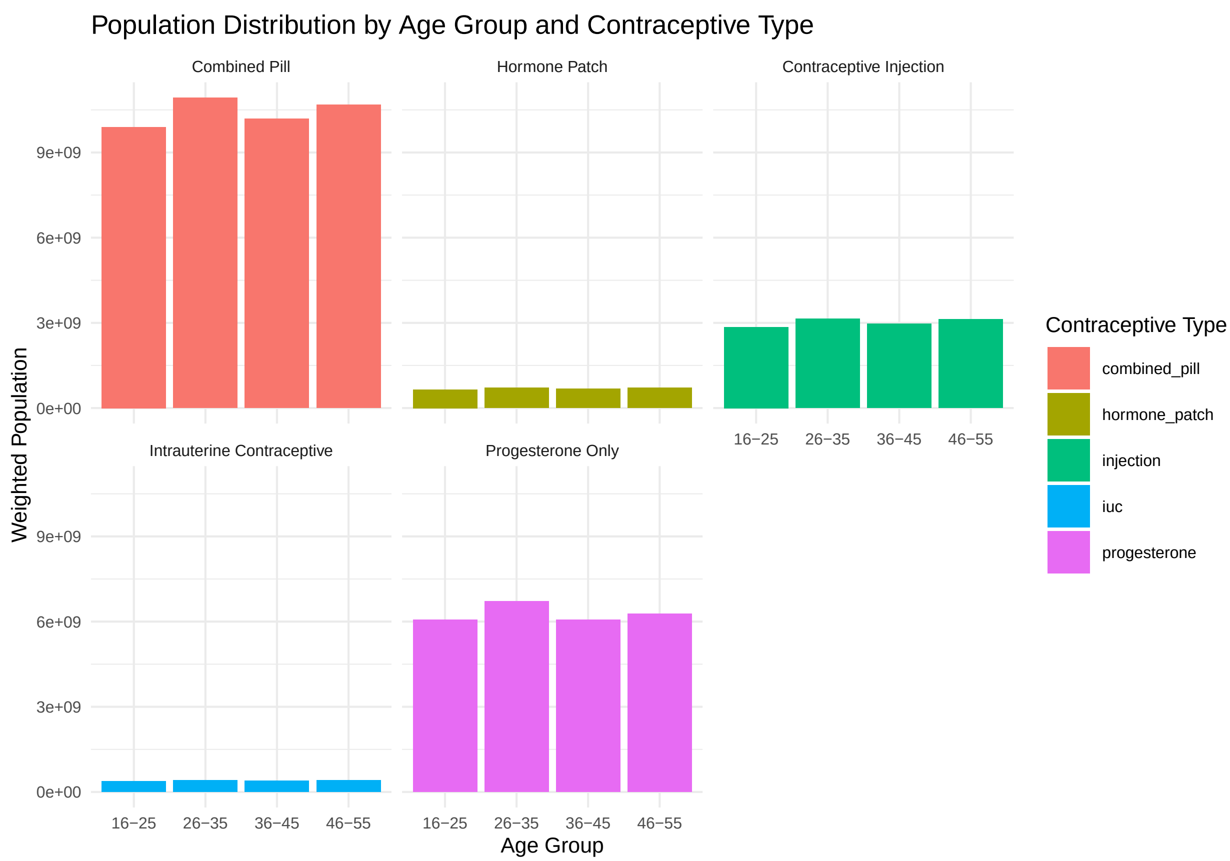


Figure 2: The distribution of age across contraceptive types are similar with age 26-35 consistently being prescribed the largest number of each contraceptive. Older age groups seem to be prescribed less progesterone only pills with the 16-25 group being prescribed slightly less than 36-45 in all methods except progesterone pills and 46-55 also being prescribed proportionally less compared to the other types.The combined pill, as expected, has the highest overall prescription and the IUC has the lowest.

```
## 9 injection            16-25   2853229451
## 10 injection           26-35   3149556353
## # i 14 more rows
```

Previous evidence shows that women under the age of 25 are most likely to use dependent forms of contraception such as the pill or condom, while women over 25 years of age used long-acting reversible contraceptives, such as IUC, implant and hormone patch(CITE). This visualisation does not closely mimic these results. This may be due to limitations of the joined dataset, as direct, person-specific comparisons between age and contraception rates were not possible so these graphs show a correlation between age distribution in healthboards and choice of contraception. Future visualisations could use a data set that collected information about age group and contraceptive use in the same cohort and could also comapare these findings to abortion rates - abortion pill was not prescribed a lot in the open NHS data set.

