Analysis of Measles Susceptibility in Glasgow 1998-2012

Group 8

Table of contents

Introduction

In 1998, a study by Wakefield et al. controversially linked the MMR vaccine to autism, triggering widespread media coverage and a drop in vaccination uptake across the UK—falling to around 80% in some areas by 2003. Although the article was partially retracted in 2004 and fully discredited by 2010, the resulting vaccine hesitancy contributed to later measles outbreaks, including one in 2013.

This report investigates measles susceptibility among young children in Glasgow from 1998 to 2012 to determine if Glasgow exhibited a change in measles susceptibility following the Wakefield article's retraction. In particular, we examine whether any change occurred around 2004, when the article's claims were repudiated. We use data from DAProject15.csv, which contains yearly counts of measles-susceptible children in Glasgow's Intermediate Zones (IZs).

Key variables are:
Y – number of pre-school children (per IZ) susceptible to measles.
N – total number of pre-school children in that IZ.
Year – the year of observation.

By analyzing the proportion of susceptible children (Y/N) over time, we aim to assess trends before and after 2004. The analysis includes an exploratory visualization of susceptibility rates over 1998–2012 and formal statistical tests (e.g. chi-square tests) comparing the pre-2004 and post-2004 periods. The results will shed light on whether the public's response in Glasgow changed following the discrediting of the Wakefield study.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to understand trends in measles susceptibility among pre-school children in Glasgow from 1998 to 2012. This includes examining year-to-year patterns, comparing susceptibility before and after 2004, and visualizing distributions to support further statistical analysis.

Table 1: Mean Measles Susceptibility by Year

Year	mean_susceptibility
1998	0.0488
2000	0.0599
2002	0.0725
2004	0.0680
2008	0.0811
2010	0.0518
2012	0.0474

Table 1 *shows the mean measles susceptibility by year. No sharp decline is observed post-2004, indicating that vaccine hesitancy may have persisted after the Wakefield article's retraction. Any recovery in vaccination rates appears gradual.*

Data Visualisation

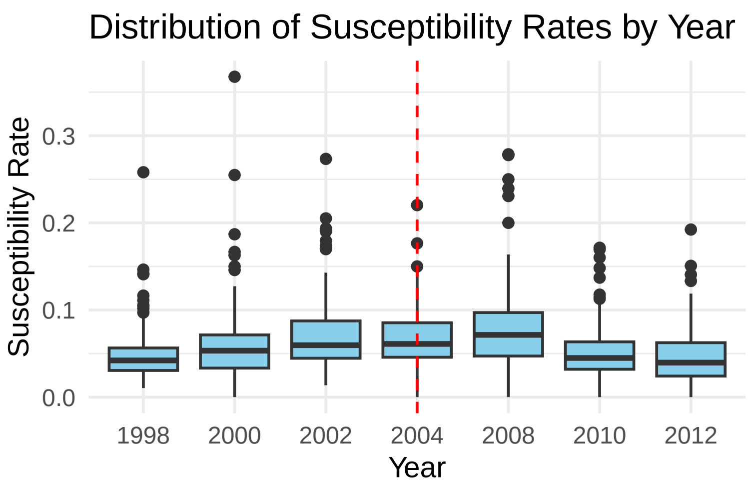

Figure 1: Boxplot of Measles Susceptibility Rates by Year in Glasgow (1998–2012)

