EVALUATING TEENAGE PREGNANCY RATES IN NHS LOTHIAN: A Comparative Analysis of Scottish Health Boards (2003-2021)

FALINYI SAMSON

2024-11-26

Teenage pregnancy remains a significant public health concern in Scotland, with implications for both maternal and child health outcomes. This analysis focuses on NHS Lothians teenage pregnancy rates from 2003 to 2021, examining trends and comparing them with other Scottish health boards to inform future interventions.

RESEARCH QUESTION: How do teenage pregancy rates (TPRs) in NHS Lothian compare to other health boards in Scotland, and what trends can be identified to inform targeted interventions?

LOADING PACKAGES

This analysis utilized the following R packages:

- tidyverse: For data manipulation and transformation,

- ggplot2: For creating visualizations, and

- stats: For statistical analysis (chi-square test)

```
library(tidyverse)
library(ggplot2)
```

LOAD DATA

This analysis uses teenage pregnancy data from the Scottish Public Health Observatory (ScotPHO), covering all Scottish health boads from 2003-2021. The dataset includes rates per 1,000 women aged 15-19,as well as confidence intervals and actual numbers of pregnancies.

```
#Load and inspect data

TPA.data <- read.csv("ScotPHO_datatab_extract_2024-11-23.csv")
```

```r
#data structure and summary statistics

str(TPA.data)
```

```
## 'data.frame':    285 obs. of  11 variables:
##  $ area_code                : chr  "S00000001" "S00000001" "S00000001" "S00000001" ...
##  $ area_type                : chr  "Scotland" "Scotland" "Scotland" "Scotland" ...
##  $ area_name                : chr  "Scotland" "Scotland" "Scotland" "Scotland" ...
##  $ year                     : int  2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 ...
##  $ period                   : chr  "2002 to 2004 calendar years; 3-year aggregates" "2003 to 2005 ca
##  $ type_definition          : chr  "Crude rate per 1,000 females aged 15-19" "Crude rate per 1,000 f
##  $ indicator                : chr  "Teenage pregnancies" "Teenage pregnancies" "Teenage pregnancies"
##  $ numerator                : num  8616 8825 9043 9209 9158 ...
##  $ measure                  : num  54.6 55.7 56.7 57.4 56.6 54.5 51.5 47.9 44.7 41.2 ...
##  $ upper_confidence_interval: num  55.7 56.8 57.9 58.6 57.8 55.7 52.6 49 45.8 42.2 ...
##  $ lower_confidence_interval: num  53.4 54.5 55.6 56.2 55.5 53.4 50.4 46.8 43.7 40.2 ...
```

```r
head(TPA.data)
```

```
##   area_code area_type area_name year
## 1 S00000001  Scotland  Scotland 2003
## 2 S00000001  Scotland  Scotland 2004
## 3 S00000001  Scotland  Scotland 2005
## 4 S00000001  Scotland  Scotland 2006
## 5 S00000001  Scotland  Scotland 2007
## 6 S00000001  Scotland  Scotland 2008
##                                        period
## 1 2002 to 2004 calendar years; 3-year aggregates
## 2 2003 to 2005 calendar years; 3-year aggregates
## 3 2004 to 2006 calendar years; 3-year aggregates
## 4 2005 to 2007 calendar years; 3-year aggregates
## 5 2006 to 2008 calendar years; 3-year aggregates
## 6 2007 to 2009 calendar years; 3-year aggregates
##                          type_definition           indicator numerator measure
## 1 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    8616.0    54.6
## 2 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    8825.3    55.7
## 3 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    9043.0    56.7
## 4 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    9209.0    57.4
## 5 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    9158.3    56.6
## 6 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    8890.3    54.5
##   upper_confidence_interval lower_confidence_interval
## 1                      55.7                      53.4
## 2                      56.8                      54.5
## 3                      57.9                      55.6
## 4                      58.6                      56.2
## 5                      57.8                      55.5
## 6                      55.7                      53.4
```

```r
summary(TPA.data)
```

```
##   area_code          area_type          area_name              year
```

```
##  Length:285         Length:285         Length:285         Min.   :2003
##  Class :character   Class :character   Class :character   1st Qu.:2007
##  Mode  :character   Mode  :character   Mode  :character   Median :2012
##                                                           Mean   :2012
##                                                           3rd Qu.:2017
##                                                           Max.   :2021
##
##     period           type_definition     indicator          numerator
##  Length:285         Length:285         Length:285         Min.   :  10.3
##  Class :character   Class :character   Class :character   1st Qu.: 143.0
##  Mode  :character   Mode  :character   Mode  :character   Median : 431.9
##                                                           Mean   : 904.4
##                                                           3rd Qu.: 791.2
##                                                           Max.   :9209.0
##                                                           NA's   :13
##     measure      upper_confidence_interval lower_confidence_interval
##  Min.   :11.90   Min.   :22.00             Min.   : 4.40
##  1st Qu.:27.80   1st Qu.:33.80             1st Qu.:22.00
##  Median :36.40   Median :43.80             Median :31.70
##  Mean   :38.52   Mean   :44.77             Mean   :33.52
##  3rd Qu.:49.80   3rd Qu.:56.10             3rd Qu.:46.60
##  Max.   :71.20   Max.   :76.10             Max.   :66.50
##
```

