Producing a generalised linear model to predict cases of measles

Denys Melnyk, Professor Keith Abrams

Introduction

Measles is a highly contagious disease caused by the *morbillivirus* virus. It is spread through airborne respiratory droplets, and can be transmitted when an infected person breathes, coughs or sneezes. It can cause severe disease, leading to complications and death. Measles primarily infects the respiratory tract and then spreads throughout the body, causing symptoms such as a high fever, cough, runny nose and a rash all over the body. Symptoms of measles usually begin 10–14 days after initial infection. A prominent rash is the most visible symptom, usually spreading over a period of 3 to 4 days. Most deaths from measles are caused by complications related to the disease and comorbidity, which can include:[1]

1. blindness;
2. encephalitis (an infection causing brain swelling and potentially brain damage);
3. severe diarrhea and related dehydration;
4. ear infections;
5. severe breathing problems including pneumonia.

According to WHO and gov.uk data, while measles can affect any age group, it is most common in children.

In the UK the MMR (Measles, Mumps and Rubella) vaccine is used to provide long-term protection against measles. It is administered in two doses, one at the age of 1 year old, and second at 3 years 4 month old.[2] According to a study conducted by Hungerford *et al.* on a sample of over 72000 children born between 1995 and 2012, children from more deprived socioeconomic groups are significantly less likely to receive both doses of the vaccine.[3]

In this report, a variation of a generalised linear model will be used to look at the theoretical number of cases in each region of England in 2022-2023, fitted on publicly available data, and assuming that the vaccination rates hit the 95% mandated by NHS and WHO.

[1] World Health Organization, Measles, Updated July 2024, cited September 2024. Available from: https://www.who.int/news-room/fact-sheets/detail/measles

[2] NHS, MMR (measles, mumps and rubella) vaccine, Updated March 2024, cited September 2024. Available from: https://www.nhs.uk/vaccinations/mmr-vaccine/

[3] HUNGERFORD D, MACPHERSON P, FARMER S, GHEBREHEWET S, SEDDON D, VIVANCOS R, et al. Effect of socioeconomic deprivation on uptake of measles, mumps and rubella vaccination in Liverpool, UK over 16 years: a longitudinal ecological study. Epidemiology and Infection. 2016;144(6):1201–11. doi:10.1017/S0950268815002599)

Data and Methods

Data

The accessed data is publicly available, posted on gov.uk and nhs.uk in yearly reports. Gov.uk publishes the data on the number of cases of measles[4,5], split by the region and the age group of the sampled patients, as well as population censuses and projections[6] for each region in England. The NHS publishes vaccination reports[7] each year covering the UK, which includes percentage of children who received the first and second doses of MMR vaccine. Data only covering England was used, and for the first dose due to limitations of the project. Initially age categories were also used, as split by the Measles reports by gov.uk, however they were discarded due to one of the age categories including over 50% of the population, rendering this part of data inefficient.