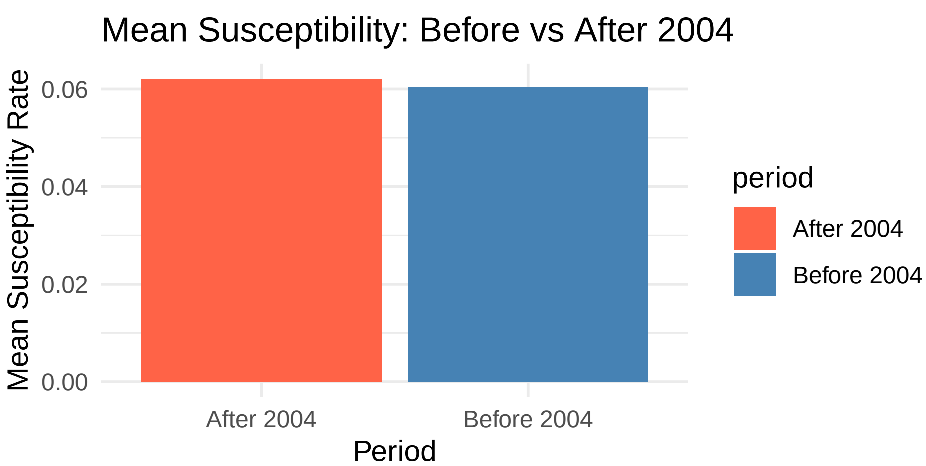

Figure 2: Mean Measles Susceptibility Before vs After 2004

Figure 1 and Figure 2 together show measles susceptibility trends across years and between periods. Figure 1 highlights yearly distributions, with the 2004 retraction marked, showing a peak in 2008 but no sharp drop post-2004. Figure 2 compares average susceptibility before and after 2004, revealing only a slight decrease. These patterns align with Table 1, suggesting vaccine hesitancy persisted beyond the article's retraction.

Formal Analysis

(a) We define **measles susceptibility** in each Intermediate Zone (IZ) and year as the proportion:

$$p_i = \frac{Y_i}{N_i}$$

where:

- (Y_i) is the number of susceptible pre-school children in the (i)-th IZ.
- (N_i) is the total number of pre-school children in the same IZ.
- (p_i) represents the proportion of susceptible children.

We aim to test whether the **proportion of susceptible children changed significantly after 2004**. We formulate two statistical models:

1. Comparison of Two Proportions (Chi-square Test)

To compare **pre-2004 vs. post-2004**, we define the null and alternative hypotheses:

$$H_0 : p_{\text{pre}} = p_{\text{post}}$$

$$H_A : p_{\text{pre}} \neq p_{\text{post}}$$

where:

- (p_{pre}) is the overall proportion of susceptible children in **1998, 2000, and 2002**.
- (p_{post}) is the overall proportion in **2004, 2008, 2010, and 2012**.

We use a **chi-square test for independence** on the contingency table,the p-value 0.7434 suggests that there is no statistically significant difference in the proportion of susceptible children between pre-2004 and post-2004.

Table 2: chi-square test for independence

statistic	p.value	parameter	method
0.11	0.74	1	Pearson's Chi-squared test with Yates' continuity correction

(b)

Following the initial chi-square test and exploratory analysis, we proceed with a more detailed statistical investigation by fitting a generalized linear mixed model (GLMM) with a binomial family and a logit link function to model changes in measles susceptibility among preschool children in Glasgow between 1998 and 2012.Let:

- Y_{it}: the number of susceptible children in intermediate zone i at year t
- N_{it}: total number of children in the same group
- p_{it}: probability that a child in zone i at time t is susceptible to measles

Then the model assumes:

$$Y_{it} \sim Binomial(N_{it}, p_{it})$$

and the logit of the probability is modeled as:

$$logit(p_{it}) = \beta_0 + \beta_1 \cdot seg1_{it} + \beta_2 \cdot seg2_{it} + \beta_3 \cdot seg3_{it} + \beta_4 \cdot I2004_{it} + \beta_5 \cdot I2010_{it} + (1|IZ)$$

where:

- β_0 is the global intercept, representing the baseline log-odds of susceptibility in the reference group .
- $\beta_1, \beta_2, \beta_3$ are the slopes for different time segments (gradual trends)
- β_4, β_5 capture abrupt level shifts in 2004 and 2010 respectively
- $seg1, seg2, seg3$ capture linear time trends within three periods: 1998–2003, 2004–2009, and 2010–2012, allowing different slopes for each segment.
- $I2004, I2010$ are binary indicators for years 2004 and 2010, respectively, capturing abrupt level shifts in susceptibility associated with the partial (2004) and full (2010) retraction of the Wakefield article.
- $(1|IZ)$ A random intercept term allowing each intermediate zone to have its own baseline level of susceptibility.