DATA CLEANING

This step will identify if there are any missing values in the dataset. The Year column will be transformed into a character format for easy manipulation and visualisation.

```
#check for missing values

missing_values <- sum(is.na(TPA.data))
cat("Number of missing values:", missing_values, "/n")
```

```
## Number of missing values: 13 /n
```

```
#conversion of year to character for initial analysis

TPA.data$year <- as.character(TPA.data$year)
TPA.data %>% head()
```

```
##   area_code area_type area_name year
## 1 S00000001  Scotland  Scotland 2003
## 2 S00000001  Scotland  Scotland 2004
## 3 S00000001  Scotland  Scotland 2005
## 4 S00000001  Scotland  Scotland 2006
## 5 S00000001  Scotland  Scotland 2007
## 6 S00000001  Scotland  Scotland 2008
##                                         period
## 1 2002 to 2004 calendar years; 3-year aggregates
## 2 2003 to 2005 calendar years; 3-year aggregates
```

```
## 3 2004 to 2006 calendar years; 3-year aggregates
## 4 2005 to 2007 calendar years; 3-year aggregates
## 5 2006 to 2008 calendar years; 3-year aggregates
## 6 2007 to 2009 calendar years; 3-year aggregates
##                           type_definition           indicator numerator measure
## 1 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    8616.0    54.6
## 2 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    8825.3    55.7
## 3 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    9043.0    56.7
## 4 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    9209.0    57.4
## 5 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    9158.3    56.6
## 6 Crude rate per 1,000 females aged 15-19 Teenage pregnancies    8890.3    54.5
##   upper_confidence_interval lower_confidence_interval
## 1                      55.7                      53.4
## 2                      56.8                      54.5
## 3                      57.9                      55.6
## 4                      58.6                      56.2
## 5                      57.8                      55.5
## 6                      55.7                      53.4
```

REGIONAL COMPARISON:

This analysis compares teenage pregnancy rates (TPRs) across Scottish health boards at three strategic time points:

