MMR Vaccine: Measles Susceptibility in Edinburgh

Project No.14

Julia Daws, Xinnan Huang, Beth MacLean, Carrie Turner

1. Introduction

Between 1998 and 2014, data were collected by The Scottish Childhood Immunisation Record System on measles, mumps and rubella (MMR) vaccination uptake. In 1998, Wakefield et al. published an article which linked the MMR vaccine with an increased risk of autism, which decreased the vaccination rates in 2003 across the United Kingdom. The article was partially withdrawn in 2004, then later discredited in 2010. This report aims to answer two main questions:

- Did Edinburgh exhibit a change in measles susceptibility following the retraction of the Wakefield article?
- Did the change, if any, in measles susceptibility occur in 2004 alongside the articles' retraction?

Throughout this report we will use both exploratory and formal methods of analysis to answer the questions posed.

2. Exploratory Analysis

We begin our exploratory analysis by creating a new column in the dataset which calculates the proportion of pre-school children that are susceptible to measles in each of Edinburgh's 101 Intermediate Zones (IZ). This is what we are interested in modelling and it is calculated using the formula:

$$Measles\ Susceptibility\ Rate = \frac{Susceptibility\ Children}{Total\ Children}$$

This computation yielded a new column, as presented in the following table.

Measles Susceptibility Data

Year	Total Children	Susceptible Children	Susceptibility Rate
1998	185	27	0.1459459
1998	141	19	0.1347518
1998	81	12	0.1481481
1998	168	27	0.1607143
1998	64	6	0.0937500
1998	194	26	0.1340206

For each year, we can calculate the mean proportion of children susceptible to measles. A summary is presented below:

Table 1: Mean proportion of pre-school children susceptible to measles by year.

1998	2000	2002	2004	2006	2008	2010	2012	2014
0.153	0.155	0.160	0.177	0.146	0.123	0.101	0.082	0.078

From Table 1, we can see that between 1998 and 2004, the mean proportion of pre-school children susceptible to measles increased. With a larger increase between 2002 and 2004. After 2004, there was a sharp decrease in the mean proportions. The mean proportion of pre-school children susceptible to measles in 2014 is approximately half of the mean proportion in 2006.

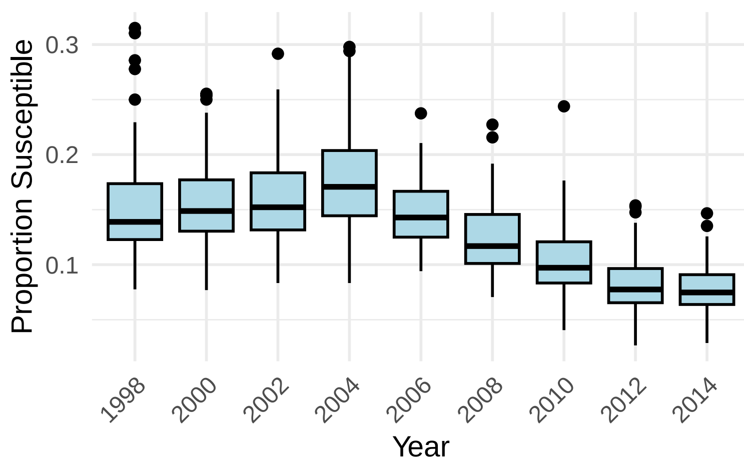

Figure 1: Distribution of mean proportion of measles susceptibility by year (1998 - 2014).

From Figure 1, we can see that the variability in the proportion of children susceptible is greater in the earlier years (1998-2004), since the boxplots are wider. The median proportion increased between 1998 and 2004, then began to decrease from 2006 onwards. Most of the boxplots appear to be symmetric, however there are a few outliers for each year.

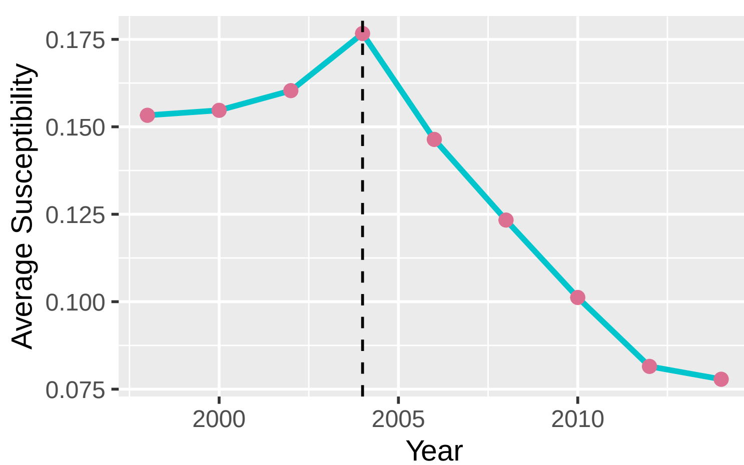

Figure 2: Mean measles susceptibility over time.

From Figure 2 we can see that there is a peak in mean susceptibility in 2004 before a sharp decrease afterwards, this aligns with all other plots and summaries we have seen before.