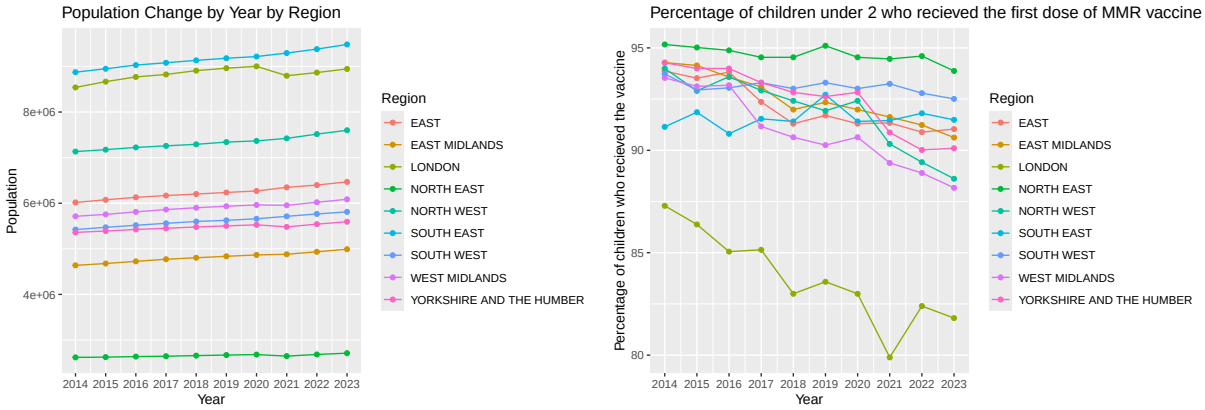

Figure 1: Left: Population by Region, Right: Vaccination Rates by Region

Trends

Figure 1 shows that the population in England has been steadily increasing, however there was a small decrease in London in 2020-2021, possibly due to pandemic. Vaccination rates, have been steadily decreasing across most regions, with a noticeable increase observed in most regions in 2019, aligning with a 2018-2019 measles outbreak. London has a significantly larger population than other regions, except for East England, as well as low vaccination rates. As of 2022, vaccination rates in London were around 10% lower than the rest of England, with the lowest vaccination rate recorded in 2021, almost 2.5% lower than in 2020. Amongst the other regions, 3 have a noticeable decrease of over 2% in vaccination rates post-pandemic, specifically West Midlands, Yorkshire and North West England.

Figure 2 shows when significant outbreaks of measles occurred, with the 2018-2019 outbreak being clearly visible on the graphs. A significant decrease in cases can be seen in 2020-2021, likely due to England-wide lockdown that occurred due to COVID-19 pandemic. Post-pandemic, the number of cases has been steadily increasing each year, with West Midlands, Yorkshire and London having the largest increase of measles infections.

[4]GOV.UK, Confirmed cases of measles in England and Wales by region and age: 2012 to 2022, Updated November 2023, cited September 2024. Available from: https://www.gov.uk/government/publications/measles-confirmed-cases/confirmed-cases-of-measles-in-england-and-wales-by-region-and-age-2012-to-2014

[5]GOV.UK, Confirmed cases of measles in England by month, age and region: 2023, Updated August 2024, cited September 2024. Available from: https://www.gov.uk/government/publications/measles-epidemiology-2023/confirmed-cases-of-measles-in-england-by-month-age-and-region-2023

[6]Office for National Statistics, Population estimates, Updated July 2024, cited September 2024. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates

[7]NHS, Childhood Vaccination Coverage Statistics, Updated September 2023, cited September 2024. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-immunisation-statistics

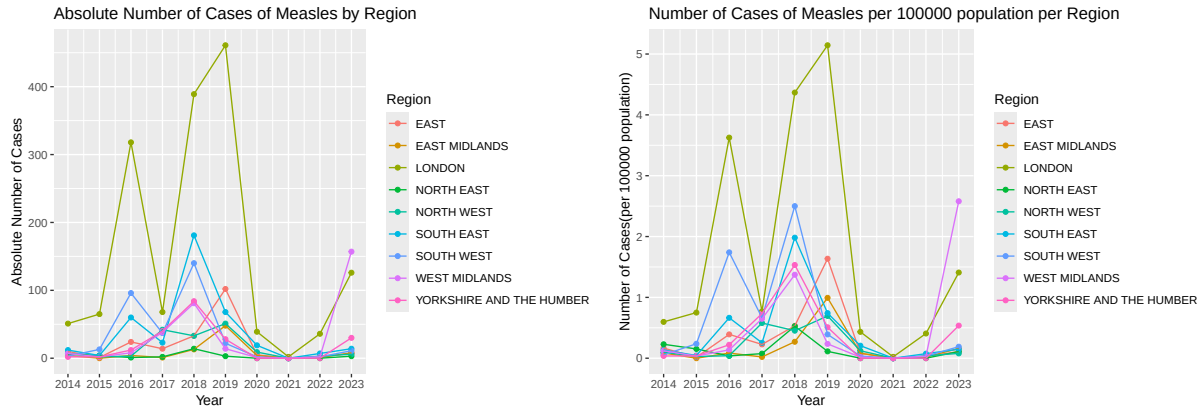

Figure 2: Number of Cases of Measles by Region. Left: Absolute number of Cases, Right: Number of Cases adjsuted by the population