Baseline (2003): Initial measurement point

Midpoint (2012): To assess progress after first decade

Endpoint (2021): Most recent available data

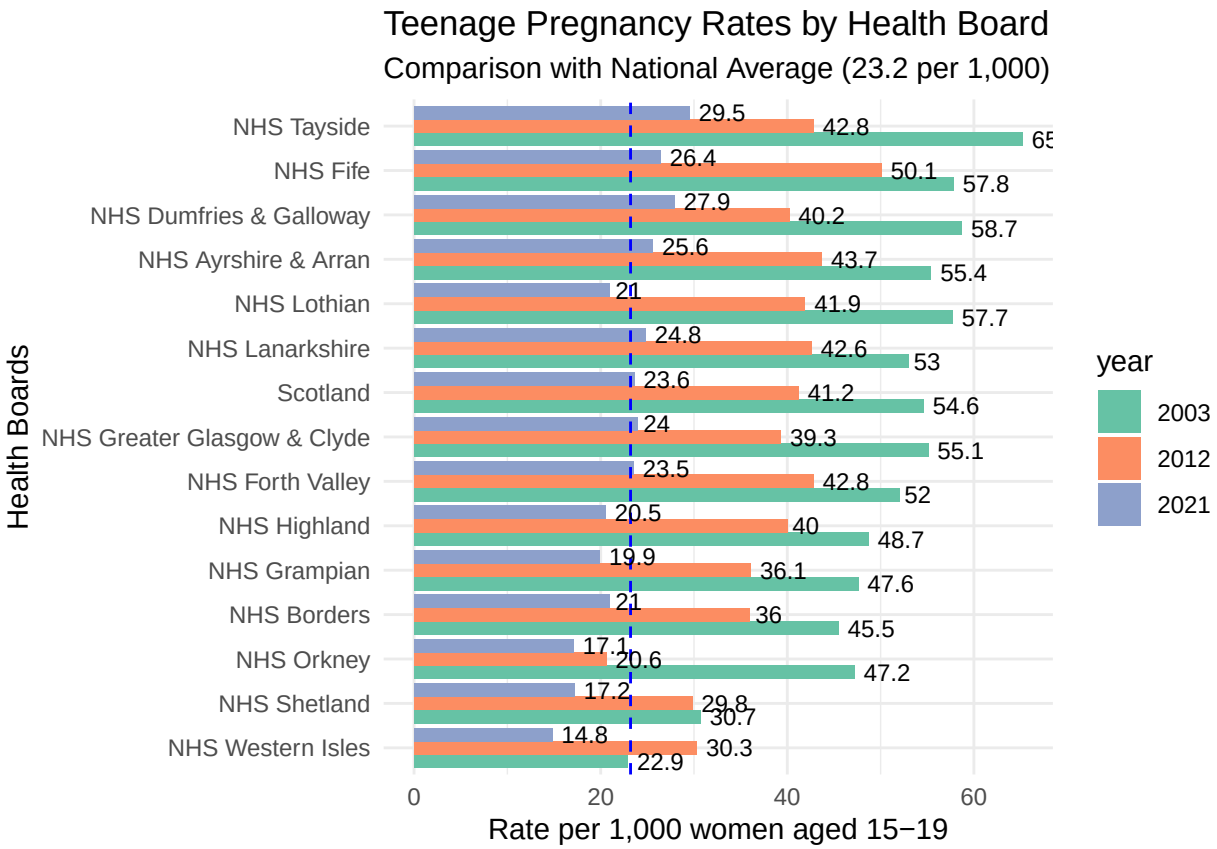
The mean TPR will be calculated for each health board to enable both cross-sectional comparison and temporal trend analysis. The national average of 23.2 per 1,000 women aged 15-19 serves as a reference point.

```
#Mean TPR for each health board by year

mean.tpa <- TPA.data %>%
  group_by(year, area_name) %>%
  summarise(mean.tpa = mean(measure), .group = 'drop') %>% ungroup()
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
#Regional comparison visualization

mean.tpa %>% filter(year %in% c(2003, 2012, 2021)) %>%
  mutate(year = as.factor(year)) %>%
ggplot(aes(reorder(area_name, mean.tpa), mean.tpa, fill = year)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_hline(yintercept = 23.2, linetype= "dashed", color = "blue") +
  geom_text(aes(label=round(mean.tpa, 1)),
```

```
          position = position_dodge(width = 0.9),
          hjust = -0.2,
          size = 3) +
coord_flip() +
labs(title = "Teenage Pregnancy Rates by Health Board",
     subtitle = "Comparison with National Average (23.2 per 1,000)",
     x = "Health Boards",
     y = "Rate per 1,000 women aged 15-19",
     fill = "year") +
theme_minimal()+
scale_fill_brewer(palette = "Set2")
```



The bar-chart above illustrates the mean teenage pregnancy rates across Scottish health boards at three strategic time points (2003, 2012, 2021). NHS Lothian's rates are shown alongside other health boards, with the blue dashed line indicating Scotland's national average (23.2 per 1,000 women aged 15-19).