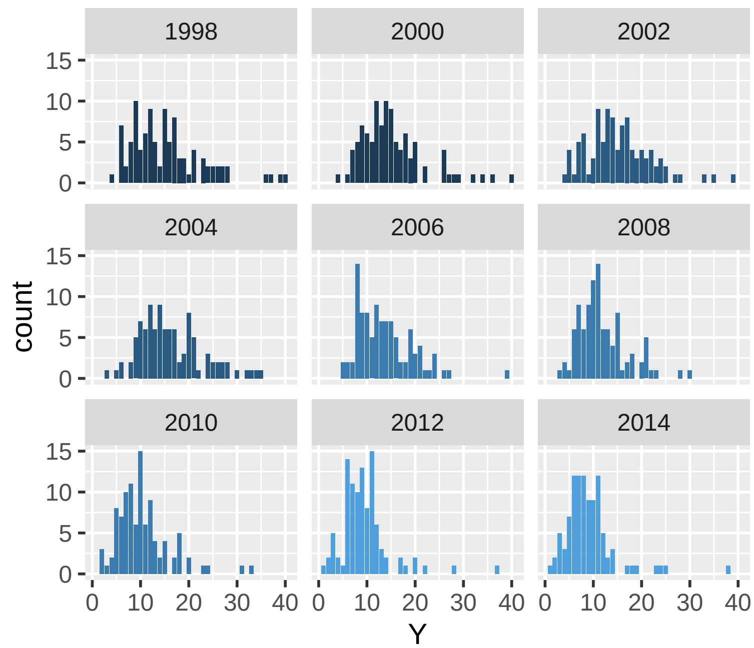

Figure 3: Histogram showing distribution of average measles suscepitibility for each year.

The histograms for the years 1998-2004 look relatively similar. From 2006 onwards, the histograms appear to be slightly right-skewed. For the earlier years (1998-2004), there is more variability in the number of children susceptible to measles in each of Edinburgh's 101 Intermediate Zones (IZ).

3. Formal data analysis

3.1 Forming a model

Here we will fit a simple linear regression model to the `measles` data.

Our best-fitting line to the data is:

$$\widehat{average\ rate} = \hat{\alpha} + \hat{\beta} x_i = 11.716 - 0.006 \cdot year$$

where:

- $\hat{\alpha}$=11.716 is the intercept coefficient and means that for a given IZ with a `Year=0`, their `Average_Rate` would be 11.716. Note that `Year=0` is not actually possible in our data set.
- $\hat{\beta}$=-0.006 is the slope coefficient associated with the explanatory variable `Average_Rate`. This summarises the relationship between our response variable, `Year`, and our explanatory variable `Average_Rate`. Since the slope coefficient is negative, this means that `Year` and `Average_Rate` have a negative linear relationship, i.e. as `Average_Rate` increases, `Year` decreases. Furthermore, for every 1 unit increase in `Average_Rate`, there is an associated decrease of, on average 0.006 units of `Year`.

The overall p-value for this model is 0.002. Since this is less than 0.05, we can say that the model fits the data sufficiently well.

Here is the model summary.

Table 2: Simple linear regression model

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	11.715770329	2.426732621	4.827796	0.001904290	5.97745952	17.454081138
Year	-0.005775266	0.001209733	-4.774000	0.002026307	-0.00863583	-0.002914702

Figure 4 is a plot of our data with a line of best fit superimposed.