Poisson Regression

Poisson regression is a generalized linear model, following a response function $g(\mu_i) = exp(\beta^T x_i) > 0$ where μ_i is an expectation of the i'th entry, x_i - vector of the predictor variables, and β - parameter vector.

Initially a poisson regression was fit, but the abundance of Regions with 0 cases made the model fail. This is due to one of the assumptions of the poisson regression is $\mu_i > 0$. To address this a modification of a poisson model, a zero-inflated poisson model, was used.

Zero-inflated poisson regression (ZIP regression) uses a distribution that is a mixture of two other distributions, X and Y, such that the probability mass function of Y is:

$$f(Y = y) = \begin{cases} p, & \text{if } y = 0 \\ 1 - p, & \text{if } y = X \\ 0, & \text{otherwise} \end{cases}$$

With X being a poisson distribution with the parameter μ. The ZIP regression uses logistic regression to model Y, and standard poisson regression to model X.

After the model was constructed and fitted, it was used to predict the number of cases of measles in each region post-pandemic, assuming vaccination rates of 95%.

Results

validity of the model & methodology issues

As discussed above, the standard poisson model failed due to violation of one of the constraints. ZIP model demonstrates a reasonable diagnostics plot (*Figure 3*). The variance of a Poisson distribution depends on the parameter μ, so a larger spread of residuals is to be expected for the larger fitted values. Due to fitted values being count data, there is a slight deviation from normality, however this is not significant for a count based model such as the one implemented.

While the model is theoretically passable, there are two major methodology issues - the time period is done in years, which greatly limits data available and makes any sort of time series model a poor alternative. Preferably, quarterly or monthly data would have been used, however this was not available. The second issue is due to measles' classification as a mostly eradicated infectious disease, with most outbreaks being localised in smaller communities rather than widespread, however this data is not publicly available.