This bar-chart reveals that NHS Lothian, like most health boards, has experienced a substantial decrease in teenage pregnancy rates over the years, with its current rates falling below the national average for Scotland. The consistent color coding and numerical labels make it easy to track changes across the three time points for each health board.

STATISTICAL ANALYSIS

This employs a chi-square test to evaluate changes in NHS Lothian's teenage pregnancy rates across three time points (2003, 2012, 2021). The chi-square test was chosen as the most appropriate statistical method as it satisfies key assumptions:

- Independent observations: Each pregnancy count is independent across years.

- Categorical data: Analysis uses discrete pregnancy counts rather than continuous rates

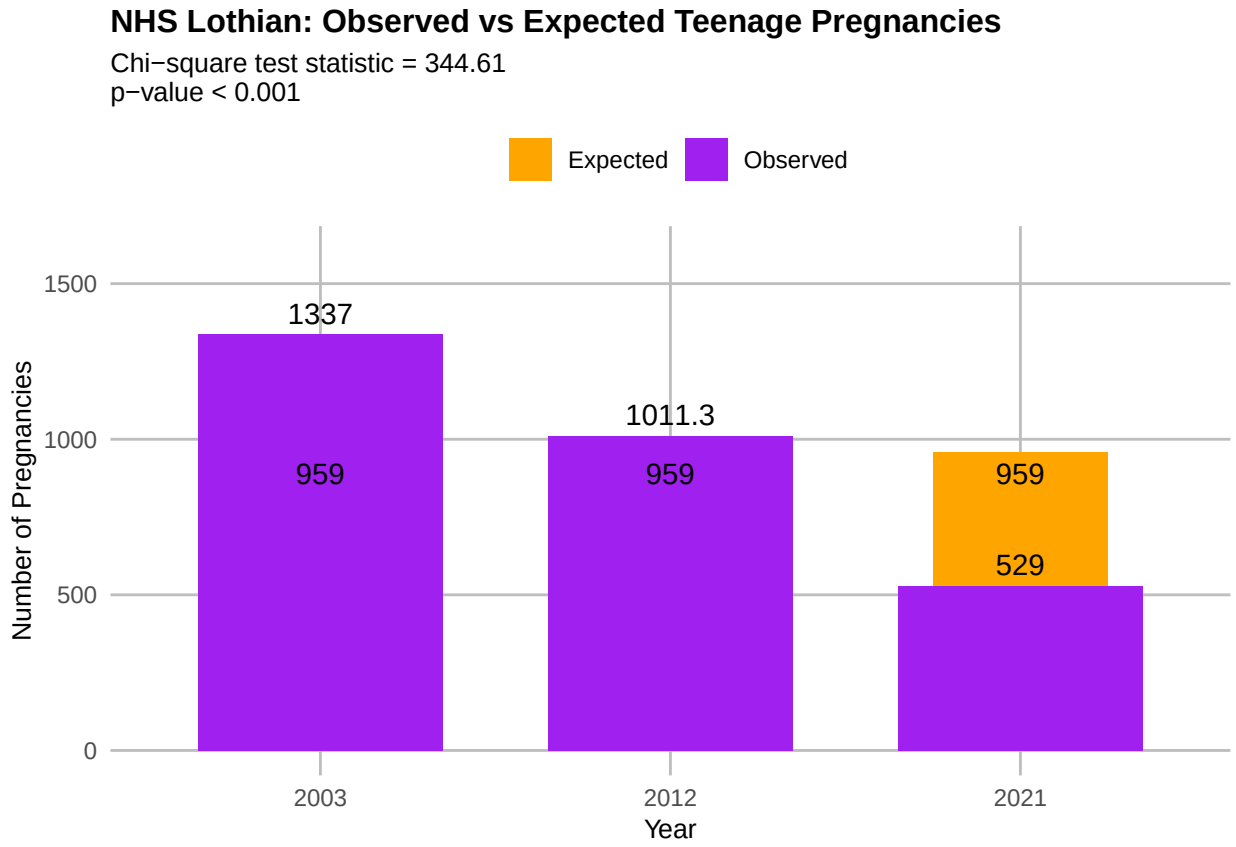
- Adequate sample size: All expected frequencies are >5.