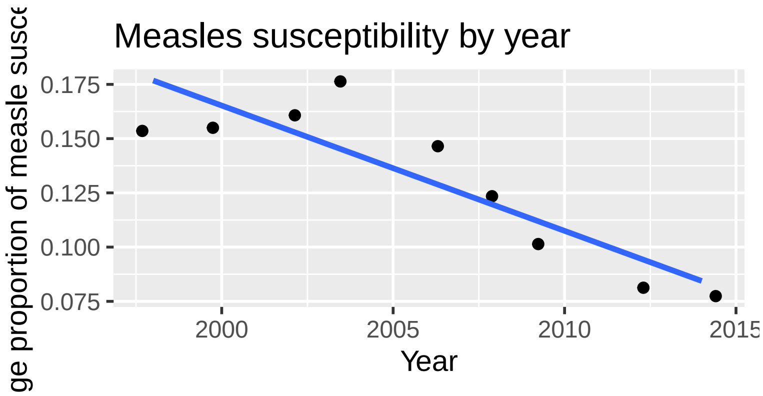

Figure 4: Jittered scatterplot of measles susceptibility by year

3.2 Assessing model fit

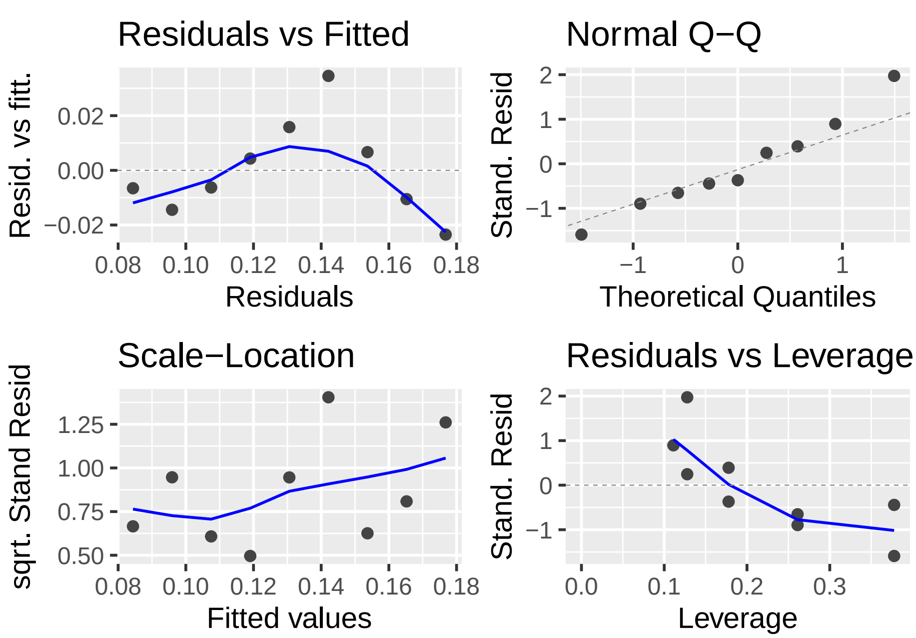

Figure 5: Assessing model assumptions

The majority of the points lie on the line in the QQ-plots. This means we can say the normality assumption is satisfied. The points lie fairly equal around the zero-line in the residuals plot, although it isn't overly clear. This means that the constant variance and zero-mean assumption is also satisfied. The independence assumption is also inherently satisfied due to the way that the data are collected.

3.3 Visualization and Analysis

A line graph was plotted based on the yearly average measles susceptibility rates, as presented in the figure below.

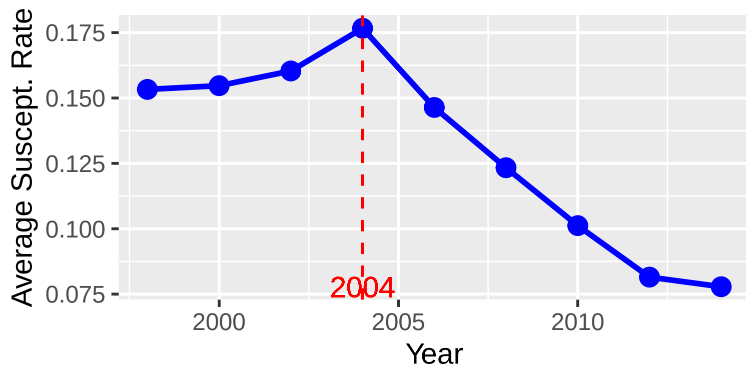

Figure 6: Line graph of measles susceptibility rate in Edinburgh (1998-2014)

This graph, similar to Figure 2, clearly indicates that the yearly average measles susceptibility rate initially increased, peaked in 2004, and then consistently declined.

Furthermore, a t-test was conducted to statistically evaluate the difference in measles susceptibility rates before and after 2004.

$$H_0 : \mu_{pre2004} = \mu_{post2004} \quad \text{and} \quad H_1 : \mu_{pre2004} \neq \mu_{post2004}$$

The table shows that p-value is less than 0.05, so we reject the null hypothesis.

t-value	Df	p-value	95% CI Lower	95% CI Upper	Pre_2004 Mean	Post_2004 Mean
12.6693	694.4696	0	0.0324	0.0442	0.1561	0.1178

4. Conclusions

During Section we could see hints that there was a change in measles susceptibility, this was initially shown in Table 1, and then once again in Figure 1, Figure 2 and Figure 3 . All of these plots also show the change happening after 2004.

In Section , a linear model with Year as a predictor, and Average Susceptibility Rate as the response. The model assumptions of normality, zero-mean, and constant variance are satisfied.The R^2 Adj. value of 0.7315 shows the model fits well but there is potential room for improvement. This could be potentially due to the number of predictors being low, meaning we are possibly missing some of the complexity of the model.

In the model summary Table 2 , both the intercept (Average Rate) and the Year variable are significant with p-values lower than 0.05. This is backed up by neither of the confidence levels containing zero.

We finally conducted a t-test to test if there is a difference between average susceptibility rates before and after 2004. This gave us a p-value of 0, and a confidence interval of $(0.0324, 0.0442)$ which doesn't include 0. These both indicate that we should reject H_0 and state that there is a significant difference in mean susceptibility rates before and after 2004.

We can hence state that there was a significant change in measles susceptibility after the retraction of the Wakefield article, and this did happen in 2004, coinciding with the retraction of the article as well.According to background information, the retraction of the article may have led to a rebound in vaccination rates, thereby reducing the risk of measles infection. Additionally, it can be observed that the difference in measles susceptibility rates between 2012 and 2014 was not significant, possibly due to reaching a threshold. It is anticipated that this figure will decrease in the future, but the trend will be very stable.