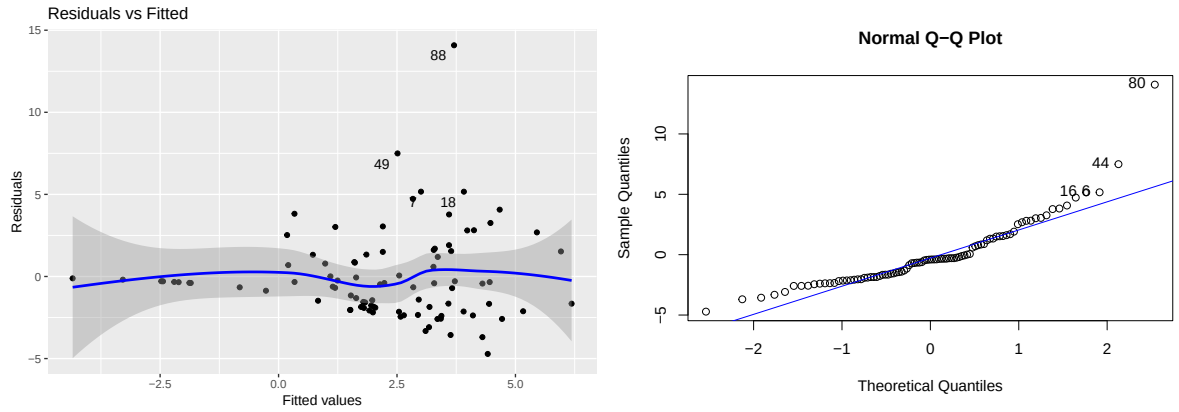

Figure 3: Diagnostics Plots. Left: Residual vs Fitted. Right: QQ Plot

Prediction with target vaccination rate post pandemic

Region	Year	Cases	Fitted Data	Predicted Data at 95% Vaccination
EAST MIDLANDS	2022	1	1	0
	2023	8	9	4
EAST	2022	2	3	1
	2023	7	18	8
LONDON	2022	36	26	2
	2023	126	174	14
NORTH EAST	2022	0	0	0
	2023	3	2	2
NORTH WEST	2022	3	3	1
	2023	6	23	7
SOUTH EAST	2022	7	4	2
	2023	14	31	15
SOUTH WEST	2022	2	4	3
	2023	11	28	17
WEST MIDLANDS	2022	2	5	1
	2023	157	40	11
YORKSHIRE AND THE HUMBER	2022	0	4	1
	2023	30	26	10

Discussion

Output results

The model tends to overshoot the value for some of the regions, however the fitted values are reasonably close to the real number of cases for the majority of regions for 2022-2023. Crucially, the output of the model provides a parameter that was used to produce the data rather than the actual predicted number of cases, so even if the predicted parameter is greater than the actual number of cases, the model is still valid. The exceptions are the few outputs for East England, North West England and West Midlands in 2023, which deviate from the actual results greatly.

Regions with low vaccination rates tend to show a drastic decrease of number of cases from the fitted values to prediction at a 95% vaccination rate. The regions which had higher actual vaccination rates tend to have a much smaller difference. The change in number of cases for 2023 is much larger than in 2022. East England is one of the regions which had a drastic decrease in actual number of cases of measles post-pandemic, but due

to the region having second largest number of cases adjusted by population, the model greatly overestimates the parameter for this region. Similar occurrence can be observed with North West England and West Midlands, due to 2018 and 2023 outbreaks respectively.

Outcomes

Through the produced model it can be seen that when the target vaccination rate of 95% is reached, there is a large decrease in theoretical incidents of measles compared to the actual data where this vaccination rate is not achieved. This is further suggested by there being a low amount of deviation between predicted and actual cases where vaccination rate is closer to the target. This is observed despite methodology issues, however this could be repeated in the future with adjustments and possibly more data to see if the observed trends still hold.

References

[1]: World Health Organization, Measles, Updated July 2024, cited September 2024. Available from: https://www.who.int/news-room/fact-sheets/detail/measles

[2]: NHS, MMR (measles, mumps and rubella) vaccine, Updated March 2024, cited September 2024. Available from: https://www.nhs.uk/vaccinations/mmr-vaccine/

[3]: HUNGERFORD D, MACPHERSON P, FARMER S, GHEBREHEWET S, SEDDON D, VIVANCOS R, et al. Effect of socioeconomic deprivation on uptake of measles, mumps and rubella vaccination in Liverpool, UK over 16 years: a longitudinal ecological study. Epidemiology and Infection. 2016;144(6):1201–11. doi:10.1017/S0950268815002599)

[4]: GOV.UK, Confirmed cases of measles in England and Wales by region and age: 2012 to 2022, Updated November 2023, cited September 2024. Available from: https://www.gov.uk/government/publications/measles-confirmed-cases/confirmed-cases-of-measles-in-england-and-wales-by-region-and-age-2012-to-2014

[5]: GOV.UK, Confirmed cases of measles in England by month, age and region: 2023, Updated August 2024, cited September 2024. Available from: https://www.gov.uk/government/publications/measles-epidemiology-2023/confirmed-cases-of-measles-in-england-by-month-age-and-region-2023

[6]: Office for National Statistics, Population estimates, Updated July 2024, cited September 2024. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates

[7]: NHS, Childhood Vaccination Coverage Statistics, Updated September 2023, cited September 2024. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-immunisation-statistics

Research supported by Warwick URRS scheme and Warwick Statistics.