Table 3 presents the fixed effects estimates from the GLMM. The intercept of –3.025 represents the baseline log-odds of measles susceptibility in 1998. The coefficients for the segmented time trends indicate a significant increase at 1% significance level during 1998–2003 ($seg1:$ =0.107, p<0.001) and continued growth in 2004–2009 ($seg2:$ =0.112, p<0.001), followed by a significant decline at 5% significance level from 2010–2012 ($seg3:$ =–0.058, p<0.05), suggesting improved vaccination uptake in later years.

The jump terms are also statistically significant. An upward shift in 2004 (=0.32, p<0.001) aligns with the partial retraction of the Wakefield article, while a notable drop in 2010 (=–0.260, p<0.001) coincides with the article's full discrediting, indicating a public recovery in vaccine confidence.

Model fit is strong, with MarginalR^2=0.18 and ConditionalR^2=0.853, demonstrating that both fixed effects and random effects substantially explain the variation in susceptibility across time and space.

Table 3: Fixed Effects Summary of Full Model

	Fixed Effect Estimates			
term	estimate	std.error	statistic	p.value
(Intercept)	-3.025	0.051	-59.823	< 1e-06
seg1	0.107	0.013	7.949	< 1e-06
seg2	0.047	0.012	3.830	0.0001284
seg3	-0.058	0.028	-2.058	0.0395488
I2004	0.320	0.052	6.120	< 1e-06
I2010	-0.260	0.053	-4.853	1.214e-06

Note:

Marginal R^2 = 0.18; Conditional R^2 = 0.853

To identify the most appropriate model for explaining measles susceptibility in Glasgow, several candidate models were fitted:

- Model 1: Null model (random intercept only)

The logit of the probability is modeled as:

$$\text{logit}(p_{it}) = \beta_0 + b_i$$

where

$$b_i \sim \mathcal{N}(0, \sigma^2)$$

is a random intercept for intermediate zone i.

- Model 2: Segmented time trends only (no random effects)

$$\text{logit}(p_{it}) = \beta_0 + \beta_1 \cdot \text{seg1}_{it} + \beta_2 \cdot \text{seg2}_{it} + \beta_3 \cdot \text{seg3}_{it}$$

- Model 3: Segmented trends with random intercepts

$$\text{logit}(p_{it}) = \beta_0 + \beta_1 \cdot \text{seg1}_{it} + \beta_2 \cdot \text{seg2}_{it} + \beta_3 \cdot \text{seg3}_{it} + b_i$$

And we compared them using AIC, BIC, and R^2, as the Table 4 shows, the fill model which included both segmented trends and binary indicators achieved the lowest AIC(4504.93)/BIC(4538.78) and highest marginal(0.18) and confitional(0.853) R^2, indicating it best captures both temporal changes and spatial variation in susceptibility. This model was therefore selected for detailed interpretation.

Table 4: Model Comparison using AIC, BIC, and R-square

	Model	AIC	BIC	Marginal_R2	Conditonal_R2
Model_null	Model_null	4702.96	4712.63	0.000	0.820
Model_trend	Model_trend	5127.01	5146.36	NA	NA
Model_trend_ID	Model_trend_ID	4541.84	4566.02	0.146	0.848
model_full	model_full	4504.93	4538.78	0.180	0.853

Figure 3 compares the actual and predicted proportions of measles-susceptible children in Glasgow from 1998–2012. The actual trend (blue) closely follows the full model's predictions (red dashed), indicating a good fit.

Susceptibility increased between 1998 and 2003, rose further after 2004, and declined sharply around 2010. These abrupt changes correspond to the full retractions of the Wakefield article, which are effectively captured by the model's jump terms (I2010).