Figure 3

A table looking at the cost of each Contraceptive Type and whether or not it differs by Age Group.. Before describing the table see if i can make it interactive (if i have time / effort)

```
# Use of NA-character_ and the "^\\d+$" because otherwise I was getting errors about it not being numer

table_contraceptive_type <- grouped_contraceptive_data %>%
  pivot_longer(cols = starts_with("age"), names_to = "age", values_to = "population") %>%
  mutate(
    age = str_remove(age, "age"),
    age = if_else(str_detect(age, "^\\d+$"), age, NA_character_),
    age = as.numeric(age)
  ) %>%
  filter(!is.na(age))

# Classifying data points by age group
table_contraceptive_type <- table_contraceptive_type %>%
  mutate(age_group = case_when(
    age >= 16 & age <= 25 ~ "16-25",
    age >= 26 & age <= 35 ~ "26-35",
    age >= 36 & age <= 45 ~ "36-45",
    age >= 46 & age <= 55 ~ "46-55",
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(age_group))

# Join costs with population data, .groups drop allows for further manipulation of data
cost_per_age_group <- table_contraceptive_type %>%
  group_by(bnf_item_description, age_group) %>%
  summarise(
    total_population = sum(population, na.rm = TRUE),
    total_cost = sum(gross_ingredient_cost, na.rm = TRUE),
    cost_per_person = total_cost / total_population,
    .groups = "drop") %>%
  filter(!is.na(bnf_item_description))


cost_per_age_group %>%
   group_by(age_group) %>%
```

```r
  gt() %>%
  cols_label(bnf_item_description = "Contraceptive Type",
             # age_group = "Age Group",
             total_population = "Total Population",
             cost_per_person = "Cost per person(£)",
            total_cost = "Total Cost(£)") %>%
  # cols_move(columns = bnf_item_description, after = age_group) %>%
  cols_align(columns = age_group, align = "left") %>%
  tab_header(
    title = md("**Cost of Contraception for Each Age Group**"),
    subtitle = md("Across Scotland in 2022")
  ) %>%
  summary_rows(columns = cost_per_person,
              fns = list("Average" = ~mean(., na.rm = TRUE)))
```

Figure 4

Scatterplot seeing if there is a correlation between proportion of Live vs Still Births and contraceptive prescription rate comparing healthboards.

Used plotly and the notation i learnt from the plotly in r website.

```r
# Necessary to join hospital code to Healthboard code because Live/Still Birth information is done by h
join_hb_hospital_birth_22 <-  left_join(births_by_hospital, hospital_and_hb, by = join_by("hospital"==

# Finding out the proportion of Live Births in 2022 for each Healthboard
summarised_births_22 <- join_hb_hospital_birth_22 %>%
  filter(financial_year == "2022/23") %>%
  group_by(health_board) %>%
  pivot_wider(names_from = outcome, values_from = smr02births) %>%
  summarise(
    total_live = sum(Live, na.rm = TRUE),
    total_still = sum(Still, na.rm = TRUE),
    proportion_alive = total_live / (total_live + total_still)
  )

summarised_births_22 <- full_join(summarised_births_22, health_boards, by = join_by("health_board"== "hb

summarised_births_22 <-   full_join(summarised_births_22, contraception_proportion, by = join_by("healtl

# To make the plot interactive I used the plotly package, the text parameter defines the tooltip conten

  plotted_births <- summarised_births_22 %>%
  ggplot(aes(x=proportion_alive, y = per_hb, colour = hb_name, text = paste(
      "Health Board:", hb_name,
      "<br>Total Live Births:", total_live,
      "<br>Total Stillbirths:", total_still,
      "<br>Proportion Alive:", round(proportion_alive,4),
      "<br>Contraceptive Rate:", round(per_hb,2)
    )))+
  geom_point()+
  scale_x_continuous(limits = c(0.9953350, 1.00))+
```