- Mutually exclusive categories: Each case belongs to only one category.

This approach will compare observed numbers (numbers of teenage pregnancies) against expected values (numbers expected if there was no real change between the years), test whether changes are statistically significant, and provide evidence for intervention effectiveness.

```
#Preparation of data for statistical analysis

stats_data <- TPA.data %>%
  filter(area_name == "NHS Lothian" &
           year %in% c(2003, 2012, 2021)) %>%
  select(year, numerator)

#calculation for observed and expected values

Observed <- as.numeric(stats_data$numerator)
Expected <- rep(mean(Observed),3)

#chi-square test for statistical analysis

chi_test <- chisq.test(Observed, p = Expected/sum(Expected))

# conversion of year to factor to create dataframe for visualization

chi_plot_data <- data.frame(
  Year = factor(c("2003", "2012", "2021")),
  Observed = Observed,
  Expected = Expected
)
#creating statistical visualization
```

```r
ggplot(chi_plot_data) +
  geom_col(aes(x = Year, y = Expected, fill = "Expected"),
           position = position_identity (), width = 0.5) +
  geom_col(aes(x = Year, y = Observed, fill = "Observed"),
           position = position_identity(),
           width = 0.7) +

#labels and titles

  labs(title = "NHS Lothian: Observed vs Expected Teenage Pregnancies",
       subtitle = paste("Chi-square test statistic =", round(chi_test$statistic, 2),
                        "\np-value < 0.001"),
       y = "Number of Pregnancies",
       x = "Year",
       fill = "") +
  scale_fill_manual(values = c("Observed" = "purple", "Expected" = "orange")) +

  geom_text(aes(x = Year, y = Observed,
                label = Observed),
            vjust = -0.5, size = 4) +
  geom_text(aes(x = Year, y = Expected, label = round(Expected,0)),
            vjust = 1.5,
            size = 4) +

  theme_minimal() +
  theme(
    plot.title = element_text(size = 12, face = "bold"),
    plot.subtitle = element_text(size = 10),
    axis.title = element_text(size = 10),
    legend.position = "top",
    panel.grid.major = element_line(color = "gray"),
    panel.grid.minor = element_blank()
    ) +
    ylim(0, max(chi_plot_data$Observed)*1.2)
```



STATISTICAL ANALYSIS RESULT

The analysis reveals significant decline from 1,337 cases (2003) to 529 cases (2021), and chi-square test statistic of 344.61 ($p < 0.001$).