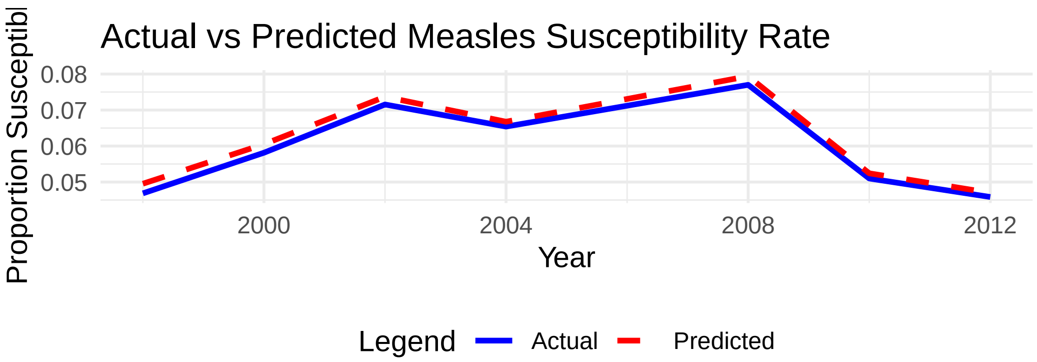

Figure 3: Actual vs Predicted Measles Susceptibility Rate

To assess the adequacy of the full generalized linear mixed model (GLMM), we examined two residual diagnostic plots. The left Figure 4 panel displays a QQ plot of simulated residuals. The residuals align closely with the 45-degree reference line, indicating that their distribution is consistent with the expected uniform distribution under a correctly specified model. This suggests that the model's distributional assumptions are appropriate, with no major deviation from uniformity.

The right Figure 4 panel shows a Pearson residuals vs fitted values plot. The residuals are symmetrically scattered around zero without noticeable curvature or funnel shape, implying that the model does not exhibit strong heteroscedasticity or systematic bias. While a few large residuals are observed at higher fitted values, these are limited and not indicative of serious model misfit.

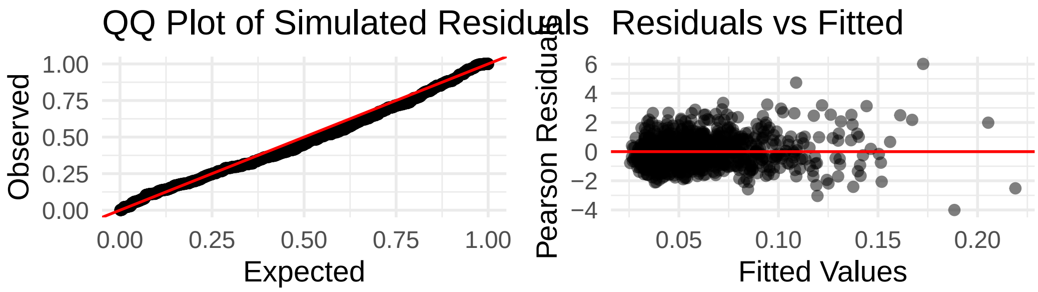

Figure 4: Residual Diagnostic Plots for GLMM(left) and Pearson Residuals vs Fitted Values for the Full Model(right)

Overall, this visual and statistical evidence strongly supports both project hypotheses. Glasgow experienced a significant change in measles susceptibility following the retraction of the Wakefield article. While the immediate effect in 2004 was an upward shift in susceptibility, a more substantial decline occurred around 2010, reflecting a delayed yet important recovery in public vaccine confidence.

Conclusion

Through this report, we have been analysing whether measles susceptibility changed in Glasgow following the partial retraction of the Wakefield article, and whether this change occurred in 2004; the year the article was retracted.

We used a simple Chi-square test for equality of proportions, which didn't lead us to finding any statistically significant difference pre/post 2004. Therefore, we fitted a more specific model; the GLMM with binomial family.

This model lead us to our findings: there was a significant *increase* in measles susceptibility from 2004-2010, indicating an initial lack of recovery from the damage due to the Wakefield article, and a prolonged hesitancy towards vaccinations following its partial retraction. However, from 2010 onwards, we found a significant decrease in susceptibility, indicating a strong and important, though late, response. This also aligns with the full discreditation of the article.

This implies the Wakefield article left a lingering effect on the societies of Glasgow; though the fear it instilled eventually dispersed as time went on.