Cost of Contraception for Each Age Group

Across Scotland in 2022

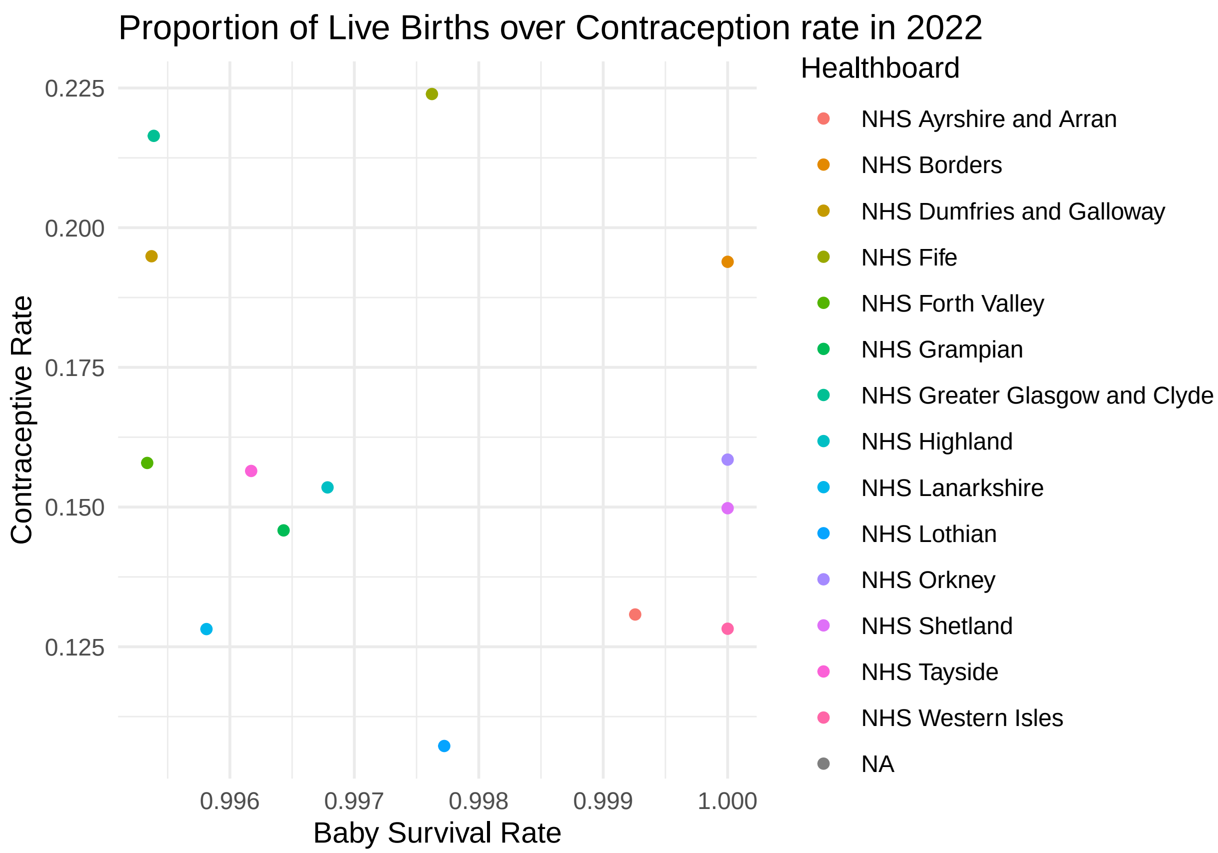
	Contraceptive Type	Total Population	Total Cost(£)	Cost per person(£)
16-25				
	combined_pill	2576007504	12685715	0.004924565
	hormone_patch	327304176	4710493	0.014391790
	injection	588446142	4911733	0.008346954
	iuc	141810234	10677633	0.075295222
	progesterone	1070642689	7393402	0.006905573
Average	—	—	—	0.02197282
26-35				
	combined_pill	2843304757	12685715	0.004461609
	hormone_patch	360890229	4710493	0.013052426
	injection	651726102	4911733	0.007536498
	iuc	154792230	10677633	0.068980420
	progesterone	1186272664	7393402	0.006232464
Average	—	—	—	0.02005268
36-45				
	combined_pill	2650831324	12685715	0.004785561
	hormone_patch	336640513	4710493	0.013992651
	injection	606524783	4911733	0.008098156
	iuc	147180101	10677633	0.072548076
	progesterone	1098257194	7393402	0.006731940
Average	—	—	—	0.02123128
46-55				
	combined_pill	2773239052	12685715	0.004574332
	hormone_patch	352446051	4710493	0.013365146
	injection	637365515	4911733	0.007706304
	iuc	154552334	10677633	0.069087492
	progesterone	1149615978	7393402	0.006431193
Average	—	—	—	0.02023289

```
  labs(title = "Proportion of Live Births over Contraception rate in 2022",
       y = "Contraceptive Rate",
       x = "Baby Survival Rate",
       colour = "Healthboard")+
    theme_minimal()


#plotted_births <-ggplotly(plotted_births, tooltip = "text")
```

```
plotted_births
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
```



NEXT STEPS : Limitations of the dataset are discussed, and "next steps" suggested, particularly in terms of data that would allow for further analysis.

Use of Generative AI : to help me sort through errors and warnings when R documentation wasn't helpful enough