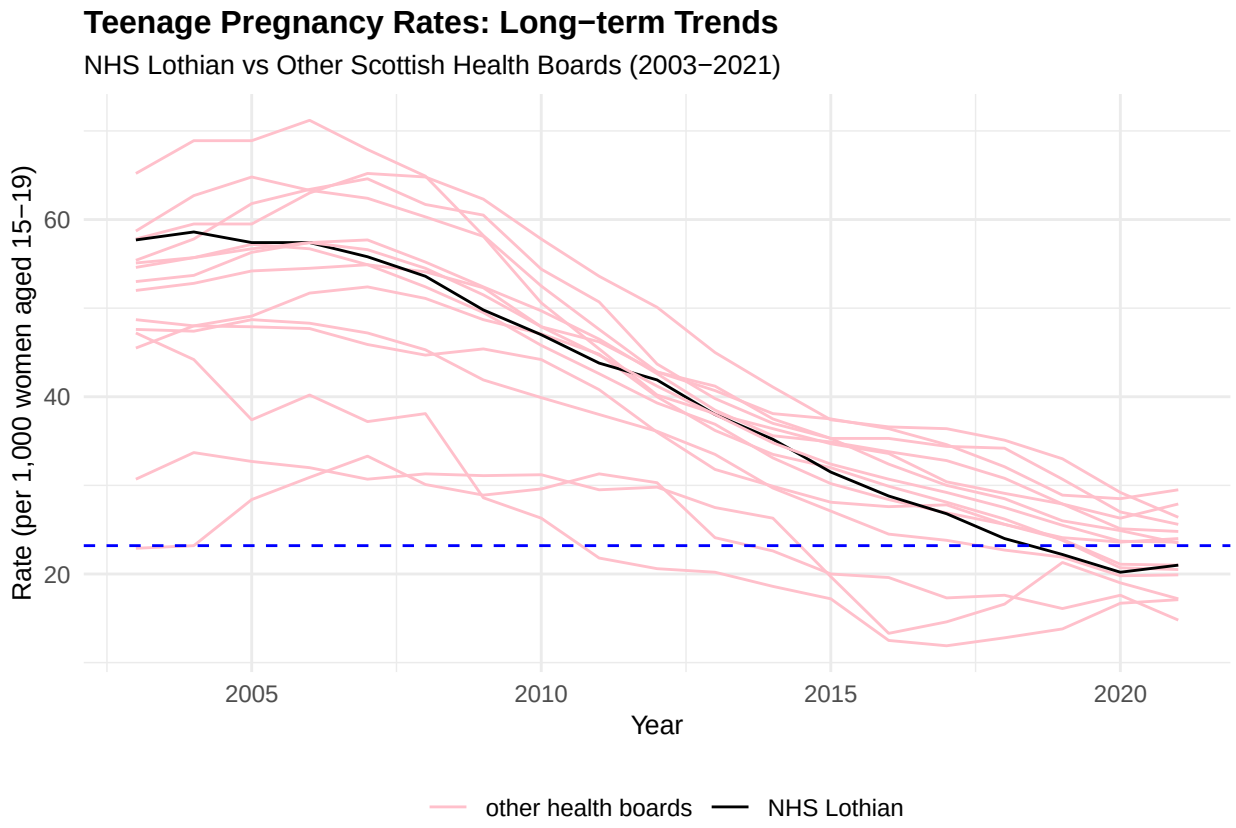
These results provide strong statistical evidence that the reduction in teenage pregnancies represents a genuine trend rather than random variation.

TREND ANALYSIS

The trend analysis will examine temporal patterns in teenage pregnancy rates from 2003 to 2021, using a line graph to visualise how NHS Lothian's rates have changed over time compared to other Scottish health boards. The line graph will aim to Track NHS Lothian's progress and compare it against other Scottish health boards, making reference to the national average for TPRs in Scotland.

```
#conversion of year variable back to numeric for trend analysis
TPA.data$year <- as.numeric(TPA.data$year)

#creating trend visualization
TPA.data %>% group_by(year) %>%
  ggplot(aes(x= year, y= measure, fill = area_name, colour = area_name == "NHS Lothian")) +
```

```
  geom_line(stat = "identity") +
  geom_hline(yintercept = 23.2, color = "blue", linetype = "dashed") +
  theme_minimal() +
  theme(legend.position = "bottom") +
  scale_color_manual(values = c("TRUE" = "black", linetype = "bold", "FALSE" = "pink"),
 labels = c("TRUE" = "NHS Lothian", "FALSE" = "other health boards")) +
  scale_size_manual(values = c("TRUE" = 1.5, "FALSE" = 0.5), guide = "none") +
  guides(colour = guide_legend(title = "")) +
  labs(title = "Teenage Pregnancy Rates: Long-term Trends",
       subtitle = "NHS Lothian vs Other Scottish Health Boards (2003-2021)",
x = "Year",
y = "Rate (per 1,000 women aged 15-19)") +
  theme(
    plot.title = element_text(size = 12, face = "bold"),
    plot.subtitle = element_text(size = 10),
    axis.title = element_text(size = 10)
    )
```



TREND ANALYSIS RESULTS

The trend analysis reveals several significant patterns. It highlights NHS Lothian's Progress (black line) from initial rate (57.7 per 1,000 (2003)), to final rate (21.0 per 1,000 (2021)). It also shows a consistent downward trajectory as well as a successful drop below national average. The decline was observed across other Scottish health boards (pink lines), suggesting the effectiveness of nationwide pregnancy prevention strategies.

CONCLUSION AND RECOMMENDATIONS

This analysis of teenage pregnancy rates in NHS Lothian (2003-2021) demonstrates significant public health progress. It reveals a substantial reduction from 57.7 to 21.0 per 1,000 women aged 15-19, with rates now below Scotland's national average of 23.2. The statistical analysis (chi-square test) confirms these improvements are significant ($p < 0.001$), and the trend analysis shows consistent decline across all health boards. These findings suggest successful implementation of prevention strategies, highlighting the importance of maintaining current approaches, while sharing best practices across health boards for continued improvement in teenage pregnancy prevention across Scotland.

SUMMARY AND CRITICAL REFLECTION

Teenage pregnancy, also known as adolescent pregnancy, refers to conception in females under the age of 20. It significantly impacts both maternal and child health outcomes (WHO, 2024). In Scotland, while rates have decreased by 60% since 2007, it continues to be a priority issue due to its association with adverse health outcomes, educational disruption, and socioeconomic disadvantages (Cook et al., 2017). The national average of 23.2 pregnancies per 1,000 women aged 15-19 recorded in 2021 indicates progress but highlights ongoing challenges (Public Health Scotland).

This analysis, targeting public health officials and policymakers within NHS Lothian and Scotland, employed three complementary visualizations following established data communication principles (Healy, 2024):

- A regional comparison bar chart which demonstrated geographic variations and trends across three strategic time points (2003, 2012, 2021).
- Statistical analysis through chi-square testing, selected for its suitability with count data and satisfaction of key assumptions (independent observations, adequate sample size, mutually exclusive categories), confirmed the significance of observed changes.
- A trend line graph which captured temporal patterns, revealing NHS Lothian's successful reduction from 57.7 to 21.0 per 1,000 women aged 15-19.

The Scottish Public Health Observatory dataset (2003-2021) provided comprehensive coverage but presented several limitations:

- Three-year aggregated periods potentially masking short-term variations.
- Absence of socioeconomic indicators.
- Lack of distinction between planned and unplanned pregnancies.
- Limited demographic profiling of teenage mothers.

The visualization approach demonstrated significant strengths through:

- Strategic time point selection.
- Consistent visual elements.
- Multiple analytical perspectives, where the chi-square test results (chi = 344.61, $p < 0.001$) provided statistical evidence for genuine trend changes rather than random variations.

However, several limitations were observed, and they include:

- Visualization complexity due to multiple health board trajectories.
- Potential oversimplification through three-point temporal analysis
- Limited representation of spatial relationships.

The findings demonstrate substantial progress in reducing teenage pregnancy rates across Scotland, with NHS Lothian showing particularly notable improvements. This evidence supports the effectiveness of implemented prevention strategies while highlighting areas for continued monitoring and intervention. Maintaining successful prevention approaches while developing targeted support for areas still above the national average and emphasizing the importance of continued cross-board collaboration in teenage pregnancy prevention efforts is highly recommended.

REFERENCES

Cook, S.M. and Cameron, S.T., 2017. Social issues of teenage pregnancy. *Obstetrics, Gynaecology & Reproductive Medicine*, *27*(11).

Healy, K. (2024) *Data Visualization: A Practical Introduction*. Princeton University Press.

Public Health Scotland, 2022. *Teenage pregnancies: Year of conception ending 31 December 2022*. Available at: https://publichealthscotland.scot/publications/teenage-pregnancies/teenage-pregnancies-year-of-conception-ending-31-december-2022/ [Accessed 28 November 2024].

Scottish Public Health Observatory 2024 *Teenage Pregnancy Data. Available at: https://scotland.shinyapps.io/ScotPHO_profiles_tool/ (Accessed 26 November 2024).

WHO (World Health Organization) (2024) *Adolescent pregnancy*. Available at: https://www.who.int/news-room/fact-sheets/detail/adolescent-pregnancy (Accessed 26 November 2024).

AI ACKNOWLEDGEMENT

I acknowledge the use of generative AI (ChatGPT) to understand some error messages. The prompt used was “i got this error while trying to knit my work, what does it mean?”. The output recieved was used to understand what was preventing my work from knitting.